A Utility Driven Change of Measure

Author:
Kendall Carolissen

Supervisors:
A/Prof. David Taylor
A/Prof. Thomas McWalter

Dissertation submitted in partial fulfillment for the M.Phil degree in Financial Mathematics at the School of Economics.

May 26, 2014

University of Cape Town

Abstract

We demonstrate how a change of probability measure can be carried out based on the risk preference of a representative investor. Using the stochastic discount factor and the Radon-Nikodým derivative, we are able to obtain the risk-neutral measure given a real world measure and a preference structure defined by a utility function. This methodology is then used to attribute the sources of skewness in the risk-neutral measure.

Keywords: Utility theory, Risk aversion, Probability measures, Risk-neutral skewness.

Plagiarism declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and to pretend it is one's own.
2. Each significant contribution to, and quotation in, this report from the work of other people has been attributed, and has been cited and referenced.
3. This project is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

May 26, 2014

Acknowledgements

I would like to express my gratitude to my supervisors, Associate Professor David Taylor and Associate Professor Thomas McWalter, for their guidance and patience during the writing of this dissertation. For their invaluable assistance, I am truly grateful.

I would also like to thank the BSM team at Old Mutual for their financial support throughout my studies. I also thank Mashudu Makananise, whose expert opinion was instrumental in the production of the final document.

Finally, I would like to thank my parents for their support and motivation throughout my studies.

Contents

Chapter 1

Introduction

Understanding the relationship between the real world measure and the risk-neutral measure (martingale measure) is fundamental to asset pricing theory. The relationship between these measures, implied by the market, holds information regarding the risk aversion in the market (Bliss and Panigirtzoglou; 2004). We seek to investigate the link that exists between risk aversion, and the real world and risk-neutral measures.

We briefly describe the probability measures discussed in this dissertation. Firstly, we note the link between probability densities and probability measures. A probability density function uniquely characterises a probability measure; this is shown in Chapter 2. Hence, if we have the density function, we have the probability measure.

We now define a real world measure and a risk-neutral measure. The real world measure is the subjective probability measure describing the distribution of an underlying asset at a point in time; also known as the *physical measure*. This measure, and the corresponding distribution, are unknown. We can, however, estimate the physical distribution. A time-series of historical asset prices is often used to estimate the physical density. The estimated distribution is used as a proxy for the unknown physical distribution. While this is a standard method for estimating the physical distribution, note that this makes the implicit assumption that the physical distribution is stationary. The estimated physical distribution obtained from a time-series of market data is referred to as the *historical distribution* or the *statistical distribution*.

The risk-neutral measure is the probability measure under which the discounted asset price is a martingale. This is also known as the equivalent martingale measure (EMM). In a complete market model, this measure is unique (Björk; 2009). The risk-neutral measure is the measure under which we price options. Therefore, given the market price of options, we can derive the market implied estimate of the risk-neutral density. The risk-neutral density can be derived from a cross-section of option prices at a specific point in time (Detlefsen et al.; 2010), where option prices can be obtained from an implied volatility surface.

The risk-neutral distribution implied by option price data exhibits negative skewness (Bakshi et al.; 2003). This skewness translates directly to the Black-Scholes implied volatility smile (Bakshi et al.; 2003).

This dissertation aims to make the following specific contributions. Firstly, we seek to investigate how to change probability measure using utility and risk aversion. We review the methodology for obtaining a risk-neutral distribution of returns from the real world distribution of returns, changing probability measure based on the risk aversion of a representative agent. We then aim to attribute the sources of skewness in the risk-neutral distribution to specific theoretical components, following the research of Bakshi et al. (2003). We investigate the effect of risk aversion and the moments of the real world distribution on the risk-neutral distribution, and thus illustrate how risk aversion and the higher order moments of the real world distribution contribute to a negatively skewed risk-neutral distribution. The result shows that negative skewness in the risk-neutral distribution is a result of a combination of negative skewness and fat-tails in the real world distribution. The extent to which the heavy tails affect the risk-neutral skewness is related to the risk preference specification of the representative investor.

Using a change of measure based on the risk aversion of a representative agent, we can draw the link amongst the physical distribution, the risk-neutral distribution and utility preference. The Radon-Nikodým theorem is used to engineer a change of measure where we incorporate the risk aversion of a representative agent via a specific utility function. The methodology here relies on the specification of a stochastic discount factor, which is used to price contingent claims in a consumption-based equilibrium model. The methodology to change probability measures using utility functions is engineered to preclude arbitrage opportunities; the risk-neutral distribution obtained is an equivalent martingale measure.

The work in this dissertation is based largely on the research of Bakshi et al. (2003). In their paper they make several significant contributions. Using a specific form for the utility function, they illustrate how the physical distribution is exponentially tilted in order to obtain the risk-neutral distribution. Bakshi et al. (2003) characterise negative risk-neutral skewness in the risk-neutral distribution in a model-free manner. By using their model-free framework, we ensure that the attribution of risk-neutral skewness is not an artifact of the asset pricing model used. Bakshi et al. (2003) attribute risk-neutral skewness to three main components; skewness and kurtosis in the physical distribution, and risk aversion. They go on to show how negative skewness in the risk-neutral measure contributes to the Black-Scholes implied volatility smile, providing theoretical motivation and empirical evidence that a greater negative skew in the risk-neutral distribution corresponds to a steeper implied volatility smile, and that fatter tailed risk-neutral distributions contribute to flatter implied volatility smiles. The work in their paper is based on the fundamental premise that we can change probability measure using the preference structure of individuals, defined by utility functions. We explore this concept in this dissertation, and attempt to motivate and describe the methodology for changing probability measures using utility functions.

The relationship between risk aversion, and the real world and risk-neutral measures is not a novel concept. Despite the model's restrictive assumptions, it is tractable, and its simplicity allows for broad applications. Extant literature pertaining to this relationship displays its application to a variety of real world market situations. The focus being mainly on deriving one specific component of the relationship given the other two, and the interpretation and empirical testing thereof.

Ait-Sahalia and Lo (2000) posit a relationship between the risk-neutral distribution, the real world distribution and a specific risk preference implied by a utility function. The relationship derived shows that, under certain conditions, any two of the three components will imply the third. Therefore a view of the physical distribution of the underlying market return, and the risk-neutral distribution implied by the options market, would imply a utility function for the market as a whole, and thus the risk preference of the market.

Similarly, Detlefsen et al. (2010) investigate empirical market utility functions implied by the market data of the DAX and the DAX options market. The findings show that the empirical utility functions illustrate a region where the representative investor is risk seeking, rather than being constantly risk averse, as is suggested by classic economic utility theory. It is shown that individual investors are heterogeneous, with utility functions that contain a point at which risk attitude switches between risk aversion and risk-seeking. This study provides empirical evidence for risk-seeking investors, particularly in bear markets.

There are also many studies seeking to take advantage of the information implicit in the risk-neutral distribution implied by the market, and the risk aversion of the market. Bliss and Panigirtzoglou (2004) investigate the "risk premium implicit in option prices", via a utility driven change of measure. Similarly, Anagnou et al. (2002) investigate the future density forecast abilities of the market implied risk-neutral distribution. In both papers the option-implied risk-neutral density and a hypothesized utility function are used to estimate the physical density of the underlying index. The resulting density is then used to forecast the distribution of future values of the underlying index.

Liu et al. (2004) adopt a different approach in assessing the forecast abilities of an implied real world distribution. In this study the real world distribution is derived from an assumed risk preference and the risk-neutral distribution implied by the options market. This distribution is then compared with the physical density obtained from a historical time series of the market. The results show that "parametric densities derived from option prices have more explanatory power than historical densities" (Liu et al.; 2004).

Employing the same strategy, Rosenberg and Engle (2002) compare the historical distribution of the underlying market values to the physical distribution implied by the observed risk-neutral distribution and an assumed preference structure. Rosenberg and Engle (2002) state that the investor's preference for payoffs is encapsulated in the relationship between the physical and risk-neutral measures.

These studies do present certain model limitations. These include the assumption of stationary and symmetrical physical distributions, erroneous aggregation of consumption data, simplified utility structures, and artifacts of parametric modeling.

The aggregation of consumption data proves to be problematic in empirical investigations. We illustrate how we can use market index prices as a proxy for consumption in the utility driven change of measure model. This alleviates the issue of erroneous or incomplete aggregate consumption data.

Following the methodology of Bakshi et al. (2003), we conduct our analysis in a model-free manner, so as to remove the effect of any results being attributed to the form of the model used. We show explicitly how Bakshi et al. (2003) extend the admissible class of utility functions to include a general utility function.

Given that the proposed change of measure methodology is reliant on utility functions, we provide a comprehensive treatment of utility theory in the appendix. We aim to motivate the use of utility functions and risk aversion coefficients as adequate and suitable measures of an individual's preferences and attitude to risk. An important section of this chapter deals with the general form of the utility function; a utility function implied by hyperbolic absolute risk aversion. This will be used in the generalisation of the utility driven change of measure.

The structure of this dissertation is as follows; the Chapter 2 presents the mathematical preliminaries for the subsequent chapters. Briefly, we show that, where it exists, a density function fully characterises its corresponding probability measure. We also include a section on the Radon-Nikodým derivative. Readers with the requisite knowledge of probability theory may skip this chapter.

In Chapter 3 we introduce the change of measure that incorporates the utility specification of a representative agent. This specification allows us to derive the relationship amongst the risk-neutral density, the physical density, and utility preferences. Some examples are provided in this chapter, based on the Black-Scholes option pricing model. The first example derives the risk-neutral density given an assumed form for the physical density and the utility function. The second example derives the utility function implied by the Black-Scholes asset pricing model.

The sources of negative skewness in the risk-neutral measure are investigated in the Chapters 4 and 5. In these chapters we relate the risk-neutral index skew to the higher order moments of the physical distribution, based on the change of measure methodology developed in Chapter 3. The outcome observed is that skewness in the risk-neutral distribution is a consequence of risk aversion and the magnitude of the higher order moments of the physical distribution. In Chapter 4, following the research of Bakshi et al. (2003), this is done using power utility. We demonstrate how the change of measure alters the moments of the physical distribution in order to obtain the risk-neutral moments. The sources of skewness in the risk-neutral measure indicate that a negatively skewed risk-neutral distribution is possible even in the case of a symmetrical physical distribution. Bakshi et al. (2003) provide a general form for the utility function, stating that the attribution of the risk-neutral skew extends to this general utility function. Chapter 5 illustrates how the general form of Bakshi et al. (2003) can be used to develop a similar attribution of the risk-neutral skew. Conclusions and further remarks are presented in Chapter 6.

Chapter 2

Mathematical Preliminaries

This chapter provides the mathematical preliminaries required for use in the subsequent chapters of this dissertation. We aim to state two main results in this chapter. The first is the relationship between a probability measure and the corresponding density function. The second result is the Radon-Nikodým theorem as it pertains to probability measures, and the application thereof. It is assumed that the reader has a basic knowledge of probability theory. Appendix B provides an introduction into the probability theory required for the understanding of this chapter. The reader with the requisite knowledge of probability theory may skip this chapter.

The first section details the relationship between a probability measure and the corresponding density function. The result here is that a probability distribution function uniquely characterises a probability measure. This allows us, in the subsequent chapters, to change probability measures by altering the form of the distribution function or, where it exists, the corresponding density function.

We consider the Radon-Nikodým theorem as the means to change probability measure. Chapter 3 will use this in order to derive a functional form for the risk-neutral density in terms of the real world density.

2.1 Probability Measures and Density Functions

This section describes the link between probability measures and the functions used to describe them. The work in this section is based on that of Jacod and Protter (2003).

Consider a probability function as defined in Definition B.12. Consider now the case where the sample space is the set of real numbers; $\Omega = \mathbb{R}$. Let $\mathfrak{B}$ be the Borel set of real numbers; that is, the σ-algebra generated by the open subsets of $\mathbb{R}$.

Definition 2.1. Given the measurable space defined by $(\mathbb{R}, \mathfrak{B})$; let $f : \mathbb{R} \to \mathbb{R}$ be a real valued function, such that f is measurable with respect to $\mathfrak{B}$. This function is called *Borel measurable* and is known as a *Borel function*.

Definition 2.2. The function defined as

$$F(x) = \mathbb{P}((-\infty, x])$$

for a probability measure $\mathbb{P}$ on the measure space $(\mathbb{R}, \mathfrak{B})$, is called the *distribution function* induced by $\mathbb{P}$ on $(\mathbb{R}, \mathfrak{B})$.

Note that the arguments taken by the distribution function F are real numbers, while the arguments taken by the probability function $\mathbb{P}$ are events, or subsets of the σ-algebra $\mathcal{F}$.

Theorem 2.3. *The distribution function induced by the probability measure* $\mathbb{P}$ *on the measure space* $(\mathbb{R}, \mathfrak{B})$*, uniquely characterizes the probability. That is, if there exists another probability measure* $\mathbb{Q}$ *such that*

$$G(x) = \mathbb{Q}((-\infty, x]), \qquad \text{for all } x \in \mathbb{R},$$

and $F = G$*, then* $\mathbb{P} = \mathbb{Q}$*.*

Proof. See Jacod and Protter (2003, Theorem 7.1, p. 39). □

The above theorem implies that we can determine a unique distribution function F from a probability measure $\mathbb{P}$. Therefore, from the distribution function F, the probability of any Borel set $A \in \mathfrak{B}$, $\mathbb{P}(A)$ can be determined. See Jacod and Protter (2003, Theorem 7.2, p. 40) for a characterisation of distribution functions.

Definition 2.4. Let the function f be a non-negative Borel measurable function, that is Riemann-integrable with $\int_{-\infty}^{\infty} f(x)dx = 1$, defined by

$$F(x) = \int_{-\infty}^{x} f(y)dy, \tag{2.1}$$

for a distribution function of a probability measure defined on the real numbers. The function f is known as the *density function* of the probability measure $\mathbb{P}$; see Definition B.44.

The following theorem relates the uniqueness of a probability measure to the Lebesgue measure of the corresponding density function. It shows that where the density function for a probability measure exists, it characterizes the probability measure entirely.

Theorem 2.5. *A function* f *on the real numbers* $\mathbb{R}$*, that is non-negative Borel measurable, is the density function of a probability measure* $\mathbb{P}$ *on the measurable space* $(\mathbb{R}, \mathfrak{B})$ *if and only if the function integrates to one over its domain*

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Where this function exists, it characterizes the probability measure $\mathbb{P}$ *entirely. Any other function* f' *which is non-negative and Borel measurable, where* $m(f' \neq f) = 0$*, is a density function for the same probability measure. The function* m *is the Lebesgue measure, defined in Section B.6.*

We also have the result that where a probability measure $\mathbb{P}$ *on the measurable space* $(\mathbb{R}, \mathfrak{B})$ *has a density function, the density is determined up to a set of Lebesgue measure zero; if* f *and* f' *are both density functions for the probability measure* $\mathbb{P}$ *on* $(\mathbb{R}, \mathfrak{B})$*, then we have* $m(f' \neq f) = 0$*.*

Proof. See Jacod and Protter (2003, Theorem 11.3, p. 78). □

Thus we have described in this section the unique relationship between a probability measure on the measure space $(\mathbb{R}, \mathfrak{B})$ and the corresponding distribution function and density function.

2.2 Radon-Nikodým Change of Measure

The ability to change between absolutely continuous probability measures is closely linked to arbitrage pricing (Björk; 2009). The Radon-Nikodým theorem is the mechanism by which such a change of measure is performed. The theorem is stated here in the context of probability measures as they pertain to asset pricing.

Definition 2.6. Consider a measurable space $(\Omega, \mathcal{F})$ that is equipped with two separate probability measures $\mathbb{P}$ and $\mathbb{Q}$, such that the triples $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega, \mathcal{F}, \mathbb{Q})$ are probability spaces. The set of possible states is given by Ω, and $\mathcal{F}$ is a σ-algebra on the subsets of Ω.

- If we have
$$\mathbb{P}(A) = 0 \quad \Rightarrow \quad \mathbb{Q}(A) = 0$$
for all $A \in \mathcal{F}$, then we say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$ on $\mathcal{F}$. We denote absolute continuity as $\mathbb{Q} \ll \mathbb{P}$.

- If $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \gg \mathbb{Q}$, then $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent measures*. We will denote equivalence between measure as $\mathbb{P} \sim \mathbb{Q}$. Therefore, for $\mathbb{P} \sim \mathbb{Q}$; $\mathbb{P}(A) = 0 \Leftrightarrow \mathbb{Q}(A) = 0$, $\forall A \in \mathcal{F}$.

Equivalent probability measures agree on states that are impossible (i.e. states that occur with zero probability). We state the Radon-Nikodým theorem for probability measures.

Theorem 2.7. *(Radon-Nikodým) Let $\mathbb{P}$ be a probability measure defined on the measurable space $(\Omega, \mathcal{F})$, and $\mathbb{Q}$ a finite measure on the same measurable space. If $\mathbb{Q} \ll \mathbb{P}$, then there exists a random variable ζ that is non-negative and integrable, where $\mathbb{E}^{\mathbb{P}}[\zeta] = 1$, such that*

$$\mathbb{Q}[A] = \mathbb{E}^{\mathbb{P}}[\mathbb{I}_A \zeta], \quad \forall A \in \mathcal{F},$$

where $\mathbb{E}^{\mathbb{P}}$ indicates the expectation with respect to the probability measure $\mathbb{P}$, and $\mathbb{I}$ is the indicator function as per definition (B.15). ζ is unique a.s.; that is, if we have another random variable ζ' satisfying the properties of ζ, then $\zeta = \zeta'$ a.s. If we have $\mathbb{Q} \sim \mathbb{P}$, then $\zeta > 0$.

Proof. See Vestrup (2003, Radon-Nikodym Theorem, p. 377). □

Definition 2.8. For $\mathbb{Q}$ defined as a probability measure on $(\Omega, \mathcal{F})$, such that $\mathbb{Q} \ll \mathbb{P}$, the random variable ζ described in Theorem 2.7 is known as the *Radon-Nikodým derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$. It is denoted

$$\zeta = \frac{d\mathbb{Q}}{d\mathbb{P}}. \tag{2.2}$$

If $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}$, then we have

$$\begin{aligned}\mathbb{Q}[A] &= \mathbb{E}^{\mathbb{P}}[\mathbb{I}_A \zeta] \\ &= \int_{\omega \in A} \zeta(\omega) \mathbb{P}(\omega) d\omega.\end{aligned}$$

Now, for a probability measure $\mathbb{Q}$ on the measurable space $(\Omega, \mathcal{F})$, where $\mathbb{Q} \sim \mathbb{P}$ and $\zeta = \frac{d\mathbb{Q}}{d\mathbb{P}} > 0$, we have

$$\begin{aligned}\mathbb{P}[A] &= \mathbb{E}^{\mathbb{P}}[\mathbb{I}_A] \\ &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{I}_A \frac{1}{d\mathbb{Q}/d\mathbb{P}} \frac{d\mathbb{Q}}{d\mathbb{P}}\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{I}_A \frac{1}{d\mathbb{Q}/d\mathbb{P}} \zeta\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\mathbb{I}_A \frac{1}{d\mathbb{Q}/d\mathbb{P}}\right],\end{aligned}$$

for all $A \in \mathcal{F}$. Given that ζ is unique, we can write

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{1}{d\mathbb{P}/d\mathbb{Q}} \text{ a.s.}$$

Definition 2.9. (Jacod and Protter; 2003) Consider a measurable space $(\Omega, \mathcal{F})$, equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$. For $\mathbb{Q}$ defined as a probability measure on this space, where $\mathbb{Q} \ll \mathbb{P}$, the *density process* of $\mathbb{Q}$ with respect to $\mathbb{P}$ is defined as

$$Z_t = \mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\,\middle|\,\mathcal{F}_t\right],$$

for all $t > 0$. $\mathcal{F}_t$ is the filtration generated up to time t, and Z_t is an integrable martingale. We can think of Z_t as the Radon-Nikodým derivative for a change of measure conditional on the information available up to time t.

Chapter 3

Utility Driven Change of Measure

This chapter presents the relationship between utility and the changing of probability measures. Using an equilibrium asset pricing model, we show that there exists a relationship between the real world measure, the risk-neutral measure, and risk aversion. This relationship allows us to change probability measure using utility functions.

The chapter begins with a treatment of the stochastic discount factor. The aim is to relate the stochastic discount factor, which prices the asset in equilibrium, to a no-arbitrage asset price, in order to derive a relationship between risk aversion (utility) and probability measures. This will allow us to incorporate utilities and risk aversion into the asset pricing procedure. The following section details how we incorporate utility of asset prices rather than utility of consumption. The third section uses stochastic discount factors together with equilibrium pricing in order to derive the relationship between the real world measure, the risk-neutral measure, and risk aversion. We include simple examples illustrating the application of the utility driven change of measure mechanism. These examples show the market risk preference implied by the Black-Scholes model (even though the Black-Scholes model is specified in a preference-free manner). The chapter concludes with a brief discussion of the pricing kernel puzzle, as well as a summary of a method for obtaining an empirical estimate for the risk-neutral distribution.

3.1 Stochastic Discount Factors

This section is based on the asset pricing framework of Cochrane (2001) and Campbell et al. (1996). The purpose of this section is to derive the stochastic discount factor (hereafter SDF). The SDF describes the relationship between asset prices and investors' preferences for consumption and saving.

Given a rational representative investor, seeking to maximise utility of consumption, we use an intertemporal equilibrium model to determine the price of an asset given its payoff. We propose the existence of an SDF, where the price of the asset at the current time is given by the expected value of the product of an asset payoff at a future time and the SDF. This equilibrium asset pricing model, incorporating a utility maximising investor, leads to a link between the SDF and the consumption preferences of investors. Thus we can use the stochastic discount factor to relate asset prices to the underlying preferences of investors (Campbell et al.; 1996). We begin by defining the SDF.

Definition 3.1. The *stochastic discount factor* is the variable used to discount an asset's payoff in order to obtain its price. The asset price is equal to the expected value of the product of the asset's payoff and the SDF (Cochrane; 2001).

The SDF provides a general methodology for asset pricing. It allows us to incorporate utility functions (more specifically, marginal utility) into the pricing of assets. This is important as it allows us to draw the link between utility functions and changes of measure.

The framework for SDF is the *basic asset pricing equation* described by Cochrane (2001). This equation states that the price of any asset is equal to its expected discounted payoff, using marginal utility to discount the asset payoff. We now define the marginal utility of consumption.

Definition 3.2. *Marginal utility of consumption* is the change in level of utility for an infinitesimal change in the level of consumption. It is the rate of change of utility with respect to consumption and can be represented mathematically as the first derivative of utility with respect to consumption; $U'(C) = \frac{\partial U(C)}{\partial C}$, for consumption level C and utility of consumption given by the function $U(C)$.

In order to derive the SDF, consider the following discrete time model. Consider an economy that has two discrete time periods, t and T, where $T > t$ and consumption is only allowed at each of these time points. Consider a representative investor who has to choose between consumption at times t and T. Assuming a market that is frictionless and complete, define an asset with a terminating payoff at time T given by x_T. The value x_T is a random variable. If an investor has one unit of this asset at time t, then at time T the value of that asset is x_T. This model can be easily extended to allow for dividends at the payoff date (Cochrane; 2001). We omit the extension here for the sake of simplicity.

Definition 3.3. At *intertemporal utility function* is a function describing the utility preferences of an investor over current consumption, and consumption at a future date.

The investor has an intertemporal utility of consumption $\mathcal{U}$ given by

$$\mathcal{U}(C_t, C_T) = U(C_t) + \beta \mathbb{E}_t[U(C_T)],$$

where $U(C)$ represents a subjective utility function of consumption and β is a *subjective discount factor*. The expectation operator subscripted by t indicates that this is the expectation taken with respect to information given up to time t. The curvature of $\mathcal{U}$ describes the investor's aversion to risk, as well as the investor's impatience (Cochrane; 2001). The degree of impatience is stated mathematically as β, the *subjective discount factor*. This impatience coefficient drives the level of *intertemporal substitution*. Intertemporal substitution is described as the trade-off between consuming now (at time t), or delaying consumption to time T; or, how the investor chooses to forgo consumption now in favour of consumption later. The subjective utility function $U(C)$ is increasing and concave, representing an investor that is non-satiated and exhibits declining marginal utility of consumption; $U'(C) \geq 0$ and $U''(C) \leq 0$. This implies a risk averse investor. Von Neumann-Morgenstern utility functions (described in Appendix A) are then viable choices for the utility function $U(C)$. The amount consumed at time step T, C_T, is a random variable.

Assume now that the investor can trade in any amount of the asset at time t at the prevailing price p_t. Assume further that this investor will act so as to maximise expected utility. We construct an optimisation problem in order to determine the amount of the asset to be held. Let initial wealth at time t be given by e_t, and let ξ be the portion invested in the asset. The optimisation problem is then

$$\max_{\{\xi\}} \left(U(C_t) + \mathbb{E}_t[\beta U(C_T)] \right), \tag{3.1}$$

subject to the constraints

$$C_t = e_t - \xi p_t,$$
$$C_T = e_T + \xi x_T.$$

Substitute the constraints into Equation (3.1) and set the derivative with respect to ξ equal to zero. This results in the first-order conditions for the optimal level of consumption,

$$p_t U'(C_t) = \mathbb{E}_t \left[\beta U'(C_T) x_T \right]. \tag{3.2}$$

We interpret Equation (3.2) as the equilibrium conditions for an optimal consumption choice. The right hand side of Equation (3.2) is the loss in utility the investor experiences if another unit of the asset is bought at time t. The left hand side of Equation (3.2) shows the increase in expected utility the investor will gain from an additional unit of the payoff at time T. The investor will trade in the asset until this equilibrium is reached (Cochrane; 2001). This then leads to the central result of this section,

$$\begin{aligned} p_t &= \mathbb{E}_t \left[\beta \frac{U'(C_T)}{U'(C_t)} x_T \right] \\ &= \mathbb{E}_t \left[M_{t,T} x_T \right]. \end{aligned}$$

Definition 3.4. (Cochrane; 2001) The *central asset pricing formula*

$$p_t = \mathbb{E}_t \left[M_{t,T} x_T \right] \tag{3.3}$$

describes the asset price p_t at time t, given an investor's consumption preference over the period t, T and the asset payoff x_T.

Definition 3.5. The term $M_{t,T}$, defined as

$$M_{t,T} = \beta \frac{U'(C_T)}{U'(C_t)}, \tag{3.4}$$

is known as the *stochastic discount factor* (Cochrane; 2001). Rosenberg and Engle (2002) refer to this term as the *asset pricing kernel.* We discuss the pricing kernel, with reference to the *pricing kernel puzzle* further on.

Definition 3.6. Consider an individual whose preference for consumption is described by the utility function $U(C)$. The *marginal rate of substitution of consumption* for such an individual is defined as the rate at which he is willing to replace consumption at time T with consumption at time t. This is stated mathematically as $\frac{U'(C_T)}{U'(C_t)}$. The SDF is the product of the marginal rate of substitution of consumption and the subjective discount factor β.

The SDF is a representation of the preferences of an investor for various outcomes. The asset price is computed as the expectation of the product of the SDF and the payoff of the asset. This asset price can then be interpreted as a preference weighted average of the payoff. If we assume a probability density for the states of the payoff outcome, the SDF describes in full the asset price as well as the expected returns and risk premia involved (Rosenberg and Engle; 2002).

The following simple example illustrates the basic pricing equation for preferences defined by a power utility function. The following section considers a SDF based on a market stock price rather than consumption.

Example 3.7. Consider the individual with consumption preference at time t and time T described by the power utility function,

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma}.$$

For the marginal utility of consumption we then have $U'(C) = C^{-\gamma}$, leading to the SDF

$$\begin{aligned} M_{t,T} &= \beta \frac{U'(C_T)}{U'(C_t)} \\ &= \beta \frac{C_T^{-\gamma}}{C_t^{-\gamma}} \\ &= \beta \left(\frac{C_T}{C_t} \right)^{-\gamma}, \end{aligned}$$

where γ is the constant coefficient of relative risk aversion. The basic pricing equation can then be written as

$$p_t = \mathbb{E}_t \left[\beta \left(\frac{C_T}{C_t} \right)^{-\gamma} x_T \right].$$

3.2 The SDF and Utility of Asset Prices

The SDF specified above is based on a representative individual's utility of *consumption.* If we consider the representative individual to be the market, then the consumption is equal to the aggregate market consumption at a point in time. However, it is exceedingly difficult to obtain a reliable estimate of aggregate market consumption (Rosenberg and Engle; 2002).

It is impossible to obtain a measure of consumption at a point in time (instantaneous consumption) (Brown and Gibbons; 1985). The estimates for aggregate consumption will then most likely be time-aggregated, rather than instantaneous. The bias introduced by using a time aggregated measure of consumption rather than instantaneous consumption is known as *temporal aggregation bias* (Brown and Gibbons; 1985).

Temporal aggregation bias results in inaccurate measures of (instantaneous) consumption at discrete time intervals. Inaccurate consumption estimates undermine the integrity of comparing consumption at discrete intervals, C_t and C_T, since the measures will be imprecise. This diminishes the power of the information contained in these variables.

In addition to this, erroneous estimates of aggregate consumption data arise as a result of sampling error, data imputation, and the problem of defining an aggregate consumption bundle, amongst other reasons (Rosenberg and Engle; 2002).

Inaccurate estimation of consumption leads to unreliable estimates for the SDF (Rosenberg and Engle; 2002). In order to avoid intertemporal aggregation bias we use a proxy for instantaneous consumption. The issue of measurement error in aggregate consumption data is circumvented by making the assumptions that, for a representative individual, we can use the measure of wealth as a proxy for the level of consumption at a point in time (Ingersoll; 1987). Considering the derivation of the SDF in the previous section, this assumption is tantamount to a representative investor who maximises the intertemporal utility of wealth over the time period (t, T). When we consider the representative investor to be the market, then the level of wealth at time t can be approximated by the market index S_t. Thus the utility derived from a level of consumption at a point in time can be represented by the utility of the market index at that point in time,

$$U(C_t) \approx U(S_t).$$

By consequence, the marginal rate of substitution of consumption can then be approximated by the marginal rate of substitution based on the asset portfolio price,

$$\frac{U'(C_T)}{U'(C_t)} \approx \frac{U'(S_T)}{U'(S_t)}.$$

Therefore, we now have the SDF defined in terms of an asset price,

$$M_{t,T} = \beta \frac{U'(S_T)}{U'(S_t)}. \tag{3.5}$$

In the following section we describe the link between the Radon-Nikodým derivative and the SDF defined by Equation (3.5). The relationship derived will allow us to change probability measure using a utility function.

3.3 Change of Measure Equation

In this section we describe a methodology for a utility driven change of probability measure within the context of asset pricing. This section is based in large part on the derivation of the equation for a utility driven change of measure in Ait-Sahalia and Lo (2000) and Detlefsen et al. (2010). We seek to link the Radon-Nikodým derivative for a change of measure from Equation (2.2) to the SDF described by Equation (3.5). This link is established by equating the price of a contingent claim priced under two different methodologies; the no-arbitrage methodology used in the Black-Scholes model, and the SDF method of Cochrane (2001), described in Section 3.2. We begin by describing the asset price according to the SDF equilibrium model.

Let S_t be the price of an asset at time t, and $V(t)$ be the price of a contingent claim at time t, the value of which is derived from S_t, the underlying asset price at that time. The contingent claim $V(t)$ will be referred to as the *option*, and S_t the asset *underlying* the option. The payoff of the contingent claim at maturity, time T where $T > t$, is described by the function $\psi(S_T)$. The value of S_T is a random variable, the density of which is given by the density function $p(S)$, which has the corresponding probability measure $\mathbb{P}$.

Consider the representative investor with preferences described by a utility function U. This investor may purchase any amount of the contingent claim $V(t)$ with a single liquidating payoff of $\psi(S_T)$ at time T, where T is a future point in time and t is the current time. The investor chooses a consumption strategy in order to maximise expected utility of initial consumption and terminal consumption. Under the assumptions stated in Section 3.1 and Section 3.2, the price of the contingent claim at time t is given by

$$V(t) = \mathbb{E}_t^{\mathbb{P}}[M_{t,T}\psi(S_T)], \tag{3.6}$$

where the SDF $M_{t,T}$ is defined over the underlying asset price, and for the subjective discount factor β, as

$$M_{t,T} = \beta \frac{U'(S_T)}{U'(S_t)}, \quad T > t. \tag{3.7}$$

The expectation in Equation (3.6) is taken with respect to the subjective real world probability measure $\mathbb{P}$, conditional on the information up to time t. Equation (3.7) is then used to define a measure change. We have priced the contingent claim under the SDF methodology, we now price the contingent claim by the no-arbitrage approach.

Consider a market consisting of a risky stock and a risk-free asset. Suppose we have the same contingent claim in this market, where the payoff at time T is dependent on the terminal value of the stock. If we assume that the market is complete and arbitrage-free, then by the first and second fundamental theorems of asset pricing, we can represent the payoff of a contingent claim, discounted at the risk-free rate, as a martingale under the risk-neutral measure (martingale measure) (Björk; 2009). Thus for a risky stock S_t and a contingent claim $V(t)$ with payoff given by $V(T) = \psi(S_T)$, the no-arbitrage price of this contingent claim is simply the expectation of the payoff at maturity, discounted at the risk-free rate,

$$V(t) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}[\psi(S_T)], \tag{3.8}$$

where the expectation is taken with respect to $\mathbb{Q}$, the risk-neutral measure. Given the assumption of a complete market, this measure is unique, by the second fundamental theorem of asset pricing (Björk; 2009). Let the risk-neutral density $q(S)$ be defined as the density function corresponding to the martingale measure $\mathbb{Q}$. We then have

$$\begin{aligned} V(t) &= e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}[\psi(S_T)] \\ &= e^{-r(T-t)}\int q(s)\psi(s)ds. \end{aligned}$$

From Equation (3.8) we have the option price defined in terms of an expectation with respect to the measure $\mathbb{Q}$. We obtain this price under the measure $\mathbb{P}$ in order to derive the relationship between the probability measures (or equivalently, the probability densities) and the SDF;

$$\begin{aligned} V(t) &= e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}\left[V(T)\right] \\ &= e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}\left[\psi(S_T)\right] \\ &= e^{-r(T-t)}\int \psi(S)q(S)dS \\ &= e^{-r(T-t)}\int \psi(S)\frac{q(S)}{p(S)}p(S)dS \\ &= \int \psi(S)\zeta(S)p(S)dS \\ &= \mathbb{E}_t^{\mathbb{P}}\left[\zeta(S_T)\psi(S_T)\right], \end{aligned}$$

where $\zeta(S_T)$ is the ratio of the densities at terminal time T, multiplied by the deterministic discount factor;

$$\zeta(S_T) = \frac{q(S_T)}{p(S_T)}e^{-r(T-t)}.$$

This is similar to the price under the equilibrium model. Comparing $\zeta(S_T)$ to the SDF given by Equation (3.7), we have $M_{t,T} = \zeta(S_T)$, implying

$$\beta\frac{U'(S_T)}{U'(S_t)} = \frac{q(S_T)}{p(S_T)}e^{-r(T-t)},$$

giving us the link between the Radon-Nikodým derivative of a change of measure and utility functions

$$\frac{q(S_T)}{p(S_T)}e^{-r(T-t)} = \beta\frac{U'(S_T)}{U'(S_t)}. \tag{3.9}$$

Equation (3.9) represents the fundamental relationship between the real-world density, the risk-neutral density, and the utility (or risk aversion) of an agent. In equilibrium, assuming the absence of arbitrage, and given any two of the risk-neutral density, real world density, and utility function, we are able to obtain the third using the relationship stated in Equation (3.9). It is highly important to note that by design, any measure $\mathbb{Q}$ derived from this relationship will be a martingale measure, therefore ensuring that arbitrage is precluded.

Using Equation (3.9), we derive a formula for the risk-neutral density function q, based on a utility function U,

$$q(S_T) = \alpha e^{r(T-t)}\beta\frac{U'(S_T)}{U'(S_t)}p(S_T), \tag{3.10}$$

where α is a constant ensuring that Equation (3.10) integrates to one. Stated more formally,

$$\begin{aligned} q(S_T) &= \frac{e^{r(T-t)}\beta\frac{U'(S_T)}{U'(S_t)}p(S_T)}{\int_0^\infty e^{r(T-t)}\beta\frac{U'(x)}{U'(S_t)}p(x)dx} \\ &= \frac{U'(S_T)p(S_T)}{\int_0^\infty U'(x)p(x)dx}. \end{aligned} \tag{3.11}$$

Note that we integrate the denominator over the support of S_T. The terminal stock price S_T can only take on positive real numbers. Equation (3.11) is the main result of this section, and will be used in the following chapter for an attribution of the risk-neutral skew. The denominator is a constant of integration, ensuring that the density function integrates to one.

We have demonstrated the link between the probability measures involved in the pricing of assets, and the preference structure of an agent. The following is a simple example that uses this relationship to derive the risk-neutral distribution given a real-world distribution and an assumed form for the utility of a representative investor.

Example 3.8. We derive the risk-neutral distribution, given a utility function and real world stock price dynamics, using the relationship described by Equation (3.11).

Assume that the stock price process follows a geometric Brownian motion (GBM) described by the stochastic differential equation (SDE)

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t^{\mathbb{P}}, \tag{3.12}$$

where $W_t^{\mathbb{P}}$ is a Brownian motion under the probability measure $\mathbb{P}$, and the drift and diffusion are given by the scalars μ and σ respectively. Solving this SDE we obtain a stock price process S_T that is a log-normally distributed random variable; under the probability measure $\mathbb{P}$ we have

$$S_T \overset{\mathbb{P}}{\sim} \ln N\left[\left(\mu - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right]. \tag{3.13}$$

Assume there exists a representative investor whose preferences can be described by a power utility function

$$U(S_T) = ak\left(\frac{S_T^{1-c}}{1-c}\right), \tag{3.14}$$

for constants a, c, k, with k defined as

$$k = U'(S_t)\frac{e^{-r(T-t)}}{S_t^{-c}}.$$

Note that S_t denotes the initial stock price, which is a constant, while S_T is the terminal stock price, which is a random variable. Define the power utility parameter c as

$$c = \frac{\mu - r}{\sigma^2}.$$

By differentiating Equation (3.14) with respect to S_T and substituting k we obtain the marginal rate of substitution

$$\frac{U'(S_T)}{U'(S_t)} = ae^{-r(T-t)}\left(\frac{S_T}{S_t}\right)^{-c}.$$

Based on the distribution of S_T under $\mathbb{P}$ given by Equation (3.13), the physical density is of the form

$$p(S_T) = \frac{1}{S_T\sqrt{2\pi\sigma^2(T-t)}}\exp\left[-\frac{\left(\ln\left(\frac{S_T}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(T-t)\right)^2}{2\sigma^2(T-t)}\right], \quad S_T > 0. \tag{3.15}$$

Using Equation (3.10) and Equation (3.15) we are able to derive the risk-neutral distribution;

$$
\begin{aligned}
q(S_T) &= e^{r(T-t)} \frac{U'(S_T)}{U'(S_t)} p(S_T) \\
&= e^{r(T-t)} a e^{-r(T-t)} \left(\frac{S_T}{S_t}\right)^{-c} p(S_T) \\
&= \frac{a}{S_T\sqrt{2\pi\sigma^2(T-t)}} \left(\frac{S_T}{S_t}\right)^{-c} \exp\left[-\frac{\left(\ln\left(\frac{S_T}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(T-t)\right)^2}{2\sigma^2(T-t)}\right] \\
&= \frac{a}{S_T\sqrt{2\pi\sigma^2(T-t)}} \exp\left[-c\ln\left(\frac{S_T}{S_t}\right) - \frac{\left(\ln\left(\frac{S_T}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(T-t)\right)^2}{2\sigma^2(T-t)}\right] \\
&= \frac{a}{S_T\sqrt{2\pi\sigma^2(T-t)}} \exp\left[\frac{r-\mu}{\sigma^2}\ln\left(\frac{S_T}{S_t}\right) - \frac{\left(\ln\left(\frac{S_T}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(T-t)\right)^2}{2\sigma^2(T-t)}\right] \\
&= \frac{a}{S_T\sqrt{2\pi\sigma^2(T-t)}} \exp\left[\frac{2(r-\mu)(T-t)\ln\left(\frac{S_T}{S_t}\right) - \left(\ln\left(\frac{S_T}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(T-t)\right)^2}{2\sigma^2(T-t)}\right] \\
&= \frac{1}{S_T\sqrt{2\pi\sigma^2(T-t)}} \exp\left[-\frac{\left(\ln\left(\frac{S_T}{S_t}\right) - \left(r - \frac{\sigma^2}{2}\right)^2\right)}{2\sigma^2(T-t)}\right].
\end{aligned}
$$

The constant $a = \exp\left[\left(\mu^2 - \mu\sigma^2 - r^2 + r\sigma^2\right)(T-t)^2\right]$ allows us to complete the square, to ensure that this is a valid density function. This is the density of a log-normally distributed random variable, with parameters $\mu = (r - \frac{\sigma^2}{2})(T-t)$ and $\sigma^* = \sigma^2(T-t)$. Therefore, under the risk-neutral measure $\mathbb{Q}$, the distribution of the stock price process is described as

$$
S_T \overset{\mathbb{Q}}{\sim} \ln N\left[\left(r - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right].
$$

The stock price dynamics in the risk-neutral world are thus described by the SDE

$$
\frac{dS_t}{S_t} = rdt + \sigma dW_t^{\mathbb{Q}}. \tag{3.16}
$$

Our risk-neutral measure is consistent with the Black-Scholes model, whereby under the real world measure the stock price process can be described by the GBM given in Equation (3.12) and under the risk-neutral measure stock price process can be described by the GBM given in Equation (3.16). This implies that the utility function implied by the Black-Scholes asset pricing model is the power utility (up to additive and multiplicative constants) of the form

$$
\begin{aligned}
U(S_T) &= \left(\frac{S_T^{1-c}}{1-c}\right) \\
c &= \frac{\mu - r}{\sigma^2}.
\end{aligned}
$$

3.4 Implied Utility Functions

We extend the findings of the previous section to posit a form for the utility function based on given real world and risk-neutral densities. The relationship described in Equation (3.9) allows us to derive a representative utility function from the real world distribution and the risk-neutral distribution. Rearranging Equation (3.9), we obtain

$$U'(S_T) = e^{-r(T-t)}\frac{1}{\beta}U'(S_t)\frac{q(S_T)}{p(S_T)},$$

where, as before, p and q are the real world and risk-neutral density functions respectively. Therefore, the form of the utility function is given by

$$U(S_T) = U(S_t) + \frac{1}{\beta}e^{-r(T-t)}U'(S_t)\int\frac{q(s)}{p(s)}ds,$$

with the integral taken over the range of admissible values of the asset, S_T. We know from Appendix A that the preferences implied by utility functions that are monotone increasing and concave are invariant to linear transformations. Therefore, ignoring additive and multiplicative constants, we have

$$U(S_T) = \int_{-\infty}^{\infty}\frac{q(s)}{p(s)}ds, \tag{3.17}$$

for the general form of a utility function given the probability density functions p and q. This describes the utility function, up to additive and multiplicative constants. Given the two probability densities, real world and risk-neutral, we can now determine a utility function for a representative investor. We apply Equation (3.17) to the asset pricing framework described by the Black-Scholes model.

Example 3.9. Using Equation (3.17) for the form of the utility function, we seek to derive the utility function implied by the Black-Scholes model. The underlying stock price dynamics are described by a GBM,

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t^{\mathbb{P}},$$

where $W_t^{\mathbb{P}}$ is a Brownian motion under the real world probability measure $\mathbb{P}$, and the drift and diffusion are given by the scalars μ and σ respectively. The solution to this SDE results in the following dynamics for the underlying stock price,

$$S_T \overset{\mathbb{P}}{\sim} \ln N\left[\left(\mu - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right],$$

under the real world probability measure $\mathbb{P}$. The resulting probability density function for the underlying asset price with respect to the real world probability measure is then the log-normal probability density;

$$p(x) = \frac{1}{x}\frac{1}{\sqrt{2\pi\tilde{\sigma}}}\exp\left(-\frac{1}{2}\left(\frac{\ln x - \tilde{\mu}}{\tilde{\sigma}}\right)^2\right), \quad x > 0,$$

with $\tilde{\mu} = (\mu + \frac{\sigma^2}{2})(T-t)$ and $\tilde{\sigma} = \sigma\sqrt{(T-t)}$. Given that the Black-Scholes model represents a complete market that exhibits no arbitrage opportunities, there exists a unique equivalent martingale measure, $\mathbb{Q}$, known as the *risk-neutral measure* (Björk; 2009). Under the risk-neutral measure, the underlying stock price dynamics are given by

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t^{\mathbb{Q}},$$

$W_t^{\mathbb{Q}}$ is a now Brownian motion under the risk-neutral probability measure $\mathbb{Q}$. Note the drift of the asset is now the risk-free rate r, while the asset diffusion remains unchanged. The resulting probability density function for the underlying asset price with respect to the risk-neutral probability measure is a log-normal probability density, given by

$$q(x) = \frac{1}{x}\frac{1}{\sqrt{2\pi\tilde{\sigma}}}\exp\left(-\frac{1}{2}\left(\frac{\ln x - \tilde{r}}{\tilde{\sigma}}\right)^2\right), \quad x > 0,$$

with $\tilde{r} = (r + \frac{\sigma^2}{2})(T-t)$ and $\tilde{\sigma} = \sigma\sqrt{(T-t)}$. We now determine the implied utility function. Considering Example 3.8, we expect that the utility function derived in this case will be the power utility. Therefore let us use the coefficient of risk aversion stated in Example 3.8. Define the constants

$$\begin{aligned} c &= \frac{\mu - r}{\sigma^2} \\ b &= e^{-r(T-t)} U'(S_t). \end{aligned}$$

Taking the ratio of the probability distributions, we have

$$\begin{aligned} \frac{q(S_T)}{p(S_T)} &= \left(\frac{S_T}{S_t}\right)^{-c} \exp\left(\frac{(\mu - r)(\mu - r - \sigma^2)(T-t)}{2\sigma^2}\right) \\ &= a\left(\frac{S_T}{S_t}\right)^{-c}, \end{aligned}$$

where $a = \exp\left(\frac{(\mu-r)(\mu-r-\sigma^2)(T-t)}{2\sigma^2}\right)$. Substitution into Equation (3.17) yields

$$\begin{aligned} U(S_T) &= \int_{-\infty}^{\infty} \left(\frac{s}{S_t}\right)^{-c} ds \\ &= \frac{1}{S_t^{-c}} \int_{-\infty}^{\infty} s^{-c} ds. \end{aligned}$$

Therefore, ignoring additive and multiplicative constants, we have the following forms for the utility function implied by the Black-Scholes model

$$U(S_T) = \frac{S_T^{1-c}}{1-c}, \quad c \neq 1 \tag{3.18}$$

$$U(S_T) = \ln(S_T), \quad c = 1. \tag{3.19}$$

This excludes the trivial case where $c = 0$, which would result in a constant utility $U(S_T) = yS_T$, for some constant y. We overlook this case as it is only possible if $\mu = r$, which would imply that the risky assets yield the risk-free rate, implying that pricing in the market is risk-neutral. Therefore it safe to assume that the risk-free rate r is not equal to the expected return on the risky asset μ.

Equation (3.18) represents a power utility function, and Equation (3.19) represents a log utility function. These both exhibit a constant relative risk aversion (CRRA) coefficient γ; coefficients of risk aversion are dealt with in Section A.4. Hence we conclude that the Black-Scholes model implies an investor with constant relative risk aversion, and a utility function that is either power utility or log utility, monotonically increasing and concave in both cases. The constant CRRA in this case, given by

$$c = \frac{\mu - r}{\sigma^2}, \tag{3.20}$$

looks similar to the Sharpe ratio; noting that in this case the excess return is standardised by the variance rather than the volatility. This implies that this model has a higher relative risk aversion than the market price of risk, given by the Sharpe ratio $\frac{\mu - r}{\sigma}$.

The next section provides a brief explanation of the empirical finding known as the *pricing kernel puzzle*.

3.5 Pricing Kernel Puzzle

Equation (3.5) defines that the stochastic discount factor, also known as the *pricing kernel*, we restate it here as a function of the terminal asset price S_T,

$$M_{t,T}(S_T) = \beta \frac{U'(S_T)}{U'(S_t)}. \tag{3.21}$$

We have shown that this is equal to the ratio of the risk-neutral density to the real world density discounted at the risk free rate,

$$M_{t,T}(S_T) = e^{-r(T-t)} \frac{q(S_T)}{p(S_T)}. \tag{3.22}$$

As we have shown above, this pricing kernel allows us to calculate the price of an option as the expected value of the the pricing kernel multiplied by the payoff of the option.

We now prove that under the assumptions of a risk averse representative investor, the pricing kernel is a decreasing function of the underlying asset price. Theorem A.10 states that an investor with utility described by the utility function $U(S)$ is risk averse if and only if the utility function $U(S)$ is concave; that is, the second derivative of the utility function is less than or equal to zero for all values of S,

$$U''(S) \leq 0, \quad \text{for all } S. \tag{3.23}$$

We assume that the representative investor is risk averse. The gradient of the pricing kernel can be described by looking at the first derivative of the pricing kernel function. From Equation (3.21) we have

$$M'_{t,T}(S_T) = \beta \frac{U''(S_T)}{U'(S_t)},$$

which is negative for all S_T, given the result in Equation (3.23), and remembering that a non-satiated investor has a positive first derivative for all S_T (see Assumption A.5). Given a positive β, this formulation indicates that $M'_{t,T}(S_T)$ is negative for all S_T, implying that the pricing kernel is a decreasing function of the terminal stock price. Given the assumption of non-satiated, risk averse investors, we would expect to observe decreasing empirical pricing kernels.

Jackwerth (2004), Detlefsen et al. (2010), Barone-Adesi, L. and Shefrin (2012) and Rosenberg and Engle (2002) provide empirical evidence to the contrary, showing that the pricing kernel is increasing for certain portions of the underlying price. This is contrary to expectations, and this finding is known as the *pricing kernel puzzle*.

The pricing kernel puzzle is the term given to the occurrence where empirical pricing kernels exhibit an upward sloping portion (Barone-Adesi, L. and Shefrin; 2012). Given that our theory indicates that the pricing kernel should be downward sloping, the occurrence of an upward sloping pricing kernel is the subject of much study. Evidence from the S&P500 implies that investors are risk-seeking for certain levels of the index, exhibiting an increasing pricing kernel in these portions, contrary to theory (Jackwerth; 2004).

Equation (3.22) allows us to estimate an empirical pricing kernel; the ratio of the empirically derived risk-neutral and real world densities, discounted at the risk free rate. Barone-Adesi, Dallo and Vovchak (2012) demonstrate a method for estimating the risk-neutral and real world densities from historical observations, ensuring that the risk-neutral distribution satisfies the no-arbitrage condition.

We find that empirical pricing kernels derived from market data before the crash of 1987 fit the expectation of a decreasing function of wealth, while after the crash this does not hold (Jackwerth; 2000). The most probable cause of the increasing portion in the pricing kernel is mispricing in the market (Jackwerth; 2000).

If our market assumptions hold, then we should expect to see a pricing kernel that is decreasing in aggregate consumption, using the market price as a proxy for aggregate consumption (Hens and Reichlin; 2012). This means that we would expect a decreasing pricing kernel in a complete market, where investors exhibit risk aversion and do not exhibit biased beliefs (Hens and Reichlin; 2012). These assumptions do not hold in practice, and as a result we cannot expect the pricing kernel to be a monotone decreasing function. "Incomplete markets, risk-seeking behaviour and incorrect beliefs can induce increasing parts in the pricing kernel and can be seen as potential solutions for the pricing kernel puzzle" (Hens and Reichlin; 2012).

Hens and Reichlin (2012) perform an empirical study in order to test these three possible explanations, using the results from the tests to gauge the validity of each of the possible solutions for the pricing kernel puzzle. Their study shows that risk-seeking behaviour is not a robust explanation. The explanation of biased beliefs is consistent with the empirical results and is robust under various aggregations. Incomplete markets allow for a pricing kernel that is increasing in certain parts, however in order for all pricing kernels to have increasing portions, we require extreme assumptions for the distribution of wealth. Barone-Adesi, L. and Shefrin (2012) show that increasing portions of the pricing kernel relate strongly to overconfidence on the part of market participants. Detlefsen et al. (2010) study empirical pricing kernels based on the DAX in 2000, 2002, and 2004. Their findings show that risk-seeking behaviour of investors may explain increasing in the pricing kernel observed in the market.

Rosenberg and Engle (2002) estimate a time-varying pricing kernel, and find evidence of time-varying risk aversion of market participants, leading to a pricing kernel that is not decreasing in wealth. Their study shows further that risk aversion in the market is strongly linked to the market cycle; "the level of risk aversion is positively correlated with indicators of recession and negatively correlated with indicators of expansion" (Rosenberg and Engle; 2002).

For a deeper discussion relating to the pricing kernel puzzle refer to Detlefsen et al. (2010), Rosenberg and Engle (2002), and Jackwerth (2004).

The following section describes the method for obtaining an empirical estimate of the risk-neutral distribution.

3.6 An Empirical Risk-neutral Distribution

We now describe the development of a method for obtaining a risk-neutral distribution of returns from a real world distribution of returns, based on information theory, where the real world distribution is taken as the historical time-series of returns. The method to follow is adapted from Stutzer (1996) and Derman and Zou (1999).

The method for obtaining a risk-neutral distribution from an empirical distribution employs results from information theory; most notably, entropy and relative entropy. The relevant theory will be explained, with references provided to relevant texts.

A summary of the process for obtaining a risk-neutral distribution from an empirical distribution follows. Firstly, we require an estimate of the real world distribution of returns. This is obtained by observing the returns on a chosen underlying stock or index, for a relevant period in history. This time-series of returns constitutes the empirical real world distribution of returns of the chosen stock or index. The empirical real world distribution is then used to determine the risk-neutral distribution by minimizing the relative entropy between the two distributions, subject to a constraint on the risk-neutral distribution. It is required that the risk-neutral distribution be consistent with the forward price of the underlying asset.

The following sections describe, respectively, the required results from information theory and the method for obtaining a risk-neutral distribution from a historical real world distribution.

Information and Entropy

The amount of information gained from the occurrence of an event should be related to the probability of the occurrence of that event (Derman and Zou; 1999). We quantify the information gained in the occurrence of a specific event by an information function. We define a function $I(p)$, that represents the information gained by the observance of the event $X = x$, where X is a random variable, x is a specific event, and p is the probability of the event x occurring. It is required that the function $I(p)$ is non-negative and decreasing in p (Derman and Zou; 1999). The function is non-negative because the occurrence of the event must provide some information. The function is decreasing because more likely the occurrence of the event is, the less information its occurrence provides. Derman and Zou (1999) show that the function that provides a measure of information is given by

$$I(p) = -\ln(p).$$

This function is used to measure entropy in a distribution.

We relate the concept of entropy to that of probability; probability provides a measure of the uncertainty of the occurrence of a specific event, while entropy provides a measure of the uncertainty in a distribution (Derman and Zou; 1999).

The entropy of a random variable X, with the probability of event x_i given by p_i, is defined as the expected value of the information given by the occurrence of an event in the distribution,

$$H(X) = -\sum_{i}^{n}(p_i \ln(p_i))$$

The probability p_i is less than or equal to one, for all i, therefore entropy is always positive. A large entropy measure implies a broad distribution, with a wide spread of probable events, while a small entropy measure implies a narrow distribution, where little information can be gained from the occurrence of an expected event (Derman and Zou; 1999).

Therefore H represents the uncertainty in the distribution; large expected information corresponds to high uncertainty, and small expected information corresponds to low uncertainty.

Definition 3.10. *Entropy* is the mathematical function that measures the uncertainty of a distribution. This is consistent with maximum entropy corresponding to maximum uncertainty. The lowest entropy of a distribution occurs when one event X_i occurs with certainty, $p_i = 1$. It can be shown that the highest entropy of a distribution occurs when $p_i = \frac{1}{n}$ for all events x_i, then we have $H = \ln(n)$, corresponding to maximal uncertainty.

We now define relative entropy, which quantifies "the information gained upon changing the distribution as a result of new information" (Derman and Zou; 1999). If we have a prior distribution P and a posterior distribution Q based on new information, relative entropy tells us the reduction in uncertainty as a result of the new information. Relative entropy between two distributions P and Q is defined as

$$\begin{aligned} S(P,Q) &= \mathbb{E}_Q\left[\ln Q - \ln P\right] \\ &= \sum_i \left(q_i \ln \frac{q_i}{p_i}\right) \\ &= -\sum_i \left(q_i \ln \frac{p_i}{q_i}\right). \end{aligned}$$

By Jensen's inequality, we have the result that the average of $-\ln \frac{p_i}{q_i}$ is greater than the log of the average of $\frac{p_i}{q_i}$. We then have

$$\begin{aligned} S(P,Q) &> -\ln \sum_i \left(q_i \frac{p_i}{q_i} \right) \\ &= -\ln \sum_i (p_i) \\ &= -\ln 1 \\ &= 0. \end{aligned}$$

Therefore we conclude that the relative entropy function $S(P,Q)$ is strictly non-negative, and is zero if and only if the distributions P and Q are identical. We can think of $S(P,Q)$ as the distance between the two distributions. We use the measure of relative entropy in the derivation of the risk-neutral distribution.

Risk-neutralised Historical Distribution

Consider a stock with a spot price S_0, with the objective, real world distribution of future returns described by the distribution function $P(S_T)$. We seek to derive a risk-neutral distribution of returns on this asset, $Q(S_T)$, from the real world distribution of returns $P(S_T)$, subject to certain constraints. We have the constraint that the expectation of the stock price under the risk-neutral distribution must be equal to the stock forward price, in order for the no-arbitrage principal to hold; that is

$$\mathbb{E}_{\mathbb{Q}}\left(S_T\right) = \int q(S) S dS = S_0 e^{rT}. \tag{3.24}$$

where r is the market risk-free rate. We also have the constraint that the risk-neutral distribution Q should integrate to one, in order for Q to be a valid probability distribution;

$$\int q(S) dS = 1. \tag{3.25}$$

The risk-neutral distribution is obtained by by minimizing the relative entropy between the real world distribution P and a risk-neutral distribution Q, subject to the constraints specified above (Derman and Zou; 1999). Therefore, to obtain the risk-neutral distribution $Q(S_T)$, we have

$$\min \left(S(P,Q) = \mathbb{E}_Q \left[\frac{Q(S)}{P(S)} \right] \right)$$

subject to the constraints specified in Equation (3.24) and Equation (3.25).

The form for the risk-neutral distribution, given these equations, is given by

$$Q(S_T) = \frac{P(S_T)}{\int P(S) e^{-\lambda S} dS} e^{-\lambda S_T} \tag{3.26}$$

where the constant λ is obtained by ensuring that the risk-neutral distribution satisfies the forward condition given by Equation (3.24). The risk-neutral distribution $Q(S_T)$ will be positive for all S_T because the real world distribution $P(S_T)$ is positive for all S_T.

Given a real world distribution obtained from historical market data, using a time series, we can obtain a risk-neutral distribution in the manner described above. This can then be used to analyse risk aversion and utility present in the market using the stochastic discount factor described earlier in this chapter. The risk-neutral distribution obtained in this manner is referred to as the *risk-neutralised historical distribution* (Derman and Zou; 1999).

The focus of the next chapters will be the derivation of the risk-neutral density, given a utility function and a real world density.

Chapter 4

Skew in the Risk-Neutral Measure: Power Utility Function

It is widely known that market risk-neutral measures are not symmetrical, and exhibit negative skewness (Bakshi et al.; 2003). In this chapter we provide motivation for the sources of skewness observed in the market implied risk-neutral measure. Using the framework provided in Chapter 3, utility driven changes of measure allow us to show that risk aversion is a source of skewness in the risk-neutral measure. We demonstrate other sources of skewness in the risk-neutral measure; namely the variance, skewness and kurtosis of the corresponding physical measure.

Such a characterisation of risk-neutral skewness is provided in Bakshi et al. (2003), for a power utility change of measure. We reproduce their argument here in greater detail. The subsequent chapter, following the work of Bakshi et al. (2003), extends this to a broader class of utility functions.

We begin by specifying the change of measure equation for preferences implied by a power utility function. This change of measure is then specified in terms of the return on an asset rather than the asset price. The final section of this chapter presents a theorem by which we can attribute the risk-neutral skewness to risk aversion and moments of the physical distribution.

4.1 The Change of Measure Equation

We present an example of a change of measure in which the distributional form of either the physical or the risk-neutral distribution is left unspecified, where the change of measure is based on the power utility function. This example shows explicitly how risk aversion can be incorporated into the risk-neutral distribution.

Recall the power utility function, based on the stock price S_T,

$$U(S_T) = \frac{S_T^{1-\gamma}}{1-\gamma} \quad \gamma > 0.$$

The marginal utility function is then given as

$$\begin{aligned} U'(S_T) &= S_T^{-\gamma} \\ &= e^{-\gamma \ln(S_T)}. \end{aligned}$$

From Equation (3.11) we determine the risk-neutral density q as

$$q(S_T) = \frac{U'(S_T)p(S_T)}{\int_0^\infty U'(x)p(x)dx} = \frac{S_T^{-\gamma}p(S_T)}{\int_0^\infty x^{-\gamma}p(x)dx}, \tag{4.1}$$

where p is the physical density, and the integral in the denominator is taken over the admissible values of the terminal stock price S_T, $S_T \in \mathbb{R}^+$. This equation allows us to obtain the risk-neutral density q of the terminal price of the underlying stock S_T, from the physical distribution of S_T.

4.2 Probability of Returns

We now seek to obtain the risk-neutral density of the *return* on the underlying stock over the period, from the physical density of the return.

Definition 4.1. The *return* of the underlying stock price over the period (t, T) is defined as the logarithmic return on the underlying stock price S_t between times t and T,

$$R_T = \ln\left(\frac{S_T}{S_t}\right). \tag{4.2}$$

If we fix the time period under consideration, notice that the initial stock price is a constant. We can therefore simplify the notation. Let the initial stock price be given by the constant k, and omit the subscripts from the terminal stock price and the return variables. We rewrite the return on the stock as

$$R = \ln\left(\frac{S}{k}\right).$$

We then have $S = ke^R$.

The aim of this section is to determine the change of measure equation in terms of the return of the underlying asset, R, rather than the asset price S. Let q_S be the risk-neutral density and p_S be the physical density function of the underlying stock price at a specific date. Similarly, let q_R and p_R be the risk-neutral and physical density functions, respectively, of the return of the underlying stock price over a set period of time, corresponding to the date of the stock S and the initial stock price k. Note, we omit the T subscript for notational simplicity. We have, from Equation (4.1),

$$q_S(S) = \frac{S^{-\gamma}p_S(S)}{\int_0^\infty x^{-\gamma}p_S(x)dx}. \tag{4.3}$$

The terminal stock price is a random variable. The return on this stock over the period considered is a transformation of this random variable. From the density function of S we obtain the density function of R. The probability of the return R being less than some constant r is

$$\begin{aligned}\mathbb{P}(R \leq r) &= \mathbb{P}\left(\ln\left(\frac{S}{k}\right) \leq r\right) \\ &= \mathbb{P}\left(S \leq ke^r\right) \\ &= \int_{-\infty}^{ke^r} p_S(x)dx,\end{aligned}$$

where $\mathbb{P}(A)$ denotes the probability of some event A. The probability density function of the return is calculated as the derivative of the probability distribution function, evaluated at the limits of the probability integral. The density of R under the physical measure is then

$$p_R(R) = p_S\left(ke^R\right) \times ke^R. \tag{4.4}$$

By similar argument, we have for the density of R under the risk-neutral measure

$$q_R(R) = q_S\left(ke^R\right) \times ke^R. \tag{4.5}$$

We now determine the risk-neutral density of return $q(R)$, from Equation (4.5), subsituting in Equation (4.3) for the risk-neutral density of stock price, as follows

$$\begin{aligned} q_R(R) &= q_S\left(ke^R\right) \times ke^R \\ &= \frac{(ke^R)^{-\gamma} p_S(ke^R)}{\int_0^\infty x^{-\gamma} p_S(x)dx} \times ke^R \\ &= \frac{k^{-\gamma} e^{-\gamma R} p_R(R)}{\int_0^\infty x^{-\gamma} p_S(x)dx}. \end{aligned}$$

The final equality follows from Equation (4.4). Note that the denominator is still specified with regard to the physical density of stock prices, p_S. In order to obtain the denominator in terms of the density of returns, we make the following substitution to the variable of integration. Let y be defined as

$$y = \ln\left(\frac{x}{k}\right),$$

we then have

$$x = ke^y \quad \Rightarrow \quad dx = ke^y dy.$$

Substitution in the integral in the denominator, and changing the limits of integration accordingly, we obtain

$$\begin{aligned} q_R(R) &= \frac{k^{-\gamma} e^{-\gamma R} p_R(R)}{\int_{-\infty}^\infty (ke^y)^{-\gamma} p_S(ke^y) ke^y dy} \\ &= \frac{k^{-\gamma} e^{-\gamma R} p_R(R)}{\int_{-\infty}^\infty k^{-\gamma} e^{-\gamma y} p_R(y)dy} \\ &= \frac{e^{-\gamma R} p_R(R)}{\int_{-\infty}^\infty e^{-\gamma y} p_R(y)dy}. \end{aligned}$$

Therefore, we have the following equation for the risk-neutral density of returns given a power utility function,

$$q_R(R) = \frac{(e^{-\gamma R}) p_R(R)}{\int_{-\infty}^\infty (e^{-\gamma y}) p_R(y)dy}. \tag{4.6}$$

Equation (4.6) shows, in the form of density functions, the formula for a change of measure; how we derive the risk-neutral density from the physical density. The physical density is *exponentially tilted* to obtain the risk-neutral density. Essentially, at each point, an exponential weighting is applied to the physical density. This weighting is a function of risk aversion (described by the CRRA coefficient γ) and the return R. This relationship shows how risk aversion determines the degree to which the risk-neutral distribution is skewed. The denominator ensures that q_R integrates to one, so that Equation (4.6) is a valid density function.

From this point on, we will work with utility of returns, and the probability densities of returns; therefore we will omit the subscript R with regard to the density functions for the remainder of this chapter. We rewrite Equation (4.6) as

$$q(R) = \frac{(e^{-\gamma R})p(R)}{\int_{-\infty}^{\infty}(e^{-\gamma y})p(y)dy}. \tag{4.7}$$

The section to follow uses Equation (4.7) in order to relate skew in the risk-neutral distribution to the risk-aversion coefficient γ.

4.3 Risk-Neutral Skewness

This section demonstrates the attribution of skewness in the risk-neutral distribution to the risk aversion of a representative agent, as well as to the higher order moments of the corresponding physical distribution. Risk-neutral skewness is defined as the third centered moment of the risk-neutral distribution. We will denote the n^{th} centered moment of the risk-neutral distribution as m_n, and the n^{th} centered moment of the corresponding physical distribution as $\bar{m}_n$. Thus m_3 and m_4 represent, respectively, the skew and kurtosis of the risk-neutral distribution, and $\bar{m}_3$ and $\bar{m}_4$ the skew and kurtosis of the physical distribution. The following theorem is adapted from Bakshi et al. (2003).

Theorem 4.2. *The skewness present in the risk-neutral distribution can be described, up to first order of γ, as a function of risk aversion and the higher order moments of the corresponding physical distribution;*

$$m_3(t,T) \approx \bar{m}_3(t,T) - \gamma\left(\bar{m}_4(t,T) - 3\right)\bar{\sigma}(t,T), \tag{4.8}$$

where $\bar{\sigma}$ is the standard deviation of the physical measure. The power utility change of measure leads to negative skewness in the risk-neutral measure, given a sufficiently fat-tailed physical distribution.

Proof. We explore the link between the skewness in the risk-neutral distribution $q(R)$ and that of the physical distribution $p(R)$. For ease of notation, we omit the time parameters. Without loss of generality, assume that the physical distribution $p(R)$ has a mean of zero (the first moment $\bar{m}_1$ is equal to zero). We define the first 3 uncentered higher order moments of the physical distribution as

$$\bar{\kappa}_2 = \int_{-\infty}^{\infty} R^2 p(R)dR$$
$$\bar{\kappa}_3 = \int_{-\infty}^{\infty} R^3 p(R)dR$$
$$\bar{\kappa}_4 = \int_{-\infty}^{\infty} R^4 p(R)dR.$$

We define the moment generating function of $p(R)$, up to order λ^4, for all λ, as

$$\begin{aligned}\bar{\mathcal{M}}[\lambda] &= \int_{-\infty}^{\infty} e^{\lambda R} p(R)dR \\ &= 1 + \frac{\lambda^2}{2}\bar{\kappa}_2 + \frac{\lambda^3}{6}\bar{\kappa}_3 + \frac{\lambda^4}{24}\bar{\kappa}_4 + o(\lambda^4).\end{aligned}$$

From Equation (4.7) we have for $\mathcal{M}[\lambda]$, the moment generating function of the risk-neutral distribution $q(R)$

$$\begin{aligned}\mathcal{M}[\lambda] &= \int_{-\infty}^{\infty} e^{\lambda R} q(R) dR \\ &= \frac{\int_{-\infty}^{\infty} e^{\lambda R} e^{-\gamma R} p(R) dR}{\int_{-\infty}^{\infty} e^{-\gamma R} p(R) dR} \\ &= \frac{\bar{\mathcal{M}}[\lambda - \gamma]}{\bar{\mathcal{M}}[-\gamma]}.\end{aligned}$$

Therefore $\mathcal{M}[\lambda]$ can be determined from the moment generating function of the physical distribution $p(R)$. We expand this form for the moment generating function, using the Taylor expansion of e,

$$\begin{aligned}\mathcal{M}[\lambda] &= \frac{\bar{\mathcal{M}}[\lambda - \gamma]}{\bar{\mathcal{M}}[-\gamma]} \\ &= \frac{\int_{-\infty}^{\infty} e^{R(\lambda-\gamma)} p(R) dR}{\int_{-\infty}^{\infty} e^{-\gamma R} p(R) dR} \\ &= \frac{\int_{-\infty}^{\infty} \left(1 + R(\lambda - \gamma) + \frac{R^2(\lambda-\gamma)^2}{2} + \frac{R^3(\lambda-\gamma)^3}{6} + \frac{R^4(\lambda-\gamma)^4}{24} + \ldots\right) p(R) dR}{\int_{-\infty}^{\infty} \left(1 - \gamma R + \frac{\gamma^2 R^2}{2} - \frac{\gamma^3 R^3}{6} + \frac{\gamma^4 R^4}{24} + \ldots\right) p(R) dR} \\ &= \frac{1 + (\lambda - \gamma)\bar{\kappa}_1 + \frac{(\lambda-\gamma)^2}{2}\bar{\kappa}_2 \frac{(\lambda-\gamma)^3}{6}\bar{\kappa}_3 + \frac{(\lambda-\gamma)^4}{24}\bar{\kappa}_4 + \ldots}{1 - \gamma\bar{\kappa}_1 + \frac{\gamma^2}{2}\bar{\kappa}_2 - \frac{\gamma^3}{6}\bar{\kappa}_3 + \frac{\gamma^4}{24}\bar{\kappa}_4 + \ldots}.\end{aligned}$$

Recall the formula for obtaining the n^{th} uncentered moment of a distribution from the moment generating function,

$$\kappa_n = \left.\frac{\partial^n \mathcal{M}[\lambda]}{\partial \lambda^n}\right|_{\lambda=0}. \tag{4.9}$$

We now calculate the uncentered moments of $q(R)$, up to the first order of γ. Note that up to first order of γ, the denominator becomes $1 + o(\gamma)$. From Equation (4.9) we have

$$\begin{aligned}\kappa_1 &= \left.\frac{\partial \mathcal{M}[\lambda]}{\partial \lambda}\right|_{\lambda=0} \\ &= \left.\frac{\bar{\kappa}_1 + (\lambda - \gamma)\bar{\kappa}_2 + \frac{(\lambda-\gamma)^2}{2}\bar{\kappa}_3 \frac{(\lambda-\gamma)^3}{6}\bar{\kappa}_4 + \ldots}{1 + o(\gamma)}\right|_{\lambda=0} \\ &= \frac{\bar{\kappa}_1 - \gamma\bar{\kappa}_2 + \frac{\gamma^2}{2}\bar{\kappa}_3 - \frac{\gamma^3}{6}\bar{\kappa}_4 + \ldots}{1 + o(\gamma)} \\ &= \frac{\bar{\kappa}_1 - \gamma\bar{\kappa}_2 + o(\gamma)}{1 + o(\gamma)} \\ &= \frac{\bar{\kappa}_1 - \gamma\bar{\kappa}_2}{1} + o(\gamma).\end{aligned}$$

The final equality follows from the properties of little-o notation. We then have the formula for the first uncentered moment of $q(R)$, up to the first order of γ, as

$$\kappa_1 \approx \bar{\kappa_1} - \gamma\bar{\kappa_2}.$$

By a similar argument, we have the second and third moments as

$$\kappa_2 \approx \bar{\kappa}_2 - \gamma\bar{\kappa}_3$$
$$\kappa_3 \approx \bar{\kappa}_3 - \gamma\bar{\kappa}_4.$$

This shows how the risk-aversion coefficient, together with a change of measure based on a power utility, alter the first three moments of the risk-neutral distribution. Therefore the centered moments of the physical measure are as follows,

$$m_2 = \frac{\bar{\kappa}_2}{\bar{\kappa}_2^2} = \sigma^2$$
$$m_3 = \frac{\bar{\kappa}_3}{\bar{\kappa}_2^3}$$
$$m_4 = \frac{\bar{\kappa}_4}{\bar{\kappa}_2^4}.$$

We now calculate the coefficient of skew for the risk-neutral measure:

$$\begin{aligned} m_3 &= \frac{\int_{-\infty}^{\infty}(R-\kappa_1)^3 q(R)dR}{\left(\int_{-\infty}^{\infty}(R-\kappa_1)^2 q(R)dR\right)^{\frac{3}{2}}} \\ &= \frac{\int_{-\infty}^{\infty}(R^3 - 3R^2\kappa_1 + 3R\kappa_1^2 - \kappa_1^3)q(R)dR}{\left(\int_{-\infty}^{\infty}(R^2 - 2R\kappa_1 + \kappa_1^2)q(R)dR\right)^{\frac{3}{2}}} \\ &= \frac{\kappa_3 - 3\kappa_1\kappa_2 + 2\kappa_1^3}{(\kappa_2 - \kappa_1^2)^{\frac{3}{2}}} \\ &\approx \frac{\bar{\kappa}_3 - \gamma\bar{\kappa}_4 - 3(\bar{\kappa}_1 - \gamma\bar{\kappa}_2)(\bar{\kappa}_2 - \gamma\bar{\kappa}_3) + 2(\bar{\kappa}_1 - \gamma\bar{\kappa}_2)^3}{(\bar{\kappa}_2 - \gamma\bar{\kappa}_3 - (\bar{\kappa}_1 - \gamma\bar{\kappa}_2)^2)^{\frac{3}{2}}} \\ &= \frac{\bar{\kappa}_3 - \gamma(\bar{\kappa}_4 - 3\bar{\kappa}_2^2)}{\bar{\kappa}_2^{\frac{3}{2}}} \\ &= \frac{\bar{\kappa}_3}{\bar{\kappa}_2^{\frac{3}{2}}} - \gamma\frac{\bar{\kappa}_4}{\bar{\kappa}_2^{\frac{3}{2}}} + \gamma\frac{\bar{\kappa}_2^2}{\bar{\kappa}_2^{\frac{3}{2}}} \\ &= \bar{m}_3 - \gamma(\bar{m}_4 - 3)\bar{\sigma}, \end{aligned}$$

where $\bar{\sigma}$ is the standard deviation of the physical measure. □

This theorem provides mathematical and economic motivation to the existence of the observed risk-neutral skew. From Equation (4.8), we see that even under symmetrical physical measure, skewness is still possible in the risk-neutral measure. The negative skewness observed in the risk-neutral measure can now be attributed to three sources: firstly, the negative skewness in the corresponding physical measure; secondly, the excess kurtosis of the corresponding physical measure; and thirdly, the risk aversion implied by a utility function.

Skew in the physical density leads to skew in the corresponding risk-neutral density even in the case of a risk-neutral investor; that is, an investor with a coefficient of risk aversion equal to zero. In the presence of a non-zero coefficient of risk aversion, it is only *excess* kurtosis that contributes to increasing the magnitude of the skew in the risk-neutral density. Excess kurtosis increases the length of the left tail of the risk-neutral density. In this framework, a physical density with a kurtosis equal to three nullifies the effect of the standard deviation of returns and, more importantly, the coefficient of risk aversion on the skew in the risk-neutral density. The magnifying effect of risk aversion (and volatility of returns) on the risk-neutral skew is diminished for levels of physical kurtosis less than three. Note that in the presence of non-zero risk aversion and excess kurtosis, the standard deviation of the physical measure (volatility of returns) contributes to the level of skew in the risk-neutral density. Greater levels of volatility in the physical measure will not correspond to negative skewness in the risk-neutral density, unless the representative agent is risk averse, and the physical density is sufficiently fat-tailed.

The following example illustrates how a physical density that is normally distributed (with no excess kurtosis) induces no skew in the risk-neutral measure.

Example 4.3. Assume that the physical density of the underlying stock return is normally distributed with mean and variance given by μ and σ^2 respectively. Equation (4.7) applied to this normally distributed density p results in the following form for the risk-neutral density,

$$
\begin{aligned}
q(R) &= A \exp\left(-\gamma R\right) \exp\left(-\frac{(R-\mu)^2}{2\sigma^2}\right) \\
&= A^* \exp\left(-\frac{\left(R-(\mu-\gamma\sigma^2)\right)^2}{2\sigma^2}\right),
\end{aligned}
$$

for positive constants A and A^*. This is the density of a mean shifted normally distributed variable. This density exhibits no skew. It follows that in the case of exponential tilting of the physical density, excess kurtosis is required to induce skew in the risk-neutral density.

Note, we have the risk-neutral density that is normally distributed, where the density depends on a risk aversion parameter γ. Given that the assumptions made conform to the Black-Scholes model, we would assume that the risk aversion parameter would be defined as that satisfying the unique preference structure of the Black-Scholes model (see Example 3.9). We substitute the constant CRRA from Example 3.9, specified in Equation (3.20). The result confirms that the risk-neutral density of returns is that which is implied in the Black-Scholes model. The CRRA derived in Example 3.9 is given as

$$
c = \frac{\mu - r}{\sigma^2},
$$

for r defined as the risk-free rate, and μ and σ defined as in this example. If we let $\gamma = c$ we have

$$
\begin{aligned}
q(R) &= A^* \exp\left(-\frac{\left(R-\left(\mu-c\sigma^2\right)\right)^2}{2\sigma^2}\right) \\
&= A^* \exp\left(-\frac{\left(R-\left(\mu-\frac{\mu-r}{\sigma^2}\sigma^2\right)\right)^2}{2\sigma^2}\right) \\
&= A^* \exp\left(-\frac{(R-r)^2}{2\sigma^2}\right),
\end{aligned}
$$

where A^* is the constant of integration. The risk-neutral density is normally distributed with mean and variance given by r and σ respectively.

The level of risk aversion increases the level of negative skew in the risk-neutral density only in the presence of excess kurtosis in the physical density. Physical densities estimated from the market are commonly symmetric and fat-tailed (Bakshi et al.; 2003). Theorem 4.2 then implies that the most probable causes of negative skew in the risk-neutral are fat-tailed physical distributions and risk aversion in the representative investor. It is commonplace that the physical densities of index returns estimated in practice exhibit excess kurtosis (Bakshi et al.; 2003).

Such an analysis is of limited use if the attribution works only under the assumption of preferences defined by the power utility. The following chapter determines whether such an attribution, based on a utility driven change of measure, can be generalised to other utility functions.

Chapter 5

Skew in the Risk-Neutral Measure: General Utility Function

Theorem 4.2 is contingent on a power utility driven change of measure. While the assumption of power utility may result in tractable equations, it limits the framework, and by extension, the robustness of the theory. We now seek to generalise this to include other forms for the utility function that drive the change of measure. The aim is to describe a change of measure for a broader class of utility functions, and then to derive a theorem such as Theorem 4.2, in order to attribute skew in the risk-neutral measure to the risk aversion of a representative agent. The methodology to follow, as well as the general utility function posited are based on the work of Bakshi et al. (2003).

5.1 The Change of Measure Equation

We begin by describing the methodology to change probability measure based on a general utility function. Firstly, we posit a general utility function. The intention is to describe a method for changing measure based on this general function.

Consider the general utility function described by its marginal utility as

$$U'(R) = \int_0^\infty e^{-z\phi R}\nu(z)dz, \tag{5.1}$$

for a scalar parameter ϕ, and a probability measure ν defined on the set of positive real numbers. This general marginal utility function nests a wide range of utility functions.

Firstly, note that for all continuous positive probability measures ν, the utility function implied by the marginal utility in Equation (5.1) is increasing and concave. We confirm this by examining the first and second derivatives of the utility function specified by this marginal utility. The first derivative of the implied utility function is given by Equation (5.1). This is positive for all $\phi, R \in \mathbb{R}$, and ν a positive probability measure. The second derivative is given as

$$\begin{aligned} U''(R) &= \frac{d}{dR}\left(\int_0^\infty e^{-z\phi R}\nu(z)dz\right) \\ &= \int_0^\infty \frac{\partial}{\partial R}\left(e^{-z\phi R}\right)\nu(z)dz \\ &= \int_0^\infty -z\phi e^{-z\phi R}\nu(z)dz \\ &= -\phi\int_0^\infty z e^{-z\phi R}\nu(z)dz, \end{aligned}$$

which is negative for all $\phi, R \in \mathbb{R}$, and ν a positive probability measure. Due to the fundamental theorem of calculus (second form), we can interchange the integration and differentiation operators, since the integrand is continuous in both R and z. Thus the utility function $U(R)$ implied by Equation (5.1) is a monotone increasing utility function. This implies a representative investor that exhibits non-satiation and risk aversion, (see Remark (A.12)).

We now demonstrate a powerful result for this general utility function. We find that all utility functions belonging to the hyperbolic absolute risk aversion (HARA) family of utility functions are nested in Equation (5.1), if we take the probability measure ν to have a density characterised by the gamma density function. The HARA family of utility functions is described in Section A.5 in the Utility Theory appendix. Recall the gamma density function

$$\nu(z) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} \qquad \text{for } z \geq 0, \tag{5.2}$$

where $\alpha, \beta > 0$ are constants, and $\Gamma(x) = (x-1)!$ is known as the *gamma function*. We demonstrate that the choice of the gamma density for the measure ν in Equation (5.1) results in HARA utility. In order to show this, we need to demonstrate that for ν given by the gamma density function, the absolute risk aversion $A(R) = -\frac{U''(R)}{U'(R)}$ is a hyperbolic function of R, of the form

$$A(w) = \frac{1}{\frac{w}{1-\gamma} + \frac{\eta}{\beta}}, \tag{5.3}$$

given in Equation (A.9). Let ν be specified by the gamma density given in Equation (5.2). Equation (5.1) then becomes

$$\begin{aligned}
U'(R) &= \int_0^\infty e^{-z\phi R} \nu(z) dz \\
&= \int_0^\infty e^{-z\phi R} \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} dz \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty z^{\alpha-1} e^{-z(\phi R + \beta)} dz \\
&= \frac{\beta^\alpha}{(\phi R + \beta)^\alpha} \int_0^\infty \frac{(\phi R + \beta)^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-z(\phi R + \beta)} dz \\
&= \frac{\beta^\alpha}{(\phi R + \beta)^\alpha}.
\end{aligned}$$

Notice that the integral in the fourth step integrates to one since it is the integral of the density function of a gamma distribution, integrated over its domain. Thus we have for the first derivative of the general utility function, with ν defined as the gamma density,

$$U'(R) = \frac{\beta^\alpha}{(\phi R + \beta)^\alpha}. \tag{5.4}$$

To calculate absolute risk aversion we require the second derivative of the utility function U, with respect to return R. Differentiating Equation (5.4) with respect to R we have

$$U''(R) = -\frac{\alpha \phi \beta^\alpha}{(\phi R + \beta)^{\alpha+1}}.$$

Therefore, the absolute risk aversion is determined as

$$\begin{aligned} A(R) &= -\frac{U''(R)}{U'(R)} \\ &= -\frac{-\alpha\phi\beta^{\alpha}}{(\phi R+\beta)^{\alpha+1}} \times \frac{(\phi R+\beta)^{\alpha}}{\beta^{\alpha}} \\ &= \frac{\alpha\phi}{\phi R+\beta} \\ &= \frac{1}{\frac{R}{\alpha}+\frac{\beta}{\alpha\phi}}. \end{aligned}$$

This is a hyperbola in the variable R, specified with three parameters, as is required for the HARA class of utility functions. Thus we have proved that the utility functions belonging to the HARA family are nested within the general utility function specified by the marginal utility in Equation (5.1).

This specification of the marginal utility then results in a variable for the coefficient of relative risk aversion that is stochastic in R,

$$\begin{aligned} \gamma(R) &= -\frac{RU''(R)}{U'(R)} \\ &= \frac{R}{\frac{R}{\alpha}+\frac{\beta}{\alpha\phi}} \\ &= \frac{1}{\frac{1}{\alpha}+\frac{\beta}{\alpha\phi R}}. \end{aligned}$$

5.2 Probability of Returns

As in the previous chapter, we require the change of measure formula in terms of the utility and density of stock returns R rather than stock prices S. In this section we derive the formula for the change of measure formula in terms of returns.

We simplify the notation by fixing the time period under consideration, thereby omitting the time subscripts, and letting the initial stock price be given by the positive constant k, as was done in Section 4.2. We have for the return over the fixed period

$$R = \ln\left(\frac{S}{k}\right).$$

Recall the physical density of stock prices from Equation (3.11), subscripted in order to avoid ambiguity,

$$q_S(S) = \frac{U'(S)p_S(S)}{\int_0^{\infty} U'(x)p_S(x)dx}. \tag{5.5}$$

We have, from Equations (4.4) and (4.5) in Section 4.2, the physical and risk-neutral density functions of asset return,

$$p_R(R) = p_S\left(ke^R\right) \times ke^R \tag{5.6}$$

$$q_R(R) = q_S\left(ke^R\right) \times ke^R, \tag{5.7}$$

where the subscript R indicates the density with respect to the asset return. A substitution of Equation

(5.5) into Equation (5.7) allows us to derive the risk-neutral density of asset returns,

$$\begin{aligned} q_R(R) &= q_S\left(ke^R\right) \times ke^R \\ &= \frac{U'(ke^R)p_S(ke^R)}{\int_0^\infty U'(x)p_S(x)dx} \times ke^R \\ &= \frac{U'(ke^R)p_R(R)}{\int_0^\infty U'(x)p_S(x)dx}. \end{aligned}$$

The final equality follows from a substitution of Equation (5.6).

At this point, define another utility function $\bar{U}$ as

$$\bar{U}(R) = \int_{-\infty}^{R} U'(ke^z)dz,$$

where U is the current utility function. We want to demonstrate that this is a valid utility function of a risk averse, non-satiated individual; satisfying the criteria of being twice differentiable and having a positive and decreasing marginal utility function, for all R. For the first derivative of $\bar{U}$, we have

$$\begin{aligned} \bar{U}'(R) &= U'(ke^R) \\ &\geq 0 \quad \text{for all } R \in \mathbb{R}, \end{aligned}$$

since $U'(x) \geq 0$ for all $x \in \mathbb{R}^+$, and the term ke^R will be positive for all $R \in \mathbb{R}$ and positive k. For the second derivative of $\bar{U}$ we have

$$\begin{aligned} \bar{U}''(R) &= U''(ke^R) \times ke^R \\ &\leq 0 \quad \text{for all } R \in \mathbb{R}, \end{aligned}$$

since $U''(x) \leq 0$ for all $x \in \mathbb{R}^+$. Therefore we have shown that $\bar{U}$ is a valid utility function, satisfying the conditions for the preference structure of an investor who is risk averse and non-satiated.

We substitute this new utility function into our formula for the risk-neutral density of returns,

$$\begin{aligned} q_R(R) &= \frac{U'(ke^R)p_R(R)}{\int_0^\infty U'(x)p_S(x)dx} \\ &= \frac{\bar{U}'(R)p_R(R)}{\int_0^\infty U'(x)p_S(x)dx}. \end{aligned}$$

Similarly to the previous chapter, we do a change of the variable of integration in the denominator in order to specify the integral in terms of the density of returns rather than the density of asset prices. Let $y = \ln\left(\frac{x}{k}\right)$. We then have,

$$x = ke^y \quad \Rightarrow \quad dx = ke^y dy.$$

Substituting this into the denominator of the physical density of returns, and appropriately changing the limits of integration, we have

$$\begin{aligned} q_R(R) &= \frac{\bar{U}'(R)p_R(R)}{\int_{-\infty}^\infty U'(ke^y)p_S(ke^y)ke^y dy} \\ &= \frac{\bar{U}'(R)p_R(R)}{\int_{-\infty}^\infty U'(ke^y)p_R(y)dy} \\ &= \frac{\bar{U}'(R)p_R(R)}{\int_{-\infty}^\infty \bar{U}'(y)p_R(y)dy}. \end{aligned}$$

This is similar to the form of Equation (4.6), the formula for the change of measure between the physical density of stock prices and the risk-neutral density of stock prices. From this point forward we will use the form for asset returns rather than stock price, and the subscripts for returns and stock prices will be omitted. We have

$$q(R) = \frac{\bar{U}'(R)p(R)}{\int_{-\infty}^{\infty} \bar{U}'(y)p(y)dy}. \tag{5.8}$$

5.3 Risk-Neutral Skewness

We extend Theorem 4.2 to the case where utility is given by the general utility function implied by Equation (5.1).

Theorem 5.1. *For a general utility function, with marginal utility specified by Equation (5.1), we have as the equation for the skewness in the risk-neutral distribution*

$$m_3 \approx \bar{m}_3 - \left(\phi \int_0^{\infty} z\nu(z)dz\right)(\bar{m}_4 - 3)\,\bar{\sigma}, \tag{5.9}$$

where $m_3, \bar{m}_3, \bar{m}_4, \bar{\sigma}, \nu(z)$ *are defined as before.*

Proof. The proof follows largely from the proof of the previous theorem. Again, for ease of notation, we will omit the time parameters. From Equation (5.8) we have the density function of the risk-neutral distribution of asset returns

$$q(R) = \frac{\bar{U}'(R)p(R)}{\int_{-\infty}^{\infty} \bar{U}'(y)p(y)dy}.$$

Substitution of the marginal utility of the general utility equation, given in Equation (5.1), yields

$$q(R) = \frac{Rz\phi \int_0^{\infty} e^{-z\phi R}\nu(z)dz}{\int_0^{\infty} e^{-z\phi R}\nu(z)dz}.$$

Without loss of generality we can assume that the physical distribution $p(R)$ has a mean of zero (the first moment m_1 is equal to zero). The proof begins by defining the first three uncentered moments of the risk-neutral distribution, with density function $q(R)$. The first moment of the risk-neutral density is obtained as follows, up to first order in z,

$$
\begin{aligned}
\kappa_1 &\equiv \int_{-\infty}^{\infty} Rq(R)dR \\
&= \frac{\int_{-\infty}^{\infty}\int_0^{\infty} Re^{-z\phi R}\nu(z)p(R)dzdR}{\int_{-\infty}^{\infty}\int_0^{\infty} e^{-z\phi R}\nu(z)p(R)dzdR} \\
&= \frac{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)R\left(1 - z\phi R + \frac{(z\phi R)^2}{2!} - \frac{(z\phi R)^3}{3!} + \ldots\right)dzdR}{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)\left(1 - z\phi R + \frac{(z\phi R)^2}{2!} - \frac{(z\phi R)^3}{3!} + \ldots\right)dzdR} \\
&= \frac{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)R\left(1 - z\phi R + o(z)\right)dzdR}{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)\left(1 - z\phi R + o(z)\right)dzdR} \\
&\approx \frac{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)R\left(1 - z\phi R\right)dzdR}{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)\left(1 - z\phi R\right)dzdR} \\
&= \frac{\int_{-\infty}^{\infty} p(R)R\int_0^{\infty} \nu(z)dzdR - \int_{-\infty}^{\infty}\int_0^{\infty} z\nu(z)\phi R^2 p(R)dzdR}{\int_{-\infty}^{\infty} p(R)\int_0^{\infty} \nu(z)dzdR - \int_{-\infty}^{\infty}\int_0^{\infty} z\nu(z)\phi R p(R)dzdR} \\
&= \frac{\bar{\kappa}_1 - \phi\int_0^{\infty} z\nu(z)dz\bar{\kappa}_2}{1 - \bar{\kappa}_1} \\
&= \bar{\kappa}_1 - \left(\phi\int_0^{\infty} z\nu(z)dz\right)\bar{\kappa}_2,
\end{aligned}
$$

where the probability measures $\nu(z)$ and $p(R)$ both integrate to one over their respective domains, and the final equality follows on from the assumption that the physical distribution has a mean of zero. By similar argument, we have for the second and third moments of the risk-neutral distribution,

$$
\begin{aligned}
\kappa_2 &\equiv \int_{-\infty}^{\infty} R^2 q(R)dR \\
&\approx \bar{\kappa}_2 - \left(\phi\int_0^{\infty} z\nu(z)dz\right)\bar{\kappa}_3,
\end{aligned}
$$

$$
\begin{aligned}
\kappa_3 &\equiv \int_{-\infty}^{\infty} R^3 q(R)dR \\
&\approx \bar{\kappa}_3 - \left(\phi\int_0^{\infty} z\nu(z)dz\right)\bar{\kappa}_4.
\end{aligned}
$$

This shows how the first three moments of the risk-neutral distribution are altered by the risk-aversion coefficient of the general utility function implied by Equation (5.1). The centered moments of the physical measure are

$$
\begin{aligned}
m_2 &= \frac{\bar{\kappa}_2}{\bar{\kappa}_2^2} = \sigma^2 \\
m_3 &= \frac{\bar{\kappa}_3}{\bar{\kappa}_2^3} \\
m_4 &= \frac{\bar{\kappa}_4}{\bar{\kappa}_2^4}.
\end{aligned}
$$

Now we calculate the coefficient of skew for the risk-neutral measure:

$$\begin{aligned}
m_3 &= \frac{\int_{-\infty}^{\infty} (R-\kappa_1)^3 q(R)dR}{\left(\int_{-\infty}^{\infty} (R-\kappa_1)^2 q(R)dR\right)^{\frac{3}{2}}} \\
&= \frac{\int_{-\infty}^{\infty} (R^3 - 3R^2\kappa_1 + 3R\kappa_1^2 - \kappa_1^3) q(R)dR}{\left(\int_{-\infty}^{\infty} (R^2 - 2R\kappa_1 + \kappa_1^2) q(R)dR\right)^{\frac{3}{2}}} \\
&= \frac{\kappa_3 - 3\kappa_1\kappa_2 + 2\kappa_1^3}{(\kappa_2 - \kappa_1^2)^{\frac{3}{2}}} \\
&\approx \frac{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)R\,(1 - z\phi R)\, dzdR}{\int_{-\infty}^{\infty}\int_0^{\infty} \nu(z)p(R)\,(1 - z\phi R)\, dzdR} \\
&= \frac{\bar{\kappa}_3 - \gamma(\bar{\kappa}_4 - 3\bar{\kappa}_2^2)}{\bar{\kappa}_2^{\frac{3}{2}}} \\
&= \frac{\bar{\kappa}_3}{\bar{\kappa}_2^{\frac{3}{2}}} - \gamma\frac{\bar{\kappa}_4}{\bar{\kappa}_2^{\frac{3}{2}}} + \gamma\frac{\bar{\kappa}_2^2}{\bar{\kappa}_2^{\frac{3}{2}}} \\
&= \bar{m}_3 - \gamma(\bar{m}_4 - 3)\bar{\sigma},
\end{aligned}$$

where $\gamma = \left(\phi \int_0^{\infty} z\nu(z)dz\right)$ for ease of notation, and $\bar{\sigma}$ is the standard deviation of the physical measure. This completes the proof. □

Theorem 5.1 demonstrates that the utility function defining risk preferences, together with the higher order moments of the physical distribution will affect the level of skewness in the risk-neutral distribution. We see, from Equation (5.9), that although the risk-neutral skew is no longer dependent on the risk aversion coefficient directly, it is still dependent on the functional form of the utility function. This ensures that the preferences of a representative investor will always influence the effect of the second and fourth moments of the physical distribution on the risk-neutral skew.

Through the general form of the utility function, implied by the marginal utility of Equation (5.1), we are able to incorporate different features of an investor's preferences, such as non-stationary utility or areas of risk-seeking behaviour, based on the selected functional form of ν (Bakshi et al.; 2003). We provide examples to demonstrate this.

Example 5.2. We demonstrate that for a certain specification of ν, as specified in Equation (5.1), we are able to obtain a utility function that exhibits risk-seeking behaviour.

Theorem A.10 states that risk aversion is defined by a utility function whereby the second derivative of the utility function is less than or equal to zero, implying that risk-seeking behaviour is categorised by a utility function whereby the second derivative of the utility function is greater than or equal to zero.

Recall that ν is a probability measure defined on the set of positive real numbers. Let ν be defined as a gamma distribution, as is demonstrated in Section 5.1. We then have $\nu \sim \Gamma(\alpha, \beta)$. We then have the marginal utility defined as in Equation (5.4), leading to the general utility function $U(R)$ calculated as,

$$\begin{aligned}
U(R) &= \int U'(R)dR \\
&= \int \frac{\beta^{\alpha}}{(\phi R + \beta)^{\alpha}} dR \\
&= -\frac{\beta^{\alpha}}{\phi(\alpha - 1)} (\phi R + \beta)^{-(\alpha-1)} + C,
\end{aligned}$$

where C is a constant. We now let $\alpha = 2$. We then have

$$U(R) = \frac{\beta^2}{\phi} \frac{1}{(\phi R + \beta)},$$

which results in the first and second derivatives of the utility function U,

$$\begin{aligned} U'(R) &= -\frac{\beta^2}{(\phi R + \beta)^2} \\ U''(R) &= \frac{2\phi\beta^2}{(\phi R + \beta)^3}. \end{aligned}$$

If we let $\beta = 1$ and $\phi = -1$ we have for the second derivative of U,

$$U''(R) = \frac{-2}{-R^3 + 3R^2 - 3R + 1}. \tag{5.10}$$

We know that the utility function U implies risk seeking behaviour if the second derivative of the utility function is greater than or equal to zero. The function defined by Equation (5.10) is positive only when its denominator is negative. It can be shown that the denominator of Equation (5.10) has only one root, $R = 1$; for all $R > 1$ the denominator is negative, therefore the function of the second derivative is positive. Therefore we can conclude that this formulation of the utility function implies investor behaviour that is risk-seeking for $R > 1$.

Example 5.3. In this example we demonstrate the case where a certain specification of ν results in a utility that is non-stationary.

Recall the marginal utility given by Equation (5.1). Let t denote time-steps, where $t \in \mathbb{N}$, and let ν be characterised by the gamma distribution, with parameters t and β, where $t, \beta > 0$. We then have that $U'(R)$ is a discrete, time-varying marginal utility function;

$$U'(R,t) = \frac{\beta^t}{(\phi R + \beta)^t},$$

As is shown in Equation (5.4). Note that $U'(R,t)$ and by implication $U(R,t)$ are a random variables in R.

Our aim is now to demonstrate that $U(R,t)$, for some probability measure ν, is non-stationary; that is, that inter-temporal variation is implied in the process $U(R,t)$.

We have that

$$\begin{aligned} U(R,t) &= \int_{R\in\mathbb{R}} U'(R,t)dR \\ &= \int_{\mathbb{R}} \frac{\beta^t}{(\phi R + \beta)^t} dR \\ &= \frac{\beta^t \left(\phi R + \beta\right)^{-(t-1)}}{-\phi(t-1)}. \end{aligned}$$

If we consider a Taylor expansion of $U(R,t)$, it is clear that the expectation of $U(R,t)$ is a function of both time and a function of the expectation of R,

$$\mathbb{E}\left[U(R,t)\right] = f\left(g\left(\mathbb{E}[R]\right), t\right).$$

This indicates that the mean of $U(R,t)$ varies with time, hence we have shown that the process $U(R,t)$ is not wide-sense stationary, for $E[R^m] \neq 0$, $m \leq t$.

Chapter 6

Conclusion

We have introduced the concept of changing probability measure using the risk aversion of a representative agent. Within a specialized market model, the risk preferences of an agent are fully characterised by a utility function. The risk-neutral measure obtained by the utility function driven change of measure is a martingale measure by design.

The concept of a utility driven change of measure is not new. The central contribution of this research is firstly an exploration into the utility driven change of measure, used by Bakshi et al. (2003) to attribute skewness in the risk-neutral distribution. We then use this framework to review the attribution of skewness in the risk-neutral measure for the power utility function, and the extension to a general utility function (as is illustrated in their paper).

We have summarised the main findings of utility theory, and have provided motivation for measures of risk aversion, as well as describing the conditions for a valid utility function of a rational individual. To aid the case of the general utility function, we have demonstrated how the family of HARA utility functions lead to a class of commonly used utility functions.

The stochastic discount factor was used, in conjunction with no-arbitrage asset pricing, to derive a utility driven change of measure.

Using the power utility function, we have described a closed-form specification of the risk-neutral density based on the constant coefficient of risk aversion and the physical density. We see that, in this case, the risk-neutral density is obtained by exponentially tilting the physical density. This formula for the risk-neutral density was then used to demonstrate how negative skewness in the risk-neutral measure is a result of volatility, negative skew, and kurtosis in the physical measure, and the representative agent's constant coefficient of relative risk aversion.

We then extended this to a more general set of utility functions. A general utility function was posited, and it was shown how this general function houses a wide variety of forms for the utility function. We demonstrated how this general utility function is used to change probability measure, ensuring that the probability measure obtained (risk-neutral) is an equivalent martingale measure. Following a similar methodology to the previous case of power utility, we described a closed-form specification for the risk-neutral density based on risk aversion and the physical density. This formula was then used to attribute skewness in the risk-neutral measure to the volatility, skew, and kurtosis of the physical measure, and the risk aversion of a representative agent. An attractive feature of the model presented is that the underlying distribution does not need to be specified; we simply rely on estimated moments of the unknown physical distribution.

The utility driven change of measure framework has a number of practical applications. Using the model, we are able to derive the risk aversion implied by the market (Detlefsen et al.; 2010). The model also allows us to obtain a physical distribution from the options market implied risk-neutral distribution and a market implied utility (Rosenberg and Engle; 2002). Another usage is the testing of the predictive power of the physical distribution against the underlying index (see Anagnou et al. (2002)).

The methodology of using utility functions to change measure allows us to incorporate investor behaviour into the prices of assets. The framework above is based on the equilibrium consumption assumptions of Cochrane (2001), as well as the canonical assumptions regarding utility and preferences (see Ingersoll (1987) and Pratt (1964)). A possible extension to this research would be to challenge these assumptions and perhaps posit (and test) a different framework for pricing a contingent claim based on intertemporal utility. This framework could allow for more realistic investor behaviour at discrete time points. It should be noted that this would lead to a different relationship between preferences and changing probability measure.

Further extensions of this research could include an investigation into the possible utility structures that are implied by the general utility function; specifically time-varying utility and utility that allows for risk-seeking behaviour. This will allow for a more complex model, albeit at the cost of simplicity and tractability.

A further extension to this research could be the application of the utility driven change of measure to market data from the JSE. One could use the methodology described in Detlefsen et al. (2010) to obtain the risk aversion implied by the JSE market data. The impact of market and economic conditions on this market implied risk aversion could then be studied. One could further estimate the risk-neutral distribution implied by the JSE option market data, and, given an assumed form for the market utility function, derive the implied physical distribution. The forecasting power of this implied physical distribution could then be tested.

Appendix A

Utility Theory

Utility theory is a framework for understanding the decisions made by individuals given their available choices (Fishburn; 1968). We make assumptions about the preferences of an individual. Utility theory allows us to represent these preferences numerically. This in turn allows inferences to be drawn regarding the decision making habits of individuals.

Utility theory is useful in a variety of areas within mathematics of finance; the most notable of which deals with the pricing of assets in incomplete markets. The application of utility theory in this dissertation involves insight into risk aversion and, in particular, the attribution of skew in the risk-neutral measure to the risk aversion of a representative investor.

This chapter will serve to illustrate the key components of utility theory that will be required for application in this dissertation. The research in this chapter is based predominantly on the works of Fishburn (1968), Pratt (1964), Ingersoll (1987), LeRoy and Werner (2000) and Cvitanic and Zapatero (2004).

A.1 Preferences Orderings and Utility Functions

This section provides the mechanisms by which we can numerically represent preferences, and thus describe a utility function for an individual. Note that we are concerned here with preferences between alternatives, rather than absolute measures of worth or satisfaction.

Utility theory is based on a preference operator $\preceq$. This operator indicates the preferred alternative between two choices. Consider the alternative choices, x, y; the statement $x \preceq y$, means that y is preferred at least as much as x. Similarly, *not* $x \preceq y$ means that it is not true that y is preferred as least as much as x. Using this operator, we can infer an indifference relationship, and a strict preference relationship, between alternatives.

Definition A.1. If we have $y \preceq x$ and not $x \preceq y$, then we say that there is *strict preference* for y over x. Strict preference means that y is preferred to x, and is denoted $x \prec y$. If we have $y \preceq x$ and $x \preceq y$, then we say that there is *indifference* between x and y. This is denoted as $x \sim y$; an individual would be equally satisfied with either alternative.

Utility theory is based on the following components; a set of alternatives, assumptions regarding how the preference operator relates the set of alternatives, and the theorems that the resulting preference relationships lead to (Fishburn; 1968). We use preference ordering to develop properties for preferences between alternatives. Let $(x_1, x_2, \ldots, x_n) \in \mathcal{X}$ be a set of available alternatives, satisfying the following axioms:

Axiom 1. *(Completeness) Preferences are fully defined on all available alternatives. For all alternatives* $x_i, x_j \in \mathcal{X}$*, either one or both of the following must hold:* $x_i \succeq x_j$*,* $x_j \succeq x_i$*.*

Axiom 2. *(Transitivity) Preferences remain consistent across available alternatives. If an investor prefers* x_i *to* x_j*, and prefers* x_j *to* x_k*, then he must prefer* x_i *to* x_k*. Formally,* $x_i \succ x_j, x_j \succ x_k \Rightarrow x_i \succ x_k$*.*

Axiom 3. *(Continuity) For all alternatives* $x \in \mathcal{X}$*, the subset of all alternatives strictly preferred is an open set, and the subset of all alternatives* x *is strictly preferred to is an open set.*

Definition A.2. A *utility function* is a map from a set of alternatives onto the real number line, $U : \mathcal{X} \to \mathbb{R}$, with the property

$$U(x_i) > U(x_j) \Leftrightarrow x_i \succ x_j \tag{A.1}$$

$$U(x_i) \geq U(x_j) \Leftrightarrow x_i \succeq x_j \tag{A.2}$$

$$U(x_i) = U(x_j) \Leftrightarrow x_i \sim x_j. \tag{A.3}$$

Utility functions provide an abstract, unobservable measure of satisfaction, given an observable level of input, such as level of consumption or a level of wealth.

The existence of such a utility function is based on the axioms stated above.

Theorem A.3. *(Existence of a Utility Function) For a preference relationship satisfying the axioms of completeness, transitivity and continuity, over a closed and convex set* $\mathcal{X}$*, with* $x_1, x_2 \in \mathcal{X}$*, there exists a continuous function* U *that maps the set of alternatives onto the real line,* $U : \mathcal{X} \to \mathbb{R}$*, satisfying the equations (A.1), (A.2) and (A.3).*

Proof. See Jaffray (1975, Proposition, p. 982). □

The ordinal nature of utility functions, as described above, results in an important property; the preference ordering of a utility function is invariant to positive linear transformations of the utility function (Pratt; 1964). This means that a utility function $U_1(x)$ and another utility function $U_2(x) = a + bU_1(x)$ will have an identical preference ordering, for a and b constants, $b > 0$ (Pratt; 1964).

Definition A.4. (Pratt; 1964) *Equivalent* utility functions are utility functions which exhibit the same preference ordering on the domain of x. The operator $\sim$ will indicate equivalence between utility functions. For two utility functions $U_1(x)$ and $U_2(x)$, where $U_1(x)$ is a positive linear transformation of $U_2(x)$, $U_1(x) = a + bU_2(x)$ for all x, where a and b are constants and $b > 0$, we have $U_1(x) \sim U_2(x)$.

Individuals make decisions so as to maximise their utility (Ingersoll; 1987). This maximisation may be subject to constraints; in the case of maximising utility of consumption, the constraint would be the budget of the individual.

Assumption A.5. We assume *non-satiation*; increased consumption (or wealth) corresponds to greater satisfaction (utility). This implies a utility function $U(x)$ that increases monotonically in x. A monotonically increasing utility function means that more is preferred to less;

$$x_i \geq x_j \Rightarrow U(x_i) \geq U(x_j) \Leftrightarrow x_i \succeq x_j.$$

Cvitanic and Zapatero (2004) state that the assumption of non-satiation conforms to the general expectation of human preference. Mathematically, this implies that the first derivative of the utility function, with respect to the variable of interest, is always positive,

$$U'(x) \geq 0, \quad \text{for all } x \in \mathcal{X}. \tag{A.4}$$

It is common to use consumption choices as the set of alternatives (Cvitanic and Zapatero; 2004). Given that utility functions are ordinal, rather than absolute measures of satisfaction, we can then use utility of wealth as a reliable proxy for an individual's utility of consumption (Ingersoll; 1987). We now consider choices based on uncertain outcomes.

A.2 Expected Utility

Definition A.6. A *risky gamble* y is defined by the vector double $(\mathbf{y}, \mathbf{p})$, with $\mathbf{y}$ representing the set of uncertain payoffs $(y_1, \ldots, y_n)$, and $\mathbf{p}$ representing the respective probabilities of the occurrence of each of these payoffs $(p_1, \ldots, p_n)$, with $0 \leq p_i \leq 1$, $i = (1, \ldots, n)$. The set of probabilities $\mathbf{p}$ sum to one, $\sum_{i=1}^{n} p_i = 1$, indicating that we are certain that one of the outcomes of $\mathbf{y}$ will occur.

The axioms stated in the previous section also apply to risky gambles. To describe a utility function that applies to uncertain outcomes, we require the following two axioms, in addition to those stated above.

Axiom 4. *(Independence) Suppose we have two risky gambles L_1 and L_2, where $L_1 = [(y_1, y_2, y_3), \boldsymbol{p})]$ and $L_2 = [(y_1, z, y_3), \boldsymbol{p})]$. If the individual is indifferent between y_2 and z, that is $y_2 \sim z$, then the individual is indifferent between the two gambles L_1 and L_2.*

Axiom 5. *(Dominance) Suppose we have two risky gambles L_1 and L_2, where $L_1 = [(y_1, y_2), (p_1, 1-p_1)]$ and $L_2 = [(y_1, y_2), (p_2, 1-p_2)]$. If we have $y_1 \succ y_2$, then we have $L_1 \succ L_2$ if and only if $(p_1 > p_2)$.*

Axioms 1 to 5 allow us to describe choice in the presence of uncertainty (Ingersoll; 1987). Utility functions derived from these axioms, in the presence of uncertain outcomes, are known as *von Neumann-Morgenstern utility functions*. We omit 'von Neumann-Morgenstern', assuming from this point forward that all utility functions used satisfy these axioms.

Theorem A.7. *(Expected Utility Theorem) Under the above axioms, investors will choose between a set of risky gambles $\mathbf{y}$ by choosing the gamble that results in the highest expected utility. Expected utility is given as*

$$\mathbb{E}[U(y)] = p_1 U(y_1) + \ldots + p_n U(y_n) = \sum_{i=1}^{n} p_i U(y_i),$$

for a valid utility function $U(y)$.

Proof. See Ingersoll (1987, Theorem 3, p. 10). □

Expected utility is used to order preferences involving uncertain outcomes (risky gambles). The axioms of preferences lead to 'expected utility' as a framework for analysing choice under uncertainty. The theory of decision making in the presence of uncertainty allows us to understand risk appetite.

A.3 Risk Aversion

An individual's risk appetite is determined by the preference between a risky gamble and a deterministic amount equal to the expected outcome of the gamble. Individuals have a specific, identifiable attitude to risk; they are either risk-averse, risk-neutral, or risk seeking.

Definition A.8. A *fair gamble* is a risky gamble y, with the expected outcome of zero; $\mathbb{E}[y] = 0$.

A risk-averse individual would require a premium to enter into a fair gamble; a risk-neutral individual would not require a premium to enter a fair gamble; and a risk-seeking individual would pay a premium to enter a fair gamble.

The risk attitude of an individual is described by their utility function, illustrating their choice under uncertainty. By comparing the utility derived from a certain outcome to the expected utility that would be gained by partaking in a risky gamble, we determine an individual's risk appetite (LeRoy and Werner; 2000).

Definition A.9. Consider an individual with a utility function $U : \mathcal{Y} \to \mathbb{R}$, where $\mathcal{Y}$ is the set of available uncertain alternatives. We compare the individual's preference between a risky gamble and a certain (deterministic) outcome equal to the average outcome of the risky gamble.

- An individual is *risk-averse* if the utility derived from the average outcome of a gamble is greater than the expected utility of the gamble, $\mathbb{E}[U(y)] \leq U[\mathbb{E}(y)]$, with strict inequality for at least one $y \in \mathcal{Y}$.
- An individual is *risk-neutral* if there is indifference between a gamble and the certain outcome $\mathbb{E}[U(y)] = U[\mathbb{E}(y)]$, for all $y \in \mathcal{Y}$.
- An individual is *risk-seeking* if there is a preference for the risky gamble over the deterministic outcome; $\mathbb{E}[U(y)] \geq U[\mathbb{E}(y)]$ for all $y \in \mathcal{Y}$, with strict inequality for at least one $y \in \mathcal{Y}$.

The following theorem describes the relationship between risk aversion and the concavity of a utility function. The theorem states concavity of the utility function implies risk-aversion (LeRoy and Werner; 2000).

Theorem A.10. *(Risk Aversion) An individual is termed 'risk-averse' if and only if the utility function describing his preferences exhibits concavity, that is, the second derivative of the utility function is less than or equal to zero,*

$$U''(y) \leq 0, \quad \text{for all } y \in \mathcal{Y}. \tag{A.5}$$

An individual is termed 'strictly risk-averse' if and only if the utility function describing his preferences is concave, for all levels of wealth; that is, the second derivative of the utility function is less than zero,

$$U''(y) < 0, \quad \text{for all } y \in \mathcal{Y}. \tag{A.6}$$

An individual is termed 'risk-neutral' if and only if the utility function describing his preferences is linear, for all levels of wealth.

$$U''(y) = 0, \quad \text{for all } y \in \mathcal{Y}.$$

Proof. See LeRoy and Werner (2000, Theorem 9.3.1, p. 84). □

Assumption A.11. Individuals are strictly risk-averse, or equivalently, individuals preferences are described fully by a utility function that is concave over all levels of wealth. Given this assumption, the utility functions used to describe preferences will satisfy Equation (A.6).

The assumption of a strictly risk-averse representative investor is common in financial economics (Ingersoll; 1987). For this reason, many of the commonly used utility functions exhibit properties that correspond to risk aversion in the representative agent.

Remark A.12. The assumptions of non-satiation and risk-aversion provide natural restrictions on the mathematical form of the utility function. Firstly, the utility function is required to be twice differentiable. Furthermore, we require that the utility function be monotone increasing and concave. Mathematically, this translates to a utility function with the first derivative greater than or equal to zero, for all levels of wealth (see Equation (A.4)), and the second derivative less than or equal to zero, for all levels of wealth (see Equation (A.6)).

A.4 Measures of Risk Aversion

An investor's attitude to risk will affect their decisions, and in turn, their valuation of risky securities (LeRoy and Werner; 2000). We are interested in providing a measure of an individual's degree of aversion to risk. Pratt (1964) provides objective measures of risk aversion, based on utility functions. In this section we work with a utility function of wealth, $U(w)$, where w represents an individual's level of wealth. The derivations of these measures of risk aversion can be found in Pratt (1964).

Definition A.13. *Absolute risk aversion* is defined as

$$A(w) = -\frac{U''(w)}{U'(w)}, \tag{A.7}$$

for an individual with preferences characterised by the utility function $U(w)$.

We require that equivalent utility functions should exhibit the same risk aversion. The measure of risk aversion should therefore be invariant to positive linear transformations of the utility function. Suppose there are two equivalent utility functions $U_1(w)$ and $U_2(w)$, $U_1(w) \sim U_2(w)$, such that

$$U_1(w) = a + bU_2(w), \quad b > 0.$$

$U_1(w)$ and $U_2(w)$ will have the same preference ordering, and therefore the same risk aversion. The absolute risk aversion measures of both of these utility functions, $A_1(w)$ and $A_2(w)$, are equal

$$A_1(w) = -\frac{bU_1''(w)}{bU_1'(w)} = -\frac{U_1''(w)}{U_1'(w)} = A_2(w).$$

Thus the absolute risk aversion measure preserves the preference ordering, and satisfies our requirements for a valid measure of an individual's level of aversion to risk.

Definition A.14. An individual's level of risk aversion relative to the current level of wealth is known as *relative risk aversion* (LeRoy and Werner; 2000). The relative risk aversion measure is defined as

$$\begin{aligned} R(w) &= -w\frac{U''(w)}{U'(w)} \\ &= wA(w). \end{aligned}$$

This measure is useful for the analysis of risks that are presented as a proportion of wealth rather than an absolute size (Ingersoll; 1987). Relative risk aversion shows how an investor's attitude to risk changes as his level of wealth changes.

The fundamental difference between absolute and relative risk aversion is illustrated in the case of an agent holding a portfolio that consists of a risk-free asset and a risky asset. If the agent exhibits decreasing absolute risk aversion, an increase in the agent's level of wealth will result in an increase the *absolute amount* of the risky asset in the portfolio. If the agent exhibits decreasing relative risk aversion, an increase in the agent's level of wealth will result in an increase the *proportional amount* of the risky asset in the portfolio.

A.5 Hyperbolic Absolute Risk Aversion

Hyperbolic absolute risk aversion (HARA) is an important property in utility theory. This section describes the form and properties of HARA utility functions. HARA utility functions represent a broad family utility functions commonly used in financial economics. The utility functions in this family are of the form

$$U(w) = \frac{1-\gamma}{\gamma}\left(\frac{\beta w}{1-\gamma} + \eta\right)^{\gamma}, \tag{A.8}$$

where β, γ and η are constants, $\eta > 0$, and $w \in \mathbb{R}^+$. Equation (A.8) is then defined on the domain $\frac{\beta w}{1-\gamma} + \eta > 0$, in order to ensure positive utility. For $\gamma < 1$ this is a lower bound on the domain and for $\gamma > 1$ it is an upper bound. Absolute risk aversion (defined by Equation (A.7)) is

$$A(w) = \frac{1}{\frac{w}{1-\gamma} + \frac{\eta}{\beta}}. \tag{A.9}$$

We have decreasing absolute risk aversion for $\gamma < 1$, and increasing absolute risk aversion for $\gamma > 1$. For β, γ and η constants, with $b > 0$, it is clear that Equation (A.9), as a function of $w : w \in \mathbb{R}^+$, is a hyperbola.

Common Utility Functions Derived from HARA Utility

The HARA utility function specification is a general form. We can form a number of commonly used utility functions by selecting specific values for β, η and γ. We now show specific utility functions obtained from the general HARA specification given in Equation (A.8). While the general form of HARA utility functions exhibit hyperbolic absolute risk aversion, the specific examples to follow are special cases and may exhibit increasing, decreasing or constant absolute risk aversion, depending on the specification.

Linear utility

Linear utility is obtained by letting $\gamma \to 1$;

$$U(w) = \beta w + \eta, \quad \beta > 0.$$

While the first derivative of this utility functions is positive, indicating monotonicity and satisfying the property of non-satiation, the second derivative is zero. Linear utility therefore corresponds to a risk-neutral individual.

Quadratic Utility

We obtain quadratic utility by setting $\gamma = 2$,

$$\begin{aligned} U(w) &= -\frac{1}{2}(\eta - \beta w)^2 \\ &= -\frac{1}{2}\left(\eta^2 - 2\beta\eta w + \beta^2 w^2\right) \\ &= -\frac{1}{2}\eta^2 + \beta\eta w - \frac{1}{2}\beta^2 w^2, \end{aligned} \tag{A.10}$$

for constants η and β. The constant term in Equation (A.10), $-\frac{1}{2}\eta^2$, does not change the preference ordering (and therefore the risk aversion). Therefore we can simplify this equation to

$$U(w) = aw - bw^2, \tag{A.11}$$

with $a = \beta\eta$ and $b = \frac{1}{2}\beta^2$, $b \geq 0$, for all $\beta \in \mathbb{R}$. The first and second derivatives of this utility function (as given in Equation (A.11)) are $U'(w) = a - 2bw$ and $U''(w) = -2b$ respectively. The second derivative is negative over its domain, satisfying the property of risk aversion. The first derivative is positive only for $w < \frac{a}{2b}$; this utility function exhibits decreasing utility for increasing wealth when $w \geq \frac{a}{2b}$. Thus the property of monotonicity is not satisfied for all values of w for this utility function.

The absolute risk aversion for the quadratic utility is

$$A(w) = -\frac{U''(w)}{U'(w)} = \frac{2b}{a - 2bw}.$$

The first derivative of the absolute risk aversion function, with respect to wealth, is $A'(w) = \frac{4b^2}{(a-2bw)^2}$ which is greater than zero for all positive w. Thus the quadratic utility function has *increasing absolute risk aversion*. This implies unrealistic behaviour as increasing absolute risk aversion does not fit empirical evidence of investor behaviour (Cvitanic and Zapatero; 2004). However, the quadratic utility function is tractable for use in portfolio selection or asset pricing (e.g. CAPM). Quadratic utility allows for mean-variance optimisation (Cvitanic and Zapatero; 2004).

Exponential Utility

In Equation (A.8), let $\eta = 1$, and let $\gamma \to -\infty$. Taking the limit of this equation results in a negative exponential utility function,

$$U(w) = -e^{-\beta w}, \tag{A.12}$$

for all positive w. We have

$$U'(w) = \beta e^{-\beta} w > 0 \quad \text{for all } w \in \mathbb{R}^+$$
$$U''(w) = -\beta^2 e^{-\beta} w < 0 \quad \text{for all } w \in \mathbb{R}^+.$$

This utility function is monotone increasing and concave over its domain, therefore satisfying the requirements of non-satiation and risk-aversion. The absolute risk aversion function is constant and the relative risk aversion increases linearly with wealth;

$$A(w) = -\frac{-\beta^2}{\beta}\frac{e^{-\beta}w}{e^{-\beta}w} = \beta$$
$$R(w) = -w\frac{-\beta^2}{\beta}\frac{e^{-\beta}w}{e^{-\beta}w} = \beta w.$$

Thus the exponential utility has the property of constant absolute risk aversion; the absolute aversion to risk is independent of the investor's level of wealth. Investors are willing to gamble the same absolute amount regardless of their current level of wealth. This utility function exhibits relative risk aversion that increases linearly with wealth.

Power Utility

We obtain the power utility function from Equation (A.8) by setting $\eta = 0$ and $\gamma < 1$,

$$\begin{aligned} U(w) &= \frac{1-\gamma}{\gamma}\left(\frac{\beta w}{1-\gamma}\right)^{\gamma} \\ &= \frac{(1-\gamma)\beta^{\gamma}}{(1-\gamma)^{\gamma}}\frac{w^{\gamma}}{\gamma} \\ &= \frac{aw^{\gamma}}{\gamma}, \end{aligned} \tag{A.13}$$

where $a = \frac{(1-\gamma)\beta^{\gamma}}{(1-\gamma)^{\gamma}}$ is a constant. The first derivative of Equation (A.13) is $U'(w) = aw^{(\gamma-1)}$, which is positive for all γ when a and w are positive. The second derivative of Equation (A.13) is $U''(w) = (\gamma-1)aw^{(\gamma-2)}$, which is negative for $\gamma < 1$, positive a and positive w. This indicates that $U(w)$ is increasing and concave for $\gamma < 1$, positive a and positive w, satisfying the requirements for non-satiation and risk-aversion.

The absolute and relative risk aversion functions are then

$$\begin{aligned} A(w) &= \frac{(\gamma-1)}{w} \\ R(w) &= (\gamma-1), \end{aligned}$$

indicating decreasing absolute risk aversion, and constant relative risk aversion (CRRA). This implies that an individual's level of relative risk aversion is not dependent on his current level of wealth. The individual is willing to forgo the same proportion of wealth to avoid a risky gamble, at all levels of wealth. However, as wealth increases, the amount the individual would pay to forgo a risky gamble decreases. A powerful property of power utility is that it allows for the aggregation of utility of individuals with different wealth levels. Thus we can represent the aggregation of utility as a single representative individual (Campbell; 2003). Another feature of the power utility function is that the coefficient of relative risk aversion is constant.

Logarithmic Utility

Logarithmic utility is obtained from the power utility function, by taking the limit as $\gamma \to 0$. Since Equation (A.8) is not defined for $\gamma = 0$, we apply l'Hōpital's rule to a utility function equivalent to the power utility given in Equation (A.13) in order to derive the logarithmic utility function. Consider the utility function given by $U(w) = \frac{aw^{\gamma}-1}{\gamma}$, and let $a = 1$. Clearly this is a form of the power utility function, and is equivalent to that of Equation (A.13). Applying l'Hōpital's rule we have

$$\begin{aligned} \lim_{\gamma \to 0} \frac{w^{\gamma}-1}{\gamma} &= \lim_{\gamma \to 0} \frac{w^{\gamma}\ln(w)}{1} \\ &= \ln(w). \end{aligned} \tag{A.14}$$

Equation (A.14) is a logarithmic utility function. We have $U'(w) = \frac{1}{w}$ and $U''(w) = -\frac{1}{w^2}$, which are positive and negative respectively, for all positive w. This indicates that this utility function is monotone increasing and concave, satisfying the conditions of non-satiation and risk-aversion.

The absolute and relative risk aversion functions are then

$$\begin{aligned} A(w) &= \frac{1}{w^2} \bigg/ \frac{1}{w} = \frac{1}{w} \\ R(w) &= 1, \end{aligned}$$

indicating decreasing absolute risk aversion, and constant relative risk aversion, for all positive w. Logarithmic utility results in the same risk aversion characteristics as power utility.

Appendix B

Probability Theory

This appendix is a summary of results from probability theory, which supplement the work in this dissertation. The results in this section are taken from Björk (2009), Jacod and Protter (2003), Shiryaev and Boas (1995), and Ross and Pekoz (2007).

B.1 Measurable Space

Let Ω be a non-empty set. Denote the collection of all subsets of Ω as 2^{Ω}. This is known as the power set of Ω. Let $\emptyset$ denote the null set.

Definition B.1. A *σ-algebra* is a collection of subsets of Ω, denoted $\mathcal{F}$, which has the properties

1. $\emptyset \in \mathcal{F}$
2. $\forall A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A σ-algebra $\mathcal{F}$ must contain the empty set, be closed under complement and closed under countable union. This implies that $\mathcal{F}$ will be closed under countable intersection as well.

Example B.2. $\mathcal{F} = \{\emptyset, \Omega\}$ is the smallest σ-algebra, and $\mathcal{F} = 2^{\Omega}$ is the largest σ-algebra on a non-empty set Ω.

Definition B.3. Let E be a collection of subsets on a sample space Ω. The σ-algebras obtained by the intersection of all σ-algebras that contain E, is the smallest σ-algebra on Ω containing E. This is known as the *σ-algebra generated by* E, and is denoted $\sigma(E)$. Stated more formally, for any σ-algebra D on Ω containing E, $\sigma(E) \subseteq D$.

Definition B.4. Consider an infinite set. If the items in this set can be arranged sequentially, then the set is termed *countably infinite*. If, however, the items in this set cannot be arranged in sequence, then the set is termed *uncountably infinite*.

Example B.5. The set of integers $\mathbb{Z}$ is countably infinite, since the integers can be arranged sequentially. The set of real numbers on the interval $(0, 1)$ is uncountably infinite, since in is impossible to arrange these real numbers sequentially.

Definition B.6. A *measurable space* is the double $(\Omega, \mathcal{F})$ where Ω is a non-empty set and $\mathcal{F}$ is a σ-algebra on the set Ω. Define *$\mathcal{F}$-measurable sets* as the subsets of Ω which are present in the σ-algebra $\mathcal{F}$).

Definition B.7. A *filtration* $\{\mathcal{F}_t\}_{t\geq 0}$ is defined as a sequence of increasing σ-algebras on the set Ω; $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n \subseteq \ldots \subseteq \mathcal{F}$. We assume that the set $\mathcal{F}_0$ contains all sets of measure zero, while $\mathcal{F}_\infty = \sigma\left(\bigcup_n \mathcal{F}_n\right) \subseteq \mathcal{F}$. The set $\mathcal{F}_n$ can be thought of as the information that is available at time n.

B.2 Probability Measures

Definition B.8. Let A and B be two sets. If the intersection of these two sets is the empty set, $A \cap B = \emptyset$, then these sets are *pairwise disjoint*.

Definition B.9. Let A and B be two sets. The *disjoint union* of A and B combines the elements of A and B, while indexing these elements by their original set. This binary operator will be denoted as $\uplus$; i.e. the disjoint union of these sets is denoted $A \uplus B$.

Example B.10. Let $A = \{1, 2, 3, 4\}$ and let $B = \{2, 4, 6, 8\}$. The disjoint union of A and B is equal to the union of $A^* = \{(1, 0), (2, 0), (3, 0), (4, 0)\}$ and $B^* = \{(2, 1), (4, 1), (6, 1), (8, 1)\}$. We then have

$$A \uplus B = A^* \cup B^* = \{(1, 0), (2, 0), (3, 0), (4, 0), (2, 1), (4, 1), (6, 1), (8, 1)\}.$$

Definition B.11. Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mathbb{P}$ be a function on this space such that $\mathbb{P} : \mathcal{F} \to \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers. The function $\mathbb{P}$ is *countably additive* if and only if

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

This means that the function $\mathbb{P}$ applied to the union of pairwise disjoint elements of the σ-algebra $\mathcal{F}$, is equal to the sum of the function $\mathbb{P}$, applied to each of the elements.

Definition B.12. Let $(\Omega, \mathcal{F})$ be a measurable space. The elements A of σ-algebra $\mathcal{F}$ are known as *events*. A *probability measure* is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ defined on a σ-algebra, satisfying

1. $\mathbb{P}(\Omega) = 1$
2. $\mathbb{P}$ is countably additive on $\mathcal{F}$ for all pairwise disjoint, countable sets A_i, $i \geq 1$; that is

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

$\mathbb{P}(A)$ is defined as the *probability* of the event A. $\mathbb{P}(A) = 1$ means that the event A will *almost surely* occur. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*; with Ω known as the *sample space*.

Definition B.13. Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mathbb{P}$ be a function on this space such that $\mathbb{P} : \mathcal{F} \to \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers. $\mathbb{P}$ is called a *finitely additive measure* if $\mathbb{P}(\Omega) < \infty$, and a *finitely additive probability measure* if $\mathbb{P}(\Omega) = 1$.

Lemma B.14. *For a probability space defined by* $(\Omega, \mathcal{F}, \mathbb{P})$*, we have*

1. $\mathbb{P}(A) \geq 0$ *for all* $A \in \mathcal{F}$.
2. $\mathbb{P}(\emptyset) = 0$.
3. $\mathbb{P}(\Omega) = 1$.

4. *$\mathbb{P}$ is countably additive; for all pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{F}$, we have*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}\left(A_i\right).$$

5. *$\mathbb{P}$ is finitely additive; for all pairwise disjoint sets $A_1, A_2, \ldots A_n \in \mathcal{F}$, we have*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}\left(A_i\right).$$

6. *For $A \subseteq B$, $A, B \in \mathcal{F}$, we have*

$$\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A).$$

In general

$$\mathbb{P}(A) = \mathbb{P}(\Omega \backslash A^c) = \mathbb{P}(\Omega) - \mathbb{P}(A).$$

7. *For $A \subseteq B$, $A, B \in \mathcal{F}$, we have*

$$\mathbb{P}(A) \leq \mathbb{P}(B)$$

In general, for all $A \in \mathcal{F}$

$$\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1.$$

8. *For all $A, B \in \mathcal{F}$, we have*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

9. *For all increasing sequences $A_1, A_2, \cdots \in \mathcal{F}$, we have* continuity from below

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

10. *For all decreasing sequences $A_1, A_2, \cdots \in \mathcal{F}$, we have* continuity from above

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

Proof. (1), (2) and (4) follow from the definition of a probability measure.

(4) To determine $\mathbb{P}(\emptyset)$, the probability of the empty set, notice that from the definition of probability measures we have $\mathbb{P}(\emptyset) \in [0,1]$ and $\mathbb{P}(\emptyset) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$; hence $\mathbb{P}(\emptyset) = 0$.

(5) If we choose a sequence $A_n, n \geq 1$ such that $A_{n+i} = \emptyset$ for $i \geq 1$, then (5) follows directly from (4) and (2).

(6) For A a subset of B, we have $B = (B \backslash A) \uplus A$, where $\uplus$ is the *disjoint union* operator. Therefore, it follows from (5) that $\mathbb{P}(B) = \mathbb{P}(B \backslash A) + \mathbb{P}(A)$, and thus we have the result.

(7) Given $A \subseteq B$, we have $\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$ from (5), with each of these terms greater or equal to zero by definition. The result follows, and the general case holds by similar argument.

(8) We can decompose the union of A and B as follows

$$A \cup B = (A \backslash (A \cap B)) \uplus (A \cap B) \uplus (B \backslash (A \cap B)).$$

Hence we have

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).\end{aligned}$$

(9) Consider an increasing sequence with $A_0 = \emptyset$, $A_1 \subseteq A_2 \subseteq \ldots \in \mathcal{F}$. The probability measure of the union of the elements of this sequence is

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \backslash A_{i-1}\right) \\ &= \sum_{i=1}^{\infty} (\mathbb{P}(A_i) - \mathbb{P}(A_i)) \\ &= \lim_{i=1}^{\infty} \sum_{i=1}^{n} (\mathbb{P}(A_i) - \mathbb{P}(A_i)) \\ &= \lim_{i=1}^{\infty} \mathbb{P}(A_n),\end{aligned}$$

the final equality following by a telescopic sum.

(10) Consider a sequence $(A_1, A_2, \ldots)$ that is decreasing; $A_1 \supseteq A_2 \supseteq \ldots \in \mathcal{F}$. The sequence constructed as $(A_1 \backslash A_i)$, $i \in (1, 2, \ldots)$ is an increasing sequence. Therefore by (9) we have

$$\lim_{n\to\infty} \mathbb{P}(A_1 \backslash A_n) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_1 \backslash A_i)\right).$$

By (5) we have

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(A_n) &= \mathbb{P}(A_1) - \lim_{n\to\infty} \mathbb{P}(A_1 \backslash A_n) \\ &= \mathbb{P}(A_1) - \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_1 \backslash A_i)\right) \\ &= \mathbb{P}(A_1) - \mathbb{P}\left(A_1 \backslash \bigcup_{i=1}^{\infty} A_i\right) \\ &= \mathbb{P}(A_1) - \left(\mathbb{P}(A_1) - \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right)\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right).\end{aligned}$$

□

B.3 Integration

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $f : \Omega \to \mathbb{R}$ be a function. The aim of this section is to provide a definition for the expression

$$\int_{\Omega} f(\omega) d\mathbb{P}(\omega).$$

Definition B.15. Define the *indicator function* $\mathbb{I}_A$ as

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is an element of } A \\ 0 & \text{if } \omega \text{ is not an element of } A, \end{cases}$$

for any $A \in \Omega$.

Definition B.16. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, define a *simple function* $f : \Omega \to \mathbb{R}$ as a function that can be expressed as

$$f(\omega) = \sum_{i=1}^{n} c_i \cdot \mathbb{I}_{A_i}(\omega),$$

for $c_1, \ldots, c_n \in \mathbb{R}$ and $A_1, \ldots, A_n \in \mathcal{F}$.

Definition B.17. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a simple function $f : \Omega \to \mathbb{R}$, the integral of f over the space Ω is defined

$$\begin{aligned} \int_\Omega f(\omega) d\mathbb{P}(\omega) &= \sum_{i=1}^{n} c_i \int_\Omega \mathbb{I}_{A_i} d\mathbb{P} \\ &= \sum_{i=1}^{n} c_i \cdot \mathbb{P}(A_i). \end{aligned}$$

It is easy to see that the following holds for all simple measurable functions f and g defined on the probability space,

$$\int (\alpha f + \beta g)\, d\mathbb{P} = \alpha \int f d\mathbb{P} + \beta \int g d\mathbb{P},$$

for $\alpha, \beta \geq 0$.

We extend this definition of the integral to a class of functions called measurable functions. These are functions which we can approximate using simple functions.

Definition B.18. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a function $f : \Omega \to \mathbb{R}$ is defined as $\mathcal{F}$-*measurable* if the following holds

$$\{\omega \in \Omega; f(\omega) \in I\} \in \mathcal{F}, \quad \forall I \subseteq \mathbb{R}.$$

This means that for every interval $I \subseteq \mathbb{R}$, we have $f^{-1}(I) \in \mathcal{F}$. Thus the function f is a *measurable function* on this probability space. A *non-negative measurable function* on this probability space is a function $g : \Omega \to \mathbb{R}^+$.

Definition B.19. Consider a non-negative measurable function g on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let φ be a class of simple functions on this space such that $0 \leq \varphi \leq f$. The integral of the function f with respect to the probability measure $\mathbb{P}$ is defined by

$$\int_\Omega f(\omega) d\mathbb{P}(\omega) := \sup_{\varphi} \int_\Omega \varphi(\omega) d\mathbb{P}(\omega).$$

Definition B.20. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with an $\mathcal{F}$-measurable function f; the function f is *integrable* if the integral of the absolute value of the function over the space Ω is finite; i.e.

$$\int_\Omega |f(\omega)| d\mathbb{P}(\omega) < \infty.$$

This is written mathematically as $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$; or $f \in \mathcal{L}^1$ when the probability space is not ambiguous.

The following properties are stated without proof. They are a result of the manner in which the integration with respect to a probability measure is defined, together with the Monotone Convergence Theorem.

Proposition B.21. *Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with non-negative measurable functions $f, g \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. We have the following properties of the integral with respect to the probability measure $\mathbb{P}$.*

1. *For all $\alpha, \beta \in \mathbb{R}$, we have*
$$\int (\alpha f(\omega) + \beta g(\omega))\, d\mathbb{P}(\omega) = \alpha \int f(\omega) d\mathbb{P}(\omega) + \beta \int g(\omega) d\mathbb{P}(\omega)$$
2. *If $f(\omega) \leq g(\omega)$, for all ω, then*
$$\int_\Omega f(\omega) d\mathbb{P}(\omega) \leq \int_\Omega g(\omega) d\mathbb{P}(\omega).$$
3. *For all integrable functions on this probability space, we have*
$$\left| \int_\Omega f(\omega) d\mathbb{P}(\omega) \right| = \int_\Omega |f(\omega)| d\mathbb{P}(\omega).$$

Definition B.22. For any measurable set A in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with an integrable function f, define the integral of f over A as
$$\int_A f(\omega) d\mathbb{P}(\omega) = \int_\Omega I_A(\omega) f(\omega) d\mathbb{P}(\omega).$$
The following terms are represent the same integral, and may be used interchangeably;
$$\begin{aligned} \int_A f(\omega) d\mathbb{P}(\omega) &= \int_A f(\omega) \mathbb{P}(d\omega) \\ &= \int_A f d\mathbb{P}. \end{aligned}$$

This defines integration for non-negative measurable functions on probability measures. Integration over functions which are not necessarily non-negative is outside of the scope of this work and therefore will not be included here.

B.4 Random Variables

Let the sample space Ω be the set of real numbers $\mathbb{R}$.

Definition B.23. The *Borel σ-algebra* is the smallest σ-algebra containing $\mathbb{R}$. It is the σ-algebra generated by open sets on $\mathbb{R}$. This is the same as the σ-algebra generated by closed sets on $\mathbb{R}$, since σ-algebras are closed under complement. The Borel σ-algebra is denoted $\mathfrak{B}(\mathbb{R})$. Define *Borel sets* as the sets within $\mathfrak{B}(\mathbb{R})$. A *Borel measure* is a measure that is defined on a σ-algebra of Borel sets.

Proposition B.24. *Consider the measurable space $\Omega, \mathcal{F}$ and the function $f : \Omega \to \mathbb{R}$. The function f is $\mathcal{F}$-measurable if and only if $f^{-1}(B) \in \mathcal{F}$ for all Borel sets $B \subseteq \mathbb{R}$.*

Proof. See Björk (2009, Proposition A.43, p. 475). □

Definition B.25. Consider the measurable space defined by $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. Let $f : \mathbb{R} \to \mathbb{R}$ be a real valued function, such that f is measurable with respect to $\mathfrak{B}(\mathbb{R})$. This function is called *Borel measurable* and is known as a *Borel function*.

Definition B.26. A *random variable* is a function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ such that

$$\forall B \in \mathfrak{B}(\mathbb{R}), \quad X^{-1}(B) \in \mathcal{F}.$$

A random variable is then a function $X : \Omega \to \mathbb{R}$ which assigns to every event $\omega \in \Omega$ a real number.

A random variable is therefore a mapping from the state space Ω to the real number line $\mathbb{R}$. It is not known what value X will take; it represents the outcome of an unknown event.

Definition B.27. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. The *probability distribution* of the random variable X on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ is defined by

$$\begin{aligned} P_X(A) &= \mathbb{P}([\omega : X(\omega) \in A]) \\ &= \mathbb{P}(X^{-1}(A)) \end{aligned}$$

for all $A \in \mathfrak{B}(\mathbb{R})$.

Definition B.28. A *distribution function* is a function $F : \mathbb{R} \to [0, 1]$ that satisfies the following conditions

1. F is a non-decreasing function,
2. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ and $F(\infty) = \lim_{x \to \infty} F(x) = 1$,
3. F is right-continuous and has a left limit for all $x \in \mathbb{R}$.

Definition B.29. The *distribution function* of a random variable X defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if a function $F_X : \mathbb{R} \to (0, 1)$ defined by

$$F_X(x) = \mathbb{P}(\omega : X(\omega) \leq x), \quad x \in \mathbb{R}.$$

We refer to the function F_X as the distribution function that corresponds to the probability measure $\mathbb{P}$.

Theorem B.30. *For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, define a function $F_X : \mathbb{R} \to [0, 1]$ as*

$$F_X(x) = \mathbb{P}(X^{-1}(-\infty, x]).$$

This function F_X is a distribution function on the set of real numbers, also known as the cumulative distribution function *of X.*

Proof. We prove F is a non-decreasing function by noting that for $s < t$, it holds that $(-\infty, s] \subseteq (-\infty, t]$, therefore is must hold that $X^{-1}(-\infty, s] \subseteq X^{-1}(-\infty, t]$. By the properties of probability measures (Lemma (B.14), (7))

$$F_X(a) = \mathbb{P}(X^{-1}(-\infty, s]) \leq \mathbb{P}(X^{-1}(-\infty, s]) = F_X(b).$$

In order to show $F_X(\infty) = \lim_{x \to \infty} F_X(x) = 1$, consider a sequence of real numbers diverging to infinity, $(x_i)_{i \in \mathbb{N}}$, (where $\mathbb{N}$ is the set of natural numbers)

$$\forall S \in \mathbb{N}, \quad \exists T \in \mathbb{N} \text{ such that } i \geq T \Rightarrow x_i \geq S.$$

Create a new sequence from $(x_i)_{i\in\mathbb{N}}$ by taking, for each $S \in \mathbb{N}$ the first elements of $(x_i)_{i\in\mathbb{N}}$, the first value x_i greater than or equal to S. The resulting set $(x_{i_S})_{S\in\mathbb{N}}$ is non-decreasing and divergent to infinity. We now have

$$\begin{aligned}
1 = \mathbb{P}(X_{-1}(\mathbb{R})) &= \left(X^{-1}\bigcup_{s=1}^{\infty}(-\infty, x_{i_s}]\right) \\
&= \left(\bigcup_{s=1}^{\infty} X^{-1}(-\infty, x_{i_s}]\right) \\
&\leq \left(\bigcup_{i=1}^{\infty} X^{-1}(-\infty, x_i]\right) \\
&\leq 1,
\end{aligned}$$

the first inequality holds because $\mathbb{P}$ is an increasing probability measure and

$$\bigcup_{s=1}^{\infty} X^{-1}(-\infty, x_{i_s}] \subseteq \bigcup_{s=1}^{\infty} X^{-1}(-\infty, x_i].$$

The prove of $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0$ is done similarly.

We now prove that the function F_X is right continuous. Let $(x_i)_{i\in\mathbb{N}}$ be a sequence of numbers on the real number line that is non-increasing, such that the sequence converges to some $x \in \mathbb{R}$. Given that the sequence $((-\infty, x_i]_{i\in\mathbb{N}})$ is non-increasing and is a sequence of Borel sets, it must hold that $(X^{-1}(-\infty, x_i])_{i\in\mathbb{N}}$ is sequence in $\mathcal{F}$ that is non-increasing. Therefore we have

$$\begin{aligned}
\lim_{i\to\infty} F_X(x_i) &= \lim_{i\to\infty} \mathbb{P}(X^{-1}(-\infty, x_i]) \\
&= \mathbb{P}\left(\bigcap_{i=1}^{\infty} X^{-1}(-\infty, x_i]\right) \\
&= \mathbb{P}\left(X^{-1}\left(\bigcap_{i=1}^{\infty}(-\infty, x_i]\right)\right) \\
&= \mathbb{P}\left(X^{-1}(-\infty, x]\right) \\
&= F_X(x).
\end{aligned}$$

This proves that F_X is continuous from the right.

To prove that F_X has a limit on the left, construct the sequence of real numbers $(x_i)_{i\in\mathbb{N}}$, such that this sequence is non-decreasing and converges to some $x \in \mathbb{R}$. We then have $(x_i)_{i\in\mathbb{N}}$ as a sequence of Borel sets that is non-decreasing, such that the sequence $(X^{-1}(-\infty, x_i])_{i\in\mathbb{N}}$ is also non-decreasing in $\mathcal{F}$. Therefore we have

$$\begin{aligned}
\lim_{i\to\infty} F_X(x_i) &= \lim_{i\to\infty} \mathbb{P}(X^{-1}(-\infty, x_i]) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{\infty} X_{-1}(-\infty, x_i]\right) \\
&= \mathbb{P}\left(X_{-1}\left(\bigcup_{i=1}^{\infty}(-\infty, x_i]\right)\right) \\
&= \mathbb{P}\left(X^{-1}(-\infty, x)\right).
\end{aligned}$$

Thus we have proved that F_X has a limit from the left, and we can conclude that the function F_X is a distribution function for a random variable X. □

Remark B.31. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. We know that the probability measure $\mathbb{P}$ is a function of events in the σ-algebra on the sample space Ω, that is, $\mathbb{P}$ is a function of events in $\mathcal{F}$. Clearly, by the definition of a random variable, the event $A = X^{-1}(-\infty, x]$ is an element of $\mathcal{F}$, for some $x \in \mathbb{R}$. Let the event represented by $X \leq x$ denote the same event. Therefore $X^{-1}(-\infty, x)$ and $(X \leq x)$ are the same event, and

$$\mathbb{P}(X^{-1}(-\infty, x]) = \mathbb{P}(X \leq x),$$

for all $x \in \mathbb{R}$.

Note that the arguments taken by the distribution function F are real numbers, while the arguments taken by the probability function $\mathbb{P}$ are events, or subsets of the σ-algebra $\mathcal{F}$.

Theorem B.32. *The distribution function induced by the probability measure* $\mathbb{P}$ *on the measure space* $(\mathbb{R}, \mathfrak{B})$*, uniquely characterizes the probability. That is, if there exists another probability measure* $\mathbb{Q}$ *such that*

$$G(x) = \mathbb{Q}((-\infty, x]), \quad \text{for all } x \in \mathbb{R},$$

and $F = G$*, then* $\mathbb{P} = \mathbb{Q}$.

Proof. See Jacod and Protter (2003, Theorem 7.1, p. 39). □

The above theorem implies that we can determine a unique distribution function F from a probability measure $\mathbb{P}$. Therefore, from the distribution function f, the probability of any Borel set $A \in \mathfrak{B}$, $\mathbb{P}(A)$ can be determined. We now characterize all distribution functions.

Theorem B.33. *A function* F*, satisfying the following criteria, is the distribution function of a probability measure on the measurable space* $(\mathbb{R}, \mathfrak{B})$*;*

1. F *is a non-decreasing function,*
2. F *is continuous from the right,*
3. *we have* $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to +\infty} F(x) = 1$.

Proof. See Jacod and Protter (2003, Theorem 7.2, p. 40). □

Definition B.34. The *density function* of a probability measure $\mathbb{P}$ on the measurable space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ is a positive function $f : \mathbb{R} \to [0, 1]$, that is Borel measurable, and satisfies

$$\mathbb{P}(X^{-1}(-\infty, x]) = \int_{-\infty}^{x} f(z)dz.$$

Where $\mathbb{P}$ is the probability measure of a random variable X, then f is the density of X. From the theorem (B.30) we have

$$\begin{aligned} F_X(x) &= \mathbb{P}(X^{-1}(-\infty, x]) \\ &= \int_{-\infty}^{x} f(z)dz, \end{aligned} \tag{B.1}$$

this relates the distribution function of a random variable to the density function. It is natural then, due to the fact that F is right continuous, that the following holds

$$\begin{aligned} \int_a^b f_X(x)dx &= F_X(b) - F_X(a) \\ &= \mathbb{P}(X^{-1}[-\infty, b)) - \mathbb{P}(X^{-1}[-\infty, a)), \end{aligned}$$

and for the density at a point

$$f_X(a) = F_X(a) - \lim_{x \to a^-} F_X(x).$$

This is equal to zero since the distribution function is continuous at all $a \in \mathbb{R}$.

Therefore, on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, the following holds for the probability measure $\mathbb{P}$

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(X^{-1}(-\infty, x]) \\ &= \int_{-\infty}^{x} f(z)dz, \end{aligned}$$

where $F(x)$ is the cumulative density function for the random variable X, and $f(x)$ is the corresponding probability density function for X. When we have F differentiable, then f is the derivative of F. If f is Riemann integrable, then by definition $\int_{-\infty}^{\infty} f(x)dx = 1$. Note, not all distribution functions allow for a density function. Equation (B.1) implies that the distribution function F is a continuous function. In fact there are non-continuous distribution functions that do not permit a density. We demonstrate below that where this density function exists it characterizes the probability measure entirely.

Definition B.35. A *continuous* random variable X is a random variable for which such a density function exists.

B.5 Expected Value

The aim in this section is to define, for a random variable, the conditional expectation, as well as the expectation with respect to a specific probability measure. We begin with the definitions for the expected value of a continuous random variable, finite expectations and integrability, and a theorem regarding the properties of expectations of random variables.

Definition B.36. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. The *expectation* or expected value of X is defined as

$$\mathbb{E}[X] = \int_{\Omega} X(\omega)\mathbb{P}(\omega)d\omega. \tag{B.2}$$

This definition is also referred to as the expectation of X *with respect to probability measure* $\mathbb{P}$. Equation (B.2) is often written as $\int_\Omega X\mathbb{P}(d\omega)$, or $\int_\Omega X d\mathbb{P}$. If the random variable X was similarly defined on a space $(\Omega, \mathcal{F}, \mathbb{Q})$, where $\mathbb{Q}$) was a probability measure different to $\mathbb{P}$), then we could substitute this new measure $\mathbb{Q}$) into equation (B.2). In the presence of another probability measure the expectation in equation (B.2) is often denoted as $\mathbb{E}_{\mathbb{P}}[X]$ in order to remove all ambiguity.

Given a random variable X, let $X^+ = \max(0, X)$ and $X^- = -\min(0, X)$. Then we have $X^+, X^- \geq 0$, and the following holds

$$\begin{aligned} X &= X^+ - X^- \\ |X| &= X^+ + X^-. \end{aligned}$$

Definition B.37. If both $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are finite, then the random variable X has a *finite expectation*. We also say that X is *integrable*. The expectation of X is given as

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]. \tag{B.3}$$

The set of all integrable random variables is denoted as $\mathcal{L}^1$, or $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ to remove all ambiguity.

Definition B.38. For a random variable X, if $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are not both equal to infinity, then we say that X *admits an expectation.* In this case the expectation is still given by equation (B.3).

Theorem B.39. *Let X and Y be two random variables.*

1. *The expectation operator is a linear map on the vector space $\mathcal{L}^1$. The expectation operator is positive, that is*
$$X \geq 0 \Rightarrow \mathbb{E}[X] \geq 0.$$
If we have two random variables X and Y such that $Y \in \mathcal{L}^1$ and $0 \leq X \leq Y$, then $X \in \mathcal{L}^1$ and $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

2. *Any random variable that is bounded is integrable. We have $X \in \mathcal{L}^1$ if and only if $|X| \in \mathcal{L}^1$, and $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.*

3. *If we have $X = Y$ almost surely (a.s.), then $\mathbb{E}[X] = \mathbb{E}[Y]$. $X = Y$ a.s. if $\mathbb{P}(X = Y) = \mathbb{P}([\omega : X(\omega) = Y(\omega)]) = 1$.*

4. *(Monotone convergence theorem): If we have a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ that are positive and increasing a.s. to X, then we have*
$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

5. *(Lebesgue's dominated convergence theorem): If we have a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ that converge a.s. to X and $|X_n \leq Y$ a.s. $\in \mathcal{L}^1$ for all n, then $X_n \in \mathcal{L}^1, X \in \mathcal{L}^1$, and $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

Proof. See (Jacod and Protter; 2003, Thm. 9.1, p. 52). □

Theorem B.40. *(Expectation Rule) Consider the random variable X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, that takes on values in (E, ε). Let $g : (E, \varepsilon) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ be a measurable function.*

1. *$h(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ if and only if $h(X) \in \mathcal{L}^1(E, \varepsilon, \mathbb{P})$.*

2. *If the function h is positive, or if the function h satisfies the condition in (1), then we have*
$$\mathbb{E}[h(X)] = \int h(x)\mathbb{P}(dx). \tag{B.4}$$

Proof. We have defined the distribution measure $\mathbb{P}$ as $\mathbb{P}(A) = \mathbb{P}(X^{-1}(A))$. Therefore we have

$$\mathbb{E}[\mathbb{I}_A(X)] = \mathbb{P}(X^{-1}(A)) \tag{B.5}$$
$$= \mathbb{P}(A) \tag{B.6}$$
$$= \int \mathbb{I}_A(x)\mathbb{P}(dx). \tag{B.7}$$

If h is a simple function, then equation (B.4) holds by (B.5) and linearity. If the function h is positive, we use the Monotone Convergence Theorem to prove the expectation of the function of the random variable. Let h be positive, and let $\{h_n\}_{n\in\mathbb{N}}$ be a sequence of simple, positive functions converging to h. We then have

$$\begin{aligned}\mathbb{E}[h(X)] &= \mathbb{E}[\lim_{n\to\infty} h_n(X)] \\ &= \lim_{n\to\infty} \mathbb{E}[h_n(X)] \\ &= \lim_{n\to\infty} \int h_n(x)\mathbb{P}(dx) \\ &= \int \lim_{n\to\infty} h_n(x)\mathbb{P}(dx) \\ &= \int h(x)\mathbb{P}(dx).\end{aligned}$$

This proves the second part of the theorem; when the function h is positive. If we apply this to the absolute value of the function $|h|$, and recall that a random variable is in $\mathcal{L}^1$ if and only if it has a finite expectation of its absolute value, then we have proved (1).

For a function h that is not positive, take the expansion of h, $h = h^+ - h^-$. The result can then be obtained by subtraction. □

Corollary B.41. *(Expectation Rule) Consider the random variable X with a density function f, such that*

$$F(x) = \int_{-\infty}^{\infty} f(u)du, \quad -\infty < x < \infty$$
$$F(x) = \mathbb{P}(X \leq x)$$

If $\mathbb{E}[|h(X)|] < \infty$ or if h is a positive function, then we have

$$\mathbb{E}[h(X)] = \int h(x)f(x)dx.$$

Proof. See (Jacod and Protter; 2003, Corollary. 11.1, p. 80). □

B.6 Lebesgue Measures and Probability Distributions

Consider a probability measure $\mathbb{P}$ defined on the space $\mathbb{R}, \mathfrak{B}(\mathbb{R})$. This probability measure is fully characterised by its distribution function F;

$$F(x) = \mathbb{P}((-\infty, x]).$$

Definition B.42. We define the *Lebesgue measure* as the set function $m : \mathfrak{B} \to [0, \infty]$, satisfying the following conditions;

1. for the sequence of pairwise disjoint Borel sets $A_1, A_2, \ldots$, we have *countable additivity*;

$$m\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i),$$

2. if we have $a, b \in \mathbb{R}$, and $a < b$, then the Lebesgue measure $m((a, b]) = b - a$.

Theorem B.43. *The Lebesgue measure exists, and is unique.*

Proof. See Jacod and Protter (2003, Theorem 11.1 and Theorem 11.2, pp. 77, 78). □

The integration theory given above holds for Lebesgue measures. For a Borel measurable function f, if f is integrable for Lebesgue measure, we write the integral as $\int f(x)dx$. The function f is integrable if $\int f^+(x)dx < \infty$ and $\int f^-(x)dx < \infty$, with $f = f^+ - f^-$. The Lebesgue integral is more general than the Riemann integral, and exists more generally. There integrals are equal when they both exist.

Definition B.44. For a probability measure $\mathbb{P}$ defined on the space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, the *density function* is a Borel measurable function f that is positive, satisfying

$$\mathbb{P}((-\infty, x]) = \int_{-\infty}^{x} f(y)dy = \int f(y)\mathbb{I}_{-\infty,x}(y)dy. \tag{B.8}$$

If the distribution measure of a random variable X is given by the probability measure $\mathbb{P}$, then f is the density function of X.

Remark B.45. It is worth noting that not all probability measures $\mathbb{P}$ that are defined on $\mathbb{R}, \mathfrak{B}(\mathbb{R})$ have densities. Equation (B.8) is contingent on the continuity of the distribution function F; F is not always continuous. Even so, there are distribution functions which are continuous, for which no density exists for the corresponding probability measure.

Theorem B.46. *A function f on the real numbers $\mathbb{R}$ that is non-negative Borel measurable is the density function of a probability measure $\mathbb{P}$ on the measurable space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ if and only if the function integrates to one over its domain*

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Where this function exists, it characterizes the probability measure $\mathbb{P}$ entirely. Any other function f' which is non-negative and Borel measurable, where $m(f' \neq f) = 0$, is a density function for the same probability measure.

We also have the result that where a probability measure $\mathbb{P}$ on the measurable space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ has a density function, the density is determined up to a set of Lebesgue measure zero; if f and f' are both density functions for the probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, then we have $m(f' \neq f) = 0$.

Proof. See Jacod and Protter (2003, Theorem 11.3, p. 78). □

Remark B.47. We have $F(x) = \int_{-\infty}^{x} f(y)dy$ for the density function and the distribution function. At every point x where the density function $f(x)$ is continuous, the distribution function is differentiable, and the distribution function and the distribution function satisfy $F'(x) = f(x)$. If it is the case that F is piecewise differentiable, then we have $f(x) = \frac{dF(x)}{dx}$ where this exists, and $f(x) = 0$ everywhere else.

We now state the expectation rule defined with respect to Lebesgue measures.

Corollary B.48. *(Expectation Rule) Consider a random variable X that takes on values in $\mathbb{R}$, with a density function f. Define a Borel measurable function g that is positive. The function g admits an integral with respect to the probability measure $\mathbb{P}$ if and only if the product of this function and the density function is integrable with respect to a Lebesgue measure. The following then holds*

$$\mathbb{E}[g(X)] = \int g(x)\mathbb{P}(dx) = \int g(x)f(x)dx.$$

Proof. See (Jacod and Protter; 2003, Corollary. 11.1, p. 80). □

B.7 Conditional Expectations

This section describes the mathematics relating to the expectation of a random variable defined on a probability space, given that we know that some other event has occurred. This section is not an exhaustive treatment of the topic, but rather covers enough information as is required for the purpose of this dissertation.

Definition B.49. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where the event $A \in \mathcal{F}$ such that $\mathbb{P}(A) > 0$. The *conditional probability* of an event $B \in \mathcal{F}$ with respect to A is denoted as $\mathbb{P}(B|A)$, and is defined as

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Now consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with the σ-algebra $\mathcal{G} \in \mathcal{F}$; that is, $\mathcal{G}$ is a *sub σ-algebra* of the σ-algebra $\mathcal{F}$. We now define conditional expectations with respect to a σ-algebra.

Definition B.50. For a non-negative random variable X, the *conditional expectation of X with respect to the σ-algebra $\mathcal{G}$* is a random variable that satisfies

1. $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable,
2. for all events $A \in \mathcal{G}$,
$$\int_A X\mathbb{P}(\omega)d\omega = \int_A \mathbb{E}[X|\mathcal{G}]\mathbb{P}(\omega)d\omega.$$

We can extend the definition of the conditional expectation from non-negative random variables to all random variables. If we have

$$\min(\mathbb{E}[X^+|\mathcal{G}], \mathbb{E}[X^-|\mathcal{G}]) < \infty,$$

almost surely, for a random variable X, then the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ exists. It is then given by

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}].$$

The derivation of the formulae stated rely on results from the Radon-Nikodým theorem, and are outside the scope of this document. The interested reader is referred to Shiryaev and Boas (1995, Definition 1, p. 213) for the derivation of the conditional expectation.

Definition B.51. The *conditional probability of an event $B \in \mathcal{F}$ with respect to a σ-algebra $\mathcal{G}$, $\mathcal{G} \in \mathcal{F}$,* is defined as the conditional expectation $\mathbb{E}[\mathbb{I}_B|\mathcal{G}]$ and is denoted $\mathbb{P}(B|\mathcal{G})$.

We now state properties of conditional expectations. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X and Y be random variables on this space, and let k be a constant. Let the conditional expectation be defined for all random variables considered, and let σ-algebra $\mathcal{G}$ be a subset of σ-algebra $\mathcal{F}$.

1. If $X = k$ a.s., then $\mathbb{E}[X|\mathcal{G}] = k$ a.s.
2. If $X \leq Y$ a.s., then $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$ a.s.
3. $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ a.s.
4. For the constants $a, b \in \mathbb{R}$, we have
$$\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}], \quad \text{a.s.}$$

5. For the trivial σ-algebra $\mathcal{F}^* = \{\emptyset, \Omega\}$, we have

$$\mathbb{E}[X|\mathcal{F}^*] = \mathbb{E}[X], \quad \text{a.s.}$$

6. $\mathbb{E}[X|\mathcal{F}] = X$, a.s.

7. $\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\right] = \mathbb{E}[X]$

8. If we have $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then
$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1\right] = \mathbb{E}[X|\mathcal{G}_1], \quad \text{a.s.}$$

9. If we have $\mathcal{G}_1 \supseteq \mathcal{G}_2$, then
$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1\right] = \mathbb{E}[X|\mathcal{G}_2], \quad \text{a.s.}$$

10. If the random variable X is independent of the σ-algebra $\mathcal{G}$, then

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X], \quad \text{a.s.}$$

11. If Y is a $\mathcal{G}$-measurable random variable, and we have $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|XY|] < \infty$, then

$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}], \quad \text{a.s.}$$

Proof. See Shiryaev and Boas (1995, Properties of conditional expectations, p. 215). □

Theorem B.52. *(Monotone Convergence Theorem) Consider a sequence of random variables $\{Y_n\}_{n\in\mathbb{N}}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. If we have $Y_n \geq 0$, for all $Y_n, n \in \mathbb{N}$, and $\{Y_n\}$ converges to Y a.s., then*

$$\lim_{n\to\infty} \mathbb{E}[Y_n|\mathcal{G}] = \mathbb{E}[Y|\mathcal{G}], \quad \text{a.s.}$$

Proof. See Jacod and Protter (2003, Theorem 23.8 (a), p. 204). □

Theorem B.53. *(Jensen's Inequality) Consider the convex function $\varphi : \mathbb{R} \to \mathbb{R}$. Let X and $\varphi(X)$ be random variables that are integrable, and let $\mathcal{G}$ be a σ-algebra. We then have*

$$\varphi([X|\mathcal{G}]) \leq [\varphi(X)|\mathcal{G}].$$

Proof. See Jacod and Protter (2003, Theorem 23.9, p. 205). □

Bibliography

Ait-Sahalia, Y. and Lo, J. W. (2000). Nonparametric risk management and implied risk aversion, *Journal of Econometrics* **94**: 9–51.

Anagnou, I., Bedendo, M., Hodges, S. D. and Tompkins, R. (2002). The relation between implied and realised density functions.
URL: *http://ssrn.com/abstract=302280*

Bakshi, G., Kapadia, N. and Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Journal of Financial Review* **16**(1): 101–143.

Barone-Adesi, G., Dallo, H. and Vovchak, V. (2012). Is the pricing kernel monotone?, *ACRN Journal of Finance and Risk Perspectives* **1**(2): 43–68.

Barone-Adesi, G., L., M. and Shefrin, H. (2012). Behavioral finance and the pricing kernel puzzle: Estimating risk aversion, optimism and overconfidence, *Swiss Finance Institute* .

Björk, T. (2009). *Arbitrage Theory in Continuous Time*, 3rd edn, Oxford University Press.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates, *The Journal of Finance* **59**(1): 407 – 446.

Brown, D. P. and Gibbons, M. R. (1985). A simple econometric approach for utility-based asset pricing models, *Journal of Finance* **40**(2): 359–381.

Campbell, J. Y. (2003). *Handbook of the Economics of Finance: Financial Markets and Asset Pricing*, Vol. 1B, North Holland, chapter 13: Consumption-based Asset Pricing.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1996). *The Econometrics of Financial Markets*, Princeton University Press.

Cochrane, J. H. (2001). *Asset Pricing*, Princeton University Press.

Cvitanic, J. and Zapatero, F. (2004). *Introduction to the Economics and Mathematics of Financial Markets*, The MIT Press.

Derman, E. and Zou, J. (1999). Strike adjusted spread: A new metric for estimating the value of equity options, *Technical report*, Quantitative Strategies Research Note, Goldman Sachs and Co.

Detlefsen, K., Härdle, W. and Moro, R. (2010). Empirical pricing kernels and investor preference, *Mathematical Methods in Economics and Finance* **3**(1): 19–48.

Fishburn, P. C. (1968). Utility theory, *Management Science* **14**(5): 335–378.

Hens, T. and Reichlin, C. (2012). Three solutions to the pricing kernel puzzle, *Review of Finance* **17**(3): 1065–1098.

Ingersoll, J. E. (1987). *Theory of Financial Decision Making*, Rowman & Littlefield Publishers.

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realised returns, *The Review of Financial Studies* **13**(2): 433–451.

Jackwerth, J. C. (2004). Option-implied risk-neutral distributions and risk aversion, *CFA Institute Research Foundation of AIMR Publications* pp. 1–86.

Jacod, J. and Protter, P. (2003). *Probability Essentials*, 2nd edn, Springer-Verlag.

Jaffray, J. (1975). The existence of a continuous utility function: An elementary proof, *Econometrica* **43**(5/6): 981–983.

LeRoy, S. F. and Werner, J. (2000). *Principles of Financial Economics*, Cambridge University Press.

Liu, X., Shackleton, M. B., Taylor, S. J. and Xu, X. (2004). Closed-form transformations from risk-neutral to real-world distributions.
URL: *http://ssrn.com/abstract=424420*

Pratt, J. W. (1964). Risk aversion in the small and large, *Econometrica* **32**: 122–136.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels, *Journal of Financial Economics* **64**(3): 341 – 372.

Ross, S. M. and Pekoz, E. A. (2007). *A Second Course in Probability*, www.ProbabilityBookstore.com, Boston, MA.

Shiryaev, A. N. and Boas, R. P. (1995). *Probability*, 2nd edn, Springer.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation, *The Journal of Finance* **51**(5): 1633–1651.

Vestrup, E. M. (2003). *The Theory of Measures and Integration*, Wiley Series in Probability and Statistics (Book 434), Wiley-Interscience.