UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Modelling of financial risk using forward-looking distributions derived from contingent claims

by

VAUGHAN VAN APPEL

THESIS

submitted in fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

in

ACTUARIAL SCIENCE

in the

FACULTY OF NATURAL & AGRICULTURAL SCIENCES

at the

UNIVERSITY OF PRETORIA

PRETORIA

SUPERVISOR: PROF. E. MARÉ

APRIL 2022

I hereby declare that this thesis has not been submitted to any higher education institution, other than the University of Pretoria, for any educational qualification whatsoever.

V van Appel

ACKNOWLEDGEMENTS

Firstly, I would like to extend my deepest gratitude to my supervisor, Prof. Eben Maré, for his unwavering guidance, support and invaluable contribution throughout this journey. I have learnt so much under your supervision, and still hope to learn much more.

Furthermore, I am blessed to have been given the opportunity to register for a PhD at the University of Pretoria in the Department of Actuarial Science, and will forever be grateful.

I would like to thank my family who has supported me throughout all my studies from undergraduate to this point in my academic career, especially my parents, Charles and Jillian; my brothers, Jacques and Grant; and my grandparents, Charlie and Marie. Without your support, this would not have been possible.

Lastly, I wish to express my gratitude to my work colleagues in the Department of Statistics at the University of Johannesburg for all your support and encouragement throughout this journey.

ABSTRACT

In this thesis, we investigate several methods for extracting the forecast distribution from historical asset returns and market-quoted option prices. Typically, risk-neutral distributions, extracted from market quoted option prices, are considered biased estimates of the forecast distribution, and therefore need to be transformed into a real-world distribution. Transformation processes often require the use of historical data and restrictive assumptions on a representative investor. Alternatively, the recovery theorem provides a theoretically appealing method to recover the real-world distribution from the risk-neutral transition probability matrix without the use of historical returns. However, estimating the risk-neutral transition probability matrix has proven to be a challenging task, as it involves solving an ill-posed problem. Therefore, we propose a regularised multivariate Markov chain in the estimation of the risk-neutral transition probability matrix to obtain a more accurate real-world forecast distribution than obtained using the univariate model.

Comparative studies on the accuracy of real-world forecast distributions are scarce in the literature. Therefore, we further backtested and compared the accuracy of the extracted distributions on the South African Top 40 index, where we found that the forward-looking real-world distribution improved forecasting in certain situations. We also proposed a forward-looking mixture model of historical and option-implied distributions to improve forecasting.

Furthermore, we implemented the extracted forecast distributions in determining safe retirement withdrawal rates. In our empirical study, we showed that the use of forward-looking distributions drastically improved the success in retirement withdrawal rates.

PUBLICATIONS FROM THIS THESIS

- Van Appel and Maré (2018). The Ross recovery theorem with a regularised multivariate Markov chain. *ORiON*, **34**(2): 133-155.
- Van Appel and Maré (2020b). The recovery theorem with application to risk management. *South African Statistical Journal*, **54**(1): 65-91.
- Van Appel and Maré (2020a). Die herwinningstelling met toepassing op risikobestuur. *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie*, **39**(1): 133.
- Van Appel and Maré (2021). Estimation of forecast distributions: A comparison of risk-neutral and real-world distributions in the South-African market. *Working Paper*.
- Van Appel et al. (2021). Quantitative guidelines for retiring (more safely) in South Africa. *South African Actuarial Journal*, **21**: 75-91.
- Van Appel and Maré (2022). Determining safe retirement withdrawal rates using forward-looking distributions. *South African Journal of Science*, **118**(3/4): 38-44.

CONTENTS

List of Figures

List of Tables

NOTATION AND ABBREVIATIONS

Ω	Sample space.
$\mathbb{P}$	Real-world probability measure.
$\mathbb{Q}$	Risk-neutral probability measure.
$\mathbb{P}(X)$	Probability of an event X.
$f_X(x)$	Probability density function of the random variable X.
$F_X(x)$	Cumulative distribution function.
$F_n(\cdot)$	Empirical distribution function of a sample consisting of n observations.
$\sim N(\mu, \sigma^2)$	Distributed normally with mean μ and variance σ^2.
$\phi(\cdot)$	Standard normal density function, i.e., $\sim N(0,1)$.
$\Phi(\cdot)$	Cumulative standard normal distribution function.
$\sim \chi^2(\nu)$	Distributed chi-squared with ν degrees of freedom.
W_t	Standard Brownian motion.
N_t	Poisson process counting the number of jumps.
$\mathbb{I}_{\{X\}}$	Indicator function. If the event, X, is true, $\mathbb{I}_{\{X\}} = 1$, otherwise $\mathbb{I}_{\{X\}} = 0$.
$\mathbb{R}$	Real numbers.
$\mathbb{R}^+$	Strictly positive real numbers.
$\mathbb{E}(X)$	Expected value of a random variable X.
$X^\top$	Transpose of matrix X.

X^{-1}	Inverse of matrix X.
pdf	Probability density function.
cdf	Cumulative distribution function.
i.i.d.	Independent and identically distributed.
$\ell(\theta)$	log-likelihood function.
LR	Likelihood ratio test statistic.
$\hat{\theta}$	Estimator for θ.
$\|\|\cdot\|\|_1$	Absolute-value (L_1) norm.
$\|\|\cdot\|\|_2$	Euclidean (L_2) norm.
SE	Standard error.
VaR	Value-at-Risk.
$CVaR$	Conditional Value-at-Risk.
TAA	Tactical asset allocation.
Top40	South African FTSE/JSE Top 40 index.

CHAPTER 1

INTRODUCTION

An important part of financial modelling is to forecast the possible return of an underlying asset. This is arguably one of the most important and fascinating topics for many researchers and financial practitioners. However, forecasting an accurate and meaningful return distribution is a challenging task.

A common approach used in the modelling of future asset returns is to use estimates based on historical data, such as, the mean and standard deviation, which under the assumption that returns are normally distributed for example, provide a complete specified return distribution for modelling purposes (see, e.g., Alexander, 2008; Hull, 2006). However, financial markets are typically quite volatile and estimates based on historical data are unstable and are unlikely to capture future asset returns accurately (see, e.g., de Vincent-Humphreys and Noss, 2012; Cont, 2001). This is particularly evident in the higher moments of the estimated distribution, which has become ever more important in forecasting future returns and assessing risk (see, e.g., Christoffersen et al., 2013; Conrad et al., 2013).

On the other hand, financial modelling is often emphasised by the pricing of contingent claims (derivative securities[1]) whose payoff extends out in time (see, Kienitz and Wetterau, 2012; Venter, 2010). Therefore, financial derivative securities are forward-looking and essentially embed information about investors' beliefs about the distribution of asset returns (see, e.g., Christoffersen et al., 2013; Hollstein et al., 2019; Dillschneider and Maurer, 2019). For this reason, financial derivative securities are frequently used to infer information. In particular, it is found that the forecasted risk-neutral higher moments (volatility, skewness and kurtosis) were strongly related to the subsequent returns (see, e.g., Christoffersen et al., 2013; Conrad et al., 2013; Dobiáš, 2007). However, such risk-neutral distributions do not provide a true re-

[1] We will use derivative securities interchangeably with contingent claims throughout this thesis.

flection of the actual probabilities assigned to certain outcomes, as the risk-neutral distribution is obtained to recover the prices of traded options in a way that avoids arbitrage. In essence, investors are risk-averse in aggregate, which leads to differences in the probabilities extracted from option prices to that of the true real-world probabilities (see, e.g., de Vincent-Humphreys and Noss, 2012; Cuesdeanu and Jackwerth, 2018a). Essentially, risk-averse investors typically assign higher probabilities to unfavourable scenarios than to favourable scenarios, and therefore over-weigh the unfavourable scenarios when deciding on the securities future value. However, an important part of modelling is to generate accurate real-world scenarios for risk and portfolio management. It is important that these scenarios closely represent economic conditions.Since derivatives are priced under the risk-neutral measure, where risk-management mostly utilises the real-world measure for forecasting, it is important that we develop techniques for extracting the forward-looking real-world distribution.[2]

While there is often a rich market of traded equity derivative prices available, and an extensive amount of literature devoted to extracting the risk-neutral probabilities from option prices (see, e.g., Breeden and Litzenberger, 1978; Aït-Sahalia and Lo, 1998; Kienitz and Wetterau, 2012; Christoffersen et al., 2013; Malz, 2014; Barone-Adesi, 2016; Crisóstomo and Couso, 2018), the use of these probabilities to forecast future returns has been hampered by the transformation of these probabilities into real-world probabilities. Real-world probabilities offer numerous advantages over their risk-neutral counterparts (see, e.g., de Vincent-Humphreys and Noss, 2012; Shackleton et al., 2010). These advantages range from (i) gaining insight into market participants actual views on future asset prices, and (ii) the comparison of the estimated risk-neutral and real-world probabilities reveals useful information about investors risk preferences. However, it is typically difficult to recover the real-world distribution of asset returns from the risk-neutral distribution, which often relies on making use of historical data and restrictive assumptions on the representative investor (see, e.g., Cuesdeanu and Jackwerth, 2018b; Bliss and Panigirtzoglou, 2004; Aït-Sahalia and Lo, 2000; Shackleton et al., 2010).

Recently, Ross (2015) described a method for extracting the real-world distribution using only risk-neutral information. This is achieved by obtaining the richer risk-neutral transition probabilities, as opposed to only the risk-neutral distribution. The risk-neutral transition probabilities represent the risk-neutral probabilities of moving from all hypothetical current states

[2]Other terms often used to describe real-world measure in the literature are: risk-adjusted, actual probability measure, objective probability measure, physical distribution or subjective distribution.

to all future states. This approach, under strict market conditions, is very appealing from a theoretical point of view. However, implementing this approach requires solving two ill-posed problems. The first, involves finding the forward-looking state price matrix and the second involves estimating the risk-neutral transition probability matrix.

The ability to accurately forecast the asset return distribution is an important and valuable contribution that can be used in many applications. For example, risk managers use the forecasted return distribution to measure the market risk, and portfolio managers often use the forecast return distribution to make optimal portfolio choices. In what follows in this thesis is the study of extracting historical, risk-neutral and real-world return distributions from market prices. Since, investors, risk managers, actuaries and policy makers are often required to model and forecast future returns for making optimal decisions, one needs to select an appropriate model for the problem at hand. In testing the appropriateness of the models, we also implement and backtest the forecast models on the South African FTSE/JSE Top 40 (Top40) index. Thereafter, these forecast models are applied to applications in financial modelling, such as, determining Value-at-Risk forecasts and determining safe retirement withdrawal rates.

Our empirical findings based on the Top40 index return series indicated that models that used real-world forward-looking information produced better Value-at-Risk forecasts, especially during economic uncertain times, than models based on historical data. Furthermore, we also found that using the estimated moments of the real-world forward-looking distribution in a tactical asset allocation framework yielded superior portfolio returns when compared to models based on historical returns, which is vitally important in obtaining safe withdrawal rates for retirees.

1.1 Research Questions

In order to build a framework for making optimal investment and risk-management decisions, it is necessary to study the statistical properties of how the underlying asset price evolves. Therefore, the following three problems form the central aim of this thesis:

Table 1.1: Research Questions.

	Research Question	Name of paper
1.	How are forecasting distributions extracted from market prices?	Van Appel and Maré (2018) Van Appel and Maré (2021)
2.	Which forecasting distribution model is the most appropriate for forecasting the South African equity market dynamics?	Van Appel and Maré (2020a) Van Appel and Maré (2020b)
3.	Do forward-looking forecast distributions provide meaningful information for decision making in risk management, portfolio optimisation and investment strategies?	Van Appel et al. (2021) Van Appel and Maré (2022)

1.2 Thesis Contribution and Layout

Part I: The Building of Forecast Distributions

The first part of this thesis is focused on the first two research questions in Table 1.1. In particular, forecasts of the entire return probability distribution is becoming increasingly popular among many financial practitioners (see, e.g., Cristóstomo, 2021; Crisóstomo and Couso, 2018; Ross, 2015; Christoffersen et al., 2013; de Vincent-Humphreys and Noss, 2012; Shackleton et al., 2010). That is, financial practitioners are not only concerned about forecasting point estimators, such as, the expected return, but also the uncertainty around the asset return. Essentially, a forecast distribution captures all of the random variable characteristics, including the tail, which is important for risk managers. Quantifying this uncertainty provides valuable information pertaining to the risk of a financial portfolio. Therefore, extracting accurate forecast distributions are vitally important for many risk management activities.

This part is organised as follows:

In Chapter 2, we provide a brief overview of commonly used methods to estimate the historical and forward-looking risk-neutral distribution from market prices. Furthermore, we also summarise methods used to transform the risk-neutral distribution into a real-world distribution (see, Van Appel and Maré, 2021). These transformation methods will make use of historical

data and restrictive assumptions about investor preference. However, using historical data along with restrictive assumptions about investor preference in the transformation process is likely to introduce some bias in the estimation of the real-world distribution.

In Chapter 3, we study the recovery theorem introduced by Ross (2015). Although this approach is theoretically very appealing, as it only makes use of risk-neutral information, it is hampered by strict market constraints and solving an ill-posed problem in estimating the risk-neutral transition matrix. However, in this chapter, we propose solving this ill-posed problem by stabilising the estimation of the risk-neutral transition probability matrix by implementing a regularised multivariate Markov chain (see, Van Appel and Maré, 2018). In addition, we carry out a comparison of methods and a robustness check to show the effectiveness of this method.

In Chapter 4, we extract, backtest and compare the weekly one-month historical, risk-neutral and real-world forecast distributions using the Top40 index through a series of distribution forecast backtests (see, Van Appel and Maré, 2020b). Comparative studies on the accuracy of real-world forecasts are scarce in the literature, and therefore is of great value. In particular, the real-world forecast distribution has many applications in finance, especially in risk management, where real-world probabilities are preferred. Therefore, it is important to obtain as accurate forecast distributions as possible. Furthermore, since the tail of the forecast distribution is vitally important for risk management, we backtest and compare the tail of the forecast distributions. It is particularly unlikely that a single forecast model will be consistently superior throughout all time periods as the models can be sensitive to the sample period, asset type, and the forecasting horizon (see, e.g., Shackleton et al., 2010). Therefore, in order to obtain more stable and accurate forecasts, we also propose the mixing of historical and forward-looking distributions using a forward-looking mixing parameter.

Part II: Modelling using the Forecast Distribution

The second part of this thesis is focused on the third research question in Table 1.1, where we show a novel application of using the forecast distribution in determining safe retirement withdrawal rates.

The topic of modelling safe retirement withdrawal rates is of considerable value to the

actuarial profession and larger community, as we only get one chance to save for retirement. In general, there are two risk factors to consider, namely, market risk and longevity risk. Market risk focusses on the risk of losses caused by the changes is the underlying asset values in a portfolio, where longevity risk is the uncertainty around the lifespan over which we need to provide portfolio withdrawals for. It it well-known that people are living longer and therefore, a retirement withdrawal strategy needs to last at-least 30-years, with substantial probability, to be viable.

The models that we will build in this part, include transaction fees and longevity. In particular, we will show that different withdrawal rates and asset allocations can drastically affect the outcome for a retiree, and therefore it is vitally important to devise optimal withdrawal strategies.

The aim for a retiree is to draw as much money for as long as possible. This is none other than an optimisation problem. Obviously, earning higher returns increases the retirement portfolio's success rates. To achieve this, Scott (1996) suggested that retirees should increase their equity allocations in their retirement portfolio. Consequently, this strategy also increases the volatility in the portfolio returns and time to ruin. Both of these properties are undesirable for retirees. Therefore, we propose a strategy of using the moments of the forward-looking forecast distribution in a tactical asset allocation framework to increase the portfolio return without increasing the volatility in returns. This strategy allows for the possibility of achieving higher success rates for retirement portfolios.

This part is organised as follows:

In Chapter 5, we consider a problem with significant practical value. In particular, we study safe retirement fund withdrawal rates for retirees using historical data in a bootstrap simulation (see, Van Appel et al., 2021). This corresponds to typical studies found in the literature (see, e.g., Cooley et al., 1998, 1999; Maré, 2016). However, we incorporate transactional costs and longevity into the simulation for a more realistic determination of the portfolio success rates. The advantage of this modelling approach is that the correlation structure between the asset classes is captured; however, the inter-temporal links such as the serial correlation is lost. Furthermore, these traditional methods for determining safe withdrawal rates do not adequately describe accurate return scenarios, as the assumption that future returns are like past returns is

not necessarily valid.

In Chapter 6, we consider an extension of Chapter 5 by using forward-looking forecast distributions to determine safe retirement withdrawal rates (see, Van Appel and Maré, 2022). More specifically, we demonstrate empirically that using the moments of the forward-looking real-world distribution in a tactical asset allocation framework achieved higher equity returns, and therefore allows for higher safe spending rates.

Part I

The Building of Forecast Distributions

CHAPTER 2

ESTIMATION OF FORECAST DISTRIBUTIONS

This chapter forms part of the working paper Van Appel and Maré (2021) and contributes partly towards our first research question.

Chapter Synopsis

Estimating the forecast distribution typically involves time-series of asset returns or extracting the risk-neutral distribution from market-quoted option prices. However, attention has recently shifted from time-series and risk-neutral forecast distributions to real-world distributions. While there is a vast literature on methods to extract the forward-looking risk-neutral distribution from option prices (see, e.g., Crisóstomo and Couso, 2018; Malz, 2014; Christoffersen et al., 2013), there are fewer studies in transforming the risk-neutral distribution into a real-world distribution, with even fewer studies investigating the accuracy of the real-world forecasts.

Estimating the real-world forecast distribution typically involves a utility-function transform (see, Bliss and Panigirtzoglou, 2004) or by an empirical study that searches for a calibration function that best fits the risk-neutral distribution obtained from a given set of observed future values (see, Liu et al., 2007; de Vincent-Humphreys and Noss, 2012). More recent developments extended the utility-function transform approach by incorporating a behavioural framework that uses investor sentiment to obtain a more accurate transformation from the risk-neutral distribution to the real-world distribution (see, Cristóstomo, 2021). The aim of this chapter is to study such transformation methods, which will then be applied and compared in the South African market in Chapter 4.

2.1 Introduction

The purpose of this chapter is to study some commonly used methods for extracting the forecast distribution of the underlying future asset price S_T (or return). In particular, we will focus on the two types of forecast distributions used in practice, namely, the real-world distribution and the risk-neutral distribution. These methods will be implemented using the South African FTSE/JSE Top 40 (Top40) index data. The forecasts will then be backtested and compared in Chapter 4. Extracting the entire forecast distribution rather than a single forecast measure, such as the expected future asset price, has numerous advantages in practice. For example, in risk management and portfolio asset allocation, the percentiles and higher moments are extremely valuable (see, e.g., Aït-Sahalia and Lo, 1998; Dobiáš, 2007; de Vincent-Humphreys and Noss, 2012; Audrino et al., 2014; Barone-Adesi, 2016; Flint and Maré, 2017; Van Appel and Maré, 2020b).

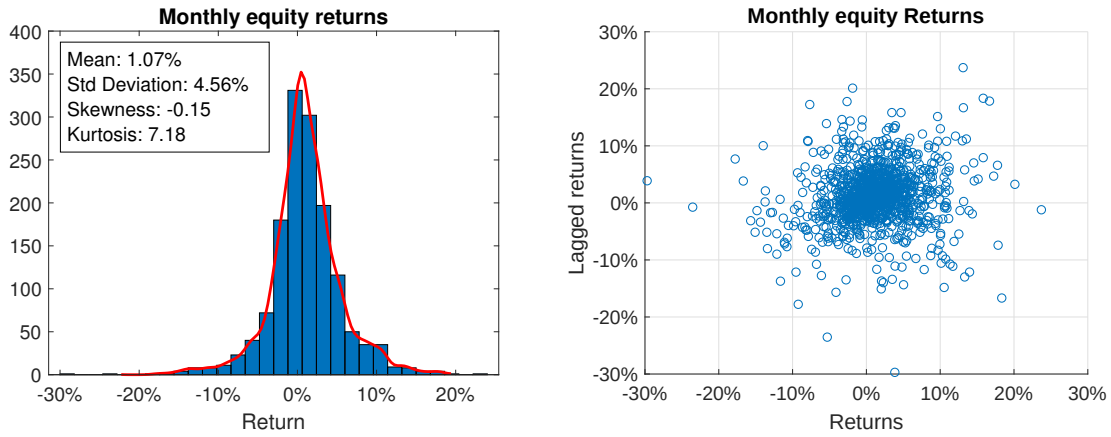
We begin this study by firstly defining the real-world distribution extracted from historical data. This is often referred to as a backward-looking distribution. Thereafter, we study methods for extracting the forward-looking distribution from European call option prices. These distributions are classified as risk-neutral distributions. Lastly, we consider methods to transform the forward-looking risk-neutral distributions into real-world distributions.

2.2 The Real-World Distribution

The real-world distribution is the distribution typically obtained from time-series analysis (see, e.g., Kienitz and Wetterau, 2012, p. 16). For example, suppose that we have identified a market invariant[1], such as the returns of an index. The historical distribution of this market invariant is often used as an input into some financial model, for example, in calculating measures of risk (see, e.g., Meucci, 2005). In Figure 2.1, we illustrate such a historical real-world distribution for a South African index over the period from 1900 to 2020.[2]

[1] Moreover market invariants can be found in Meucci (2005)

[2] Data sourced from Firer and McLeod (1999), Firer and Staunton (2002) and I-Net Bridge.



(a) Equity returns

(b) Lagged time series for the returns

Figure 2.1: Monthly returns for a South African index from 1900 to 2020.

From the moments of the distribution shown in Figure 2.1a, we observe that the returns do not follow a Gaussian distribution, which is commonly assumed in many financial models. However, Figure 2.1b suggests that the index returns are invariant (see, Meucci, 2005). Therefore, it is often better to apply probabilistic concepts on the returns rather than the realised prices of the index.

Typically, securities have return distributions with so-called skewness and fat tails. Moreover, Cont (2001) presented some statistical stylised facts, which emerged in empirical studies of most asset returns, where he showed that it is particularly difficult to reproduce many of these properties with a parametric model, requiring at least four parameters in the return distribution (i.e., a location parameter, scale parameter, a parameter describing the decay of the tails, and an asymmetry parameter to allow different behaviours between the tails). This occurrence has persuaded many risk managers to use historical simulation, rather than using a parametric model building approach, to extract the return distribution (see, e.g., Pérignon and Smith, 2010).

2.2.1 Historical Simulation

Historical simulation involves creating a database of the daily/weekly/monthly change in the asset over a period of time. For example, suppose that we have recorded 100 days of daily returns and we are interested in the 5th percentile of the daily return density (i.e., $\text{VaR}_{(0.95)}$).

This would correspond to the 5th worst change out of the 100 days of asset value returns. This method will be referred to as the historical simulation method in this thesis. The first drawback in using past data for simulation is that the forecast density will be slow to react to market shifts. Therefore, Hull and White (1998a) incorporated a volatility updating approach to adjust the historical database using the following exponential weighted moving average (EWMA) model:

$$\sigma_t^2 = \alpha\sigma_{t-1}^2 + (1-\alpha)R_{t-1}^2 \tag{2.1}$$

with $\alpha = 0.94$. Hence, suppose we are estimating the return density at time $N-1$ for time N. Let:

R_t : be the historical return on day t for the period covered by the historical sample ($t < N$)

$\hat{\sigma}_t^2$: be the historical EWMA estimate of the variance of the return for period t forecasted at the end of period $t-1$

$\hat{\sigma}_N^2$: is the EWMA estimate of the variance for period N. This estimate is made at the end of period $N-1$.

In the Hull and White (1998a) volatility updating approach, the original historical return data, R_t, are adjusted by multiplying the historical return data by the ratio of the current volatility to the volatility at the time of the observation, that is,

$$R_t^* = R_t \frac{\hat{\sigma}_N}{\hat{\sigma}_t}. \tag{2.2}$$

This method will be referred to as the historical-HW method for the remainder of this thesis. The second drawback with historical simulation methods is that there may not be enough data available in the market to form a historical database, especially for new securities.

Since historical simulation requires a large database of past returns, which possibly consists of returns when the market was in a different economic environment, we next consider using a forward-looking approach to extract the return density forecast.[3]

[3]These concerns are also reported in Bliss and Panigirtzoglou (2004) and Cuesdeanu and Jackwerth (2018a) about using historical data.

2.3 Forward-Looking Densities

This method uses derivative prices to extract the probability density function of the future asset price S_T from European call option prices. In particular, a European-style call option is a contract that gives the holder the right (but not the obligation) to purchase a prescribed asset from the writer, for a prescribed price at a prescribed time, in the future. Therefore, the price of a call option is determined by its expected future pay-off discounted by the risk-free interest rate (see, e.g., Hull, 2006; Kienitz and Wetterau, 2012; Björk, 2009):

$$\mathrm{C}(T,K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[\max(S_T - K)] = e^{-rT}\int_K^{\infty}(x-K)f_{S_T}^{\mathbb{Q}}(x)\,\mathrm{d}x, \tag{2.3}$$

where T denotes the time to expiry (tenor), K the predetermined strike price, r the risk-free interest rate, x the asset price at expiry, and $f_{S_T}^{\mathbb{Q}}(x)$ the probability density function of the future asset price under the risk-neutral measure $\mathbb{Q}$. These options are typically traded on a number of official exchanges.

Since the option pay-off extends out in time, it is well-known that option prices capture some market sentiment based on supply and demand considerations (see, e.g., Christoffersen et al., 2013; Hollstein et al., 2019; Dillschneider and Maurer, 2019; Shackleton et al., 2010; Crisóstomo and Couso, 2018). This is known as forward-looking information. Christoffersen et al. (2013) provided a description of situations when option-implied forecasts are likely to be most useful, such as, when the option market is highly liquid.

There are various methods of going about extracting the risk-neutral density, $f_{S_T}^{\mathbb{Q}}(x)$ from option prices.[4] In this study, we will consider two such approaches. The first is based on making use of a continuous stochastic diffusion processes, such as the Black-Scholes, Heston and Bates models, to model the underlying stock process and the second is based on the model-free approach proposed by Breeden and Litzenberger (1978).

[4] Although, in this thesis, we extract the risk-neutral density from call option prices, we can extract the risk-neutral density from put option prices (see, e.g., Barone-Adesi, 2016) or by making use of the put-call parity.

2.3.1 The Black-Scholes Model

Black and Scholes (1973) assumed the following real-world stock price process:

$$\mathrm{d}S_t = (r+\mu)S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t, \quad 0 \leq t \leq T, \tag{2.4}$$

where r denotes the risk-free rate, μ the equity risk premium, σ the volatility of the stock, and W_t a standard Brownian motion.

If we let $Y = \log(S_t)$, then by using Ito's lemma on (2.4), the log stock price process is (see, e.g., Christoffersen et al., 2013):

$$\mathrm{d}\log S_t = \left(r+\mu-\frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{2.5}$$

Integrating over (2.5) and solving for S_T yields the solution to the stochastic differential equation in (2.4), i.e.,

$$S_T = S_0 \exp\left[\left(r+\mu-\frac{\sigma^2}{2}\right)T + \sigma W_T\right], \tag{2.6}$$

which implies that the real-world density for S_T, denoted by $f^{\mathbb{P}}_{S_T}(x)$ is log-normally distributed:

$$f^{\mathbb{P}}_{S_T}(x) = \frac{1}{x\sqrt{2\pi\sigma^2 T}} \exp\left[-\frac{1}{2\sigma^2 T}\left\{\log x - \log S_0 - \left(r+\mu-\frac{\sigma^2}{2}\right)T\right\}^2\right]. \tag{2.7}$$

While the risk-premium is not directly observable, it can be estimated, which forms the base of transforming the risk-neutral density into a real-world density. Under the risk-neutral measure, $\mu = 0$, and we have that S_T is also log-normally distributed, but with a different mean. That is,

$$f^{\mathbb{Q}}_{S_T}(x) = \frac{1}{x\sqrt{2\pi\sigma^2 T}} \exp\left[-\frac{1}{2\sigma^2 T}\left\{\log x - \log S_0 - \left(r-\frac{\sigma^2}{2}\right)T\right\}^2\right]. \tag{2.8}$$

In the Black-Scholes framework, the risk-neutral density will have the correct form and volatility, but with the mean typically adjusted downward due to the removal of the equity premium μ. That is, the equity risk premium, μ will be positive if the underlying asset is positively correlated with overall wealth in the economy, implying that the risk-neutral forecasts will be a downward biased forecast of the future asset price (see, e.g., Christoffersen et al., 2013). Since the risk-neutral mean of the asset return is the risk-free rate, option prices provide no predictive power for the expected return. However, higher order moments have proven to

be powerful in determining movements in the asset price (see, e.g., Bates, 1991; Conrad et al., 2013).

Under the risk-neutral measure and the assumption that the future asset price is log-normally distributed, the closed-form solution of (2.3), known as the Black-Scholes option pricing formula, can be found as:

$$\mathrm{C}(T,K,S_0,r,\sigma^2) = S_0\Phi(d) - Ke^{-rT}\Phi\left(d-\sigma\sqrt{T}\right), \tag{2.9}$$

where

$$d = \frac{\log\left(\frac{S_0}{K}\right) + T\left(r+\frac{1}{2}\sigma^2\right)}{\sigma\sqrt{T}}, \tag{2.10}$$

and $\Phi(\cdot)$ represents the cumulative standard normal distribution function. The Black-Scholes option pricing formula only has one unknown parameter, namely the volatility, which we imply from the market-quoted option prices. That is, the implied volatility, denoted by $\sigma^{(\mathrm{IV})}$, is the volatility that one has to substitute into the Black-Scholes option pricing formula to reproduce market option prices. In particular, if one assumes that the given underlying asset is log-normally distributed, then by the Black and Scholes (1973) framework, the implied volatility, $\sigma^{(\mathrm{IV})}$, should be independent of the expiry T and the strike K. Unfortunately, it is well-known that this is not the case. In reality, the implied volatility is found to be dependent on both quantities: $\sigma^{(\mathrm{IV})} = \sigma^{(\mathrm{IV})}(T,K)$. Therefore, as quoted in Rebonato (2002),

> *"Thus implied volatility is the wrong number to put in the wrong formula to get the right price".*

Typically, it is common practice to quote option prices in terms of the implied volatility. This is done as there is a one-to-one relation between option prices and the implied volatility, since there is a large agreement of what the other parameters are to be in the Black-Scholes option pricing formula, such as the risk-free rate. Therefore, the Black-Scholes formula is simply used as a metric to express prices as volatilities. That is, given the call option price, the implied volatility can uniquely be determined and vice versa.

In particular, the implied volatility is not only a property of the underlying, but carries information about the derivative security. Therefore, the option implied volatility will often be

higher than the volatility of the underlying as it incorporates market risk. (see, e.g., Bakshi and Kapadia, 2003).

Due to the forward-looking nature of derivative prices and hence implied volatilities, these measures are often used to encapsulate the markets perception of the underlying asset price in the future (see, e.g., Shackleton et al., 2010; Christoffersen et al., 2013; Ross, 2015; Crisóstomo and Couso, 2018; Hollstein et al., 2019).

Since the volatility is not constant as assumed in the Black-Scholes model, we therefore consider two stochastic volatility models next.

2.3.2 Stochastic Volatility Models

In this section, we consider two stochastic volatility models to extract the risk-neutral density of the future asset price S_T, namely the Heston (1993) model and Bates (1996) model. The risk-neutral dynamics for the Heston model is given by:

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sqrt{V_t}S_t\,\mathrm{d}W_t^{(1)} \tag{2.11}$$

$$\mathrm{d}V_t = \kappa(\theta - V_t)\,\mathrm{d}t + \nu\sqrt{V_t}\,\mathrm{d}W_t^{(2)}, \tag{2.12}$$

where the parameter r represents the risk-free rate, κ models the speed of mean reversion of the variance, θ the long term variance, ν to volatility of variance, ρ the correlation between the two driving Brownian motions $W_t^{(1)}$ and $W_t^{(2)}$, S_0 the spot rate, and V_0 the spot variance.

It is well-known that adding jumps to the spot price process could improve the agreement between theoretical and observed option prices, especially in stressed markets (see, e.g., Crisóstomo and Couso, 2018). Therefore, the Bates model is simply an extension of the Heston model with independent jumps added to the security price dynamics in (2.11), giving the following risk-neutral dynamics:

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sqrt{V_t}S_t\,\mathrm{d}W_t^{(1)} + (Y-1)S_t\,\mathrm{d}N_t \tag{2.13}$$

$$\mathrm{d}V_t = \kappa(\theta - V_t)\,\mathrm{d}t + \nu\sqrt{V_t}\,\mathrm{d}W_t^{(2)}, \tag{2.14}$$

where N_t is a Poisson process, which models the number of jumps with intensity $\lambda > 0$ and Y is the jump size distribution, which, in this case, is a log-normal distribution with average jump size μ_J, and jump volatility σ_J.

Heston (1993) and Bates (1996) provided analytical expressions for the characteristic function of $\log(S_T)$, which is then used to obtain the cumulative distribution function and risk-neutral density function of S_T, denoted by $F^{\mathbb{Q}}_{S_T}(x)$ and $f^{\mathbb{Q}}_{S_T}(x)$ respectively, for positive values of x (see, e.g., Shackleton et al., 2010):

$$F^{\mathbb{Q}}_{S_T}(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^{\infty} \mathrm{Re}\left[\frac{\exp(-iu\log(x))\psi(u)}{iu}\right]\mathrm{d}u \tag{2.15}$$

$$f^{\mathbb{Q}}_{S_T}(x) = \frac{1}{\pi x}\int_0^{\infty} \mathrm{Re}\left[\exp(-iu\log(x))\psi(u)\right]\mathrm{d}u, \tag{2.16}$$

where $\psi(u) = \mathbb{E}\left[\exp(iu\log(S_T))\right]$ denotes the characteristic function of $\log(S_T)$.

Furthermore, using the characteristic function, the closed-form solution for the price of a European call option, $\mathrm{C}(T,K)$, can then be derived on an asset with stochastic volatility (see, e.g., Heston, 1993; Bates, 1996; Kienitz and Wetterau, 2012; Carr and Madan, 1999) and is given in Appendix A for convenience.

One of the main drawbacks with the model-building approaches above is that the unknown model parameters needs to be estimated from the market-quoted call option prices in order to obtain the risk-neutral density function. This processes involves a computationally expensive process of numerically minimising the difference between market prices and model prices.

2.3.3 Calibration Procedure

In this section, we present methods in estimating model parameters for the Black-Scholes, Heston and Bates models. In particular, model parameters are typically derived from quoted option prices.

The calibration procedure estimates the unknown parameters, Θ, by finding the parameter values that minimise the difference (or error) between market prices and model prices. In order to find the optimal parameter set, we firstly quantify three commonly used error functions used in the literature (see, e.g., Kienitz and Wetterau, 2012, p. 435):

1. RMSE (Root-mean-square-error)

$$RMSE(\Theta) = \frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{N}\left(\mathrm{C}_i - \hat{\mathrm{C}}_i(\Theta)\right)^2} \tag{2.17}$$

2. AAE (Average Absolute error)

$$AAE(\Theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left| C_i - \hat{C}_i(\Theta) \right| \tag{2.18}$$

3. ARPE (Average Relative Percentage error)

$$APE(\Theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\left| C_i - \hat{C}_i(\Theta) \right|}{C_i}, \tag{2.19}$$

where N represents the number of quoted option prices, C_i denotes the i-th call option price obtained from the market, and $\hat{C}_i$ the ith model price obtained with parameter set Θ. This procedure relies on the condition that there are enough traded option prices available to perform such a task. Secondly, we need to run an optimisation scheme to determine the parameter values that minimise the error function. Moreover such optimisation schemes can be found in Kienitz and Wetterau (2012, Chapter 9) and are readily available in most statistical software packages. In particular, we found that the RMSE function yielded the best results in the calibration of the Heston and Bates parameters using the Top40 index call prices, and was therefore used in this thesis.

2.3.4 Model Free Risk-Neutral Density

Breeden and Litzenberger (1978) showed that the implied risk-neutral probability density function of the future asset price, S_T can be extracted from a set of European-style option prices. That is, the cumulative distribution function (cdf), denoted by $F^{\mathbb{Q}}_{S_T}(x)$, can be obtained by taking the first order partial derivative of $C(T,K)$ with respect to the strike K, i.e.,

$$\frac{\partial C(T,K)}{\partial K} = -e^{-rT}\left[1 - F^{\mathbb{Q}}_{S_T}(K)\right]. \tag{2.20}$$

Rearranging (2.20), yields an expression for the implied risk-neutral cdf,

$$F^{\mathbb{Q}}_{S_T}(x) = 1 + e^{rT} \frac{\partial C(T,K)}{\partial K}\bigg|_{K=x}. \tag{2.21}$$

Thereafter, the conditional probability density function (pdf) is obtained by taking the partial derivative of (2.21) with respect to K, as follows:

$$f^{\mathbb{Q}}_{S_T}(x) = e^{rT} \frac{\partial^2 C(T,K)}{\partial K^2}\bigg|_{K=x}. \tag{2.22}$$

In practice, a continuum of traded strikes is not directly observed in the markets, especially in South Africa where option price data are sparse and noisy. Therefore, one typically first needs to interpolate and extrapolate call option prices over a dense strike range (see, e.g., Malz, 2014; Flint and Maré, 2017). It is typically more desirable to interpolate and extrapolate over the implied volatility, where the Black-Scholes option pricing formula is used to move between prices and implied volatility. In this study, we used the so-called stochastic volatility inspired (SVI) model to extract a dense implied volatility surface, which is then used to compute a dense set of call option prices across the full strike range for each expiry date (see, e.g., Gatheral and Jacquier, 2014; Flint and Maré, 2017). Thereafter, we numerically approximated the risk neutral distribution by taking the second difference along the interpolate and extrapolate call option price at tenor T (see, e.g., Christoffersen et al., 2013; Figlewski, 2010). The risk-neutral density forecast extracted using this model-free approach will be referred to as the RND model for the remainder of this thesis.

Now that we have obtained estimates of the risk-neutral density, we can proceed with transformation methods to estimate the real-world density. Shackleton et al. (2010) found in an empirical study that the real-world distribution improved forecasting performance for both two- and four-week horizons.

2.3.5 Risk-Premium Transformation

Real-world probabilities differ from risk-neutral probabilities in that investors require a premium that compensates them for carrying risk. The transformation from risk-neutral to real-world densities typically rely on assumptions (see, e.g., Bliss and Panigirtzoglou, 2004; Shackleton et al., 2010; Ross, 2015; Dillschneider and Maurer, 2019; Cuesdeanu and Jackwerth, 2018b). Furthermore, the link between option-implied risk-neutral densities and real-world densities is not unique and a pricing kernel (or stochastic discount factor), M_T, must be assumed to link the two densities (see, e.g., Bliss and Panigirtzoglou, 2004). The pricing kernel is defined as (see, e.g., Christoffersen et al., 2013):

$$M_T = e^{-rT} \frac{f^{\mathbb{Q}}_{S_T}(x)}{f^{\mathbb{P}}_{S_T}(x)}, \tag{2.23}$$

where $f^{\mathbb{Q}}_{S_T}(x)$ and $f^{\mathbb{P}}_{S_T}(x)$ represents the risk-neutral and real-world densities, respectively. Then from (2.3) we have that,

$$\mathrm{C}(T,K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[\max(S_T - K, 0)] \tag{2.24}$$

$$= \mathbb{E}^{\mathbb{P}}[M_T \max(S_T - K, 0)]. \tag{2.25}$$

The pricing kernel, M_T, describes how, in equilibrium, investors trade off the current (known) option prices versus the future (stochastic) pay-off (see, Christoffersen et al., 2013).

The literature provides us with guidance on how to estimate the pricing kernel (see, e.g., Aït-Sahalia and Lo, 2000; Bliss and Panigirtzoglou, 2004; Bakshi et al., 2003; Ross, 2015; Cuesdeanu and Jackwerth, 2018b). For example, Aït-Sahalia and Lo (2000) determined the pricing kernel by separately deriving the risk-neutral density from option prices and the real-world density from historical prices to infer the pricing kernel. However, Bliss and Panigirtzoglou (2004) assumed a constant risk aversion preference and estimated the elasticity parameter by comparing the predictions obtained from this with historical data. This method is also one of the most widely used transformation methods to obtain a real-world density from the risk-neutral density. The form for the pricing kernel can be motivated by a representative agent with a particular utility function of terminal wealth, $\upsilon(x)$ (see, e.g., Bliss and Panigirtzoglou, 2004). In particular, the risk-neutral density, real-world density and utility function are linked by (see, e.g., Christoffersen et al., 2013):

$$f^{\mathbb{P}}_{S_T}(x) = \frac{f^{\mathbb{Q}}_{S_T}(x)/\upsilon'(x)}{\int_0^\infty f^{\mathbb{Q}}_{S_T}(y)/\upsilon'(y)\,\mathrm{d}y}, \tag{2.26}$$

where $\upsilon'(x)$ is the first derivative of the utility function. That is, when we substitute (2.26) into (2.23), we have that,

$$M_T \propto \upsilon'(x). \tag{2.27}$$

For example, assuming a power utility with a constant relative risk aversion (CRRA) parameter γ, we have

$$\upsilon(x) = \begin{cases} \frac{1}{1-\gamma}x^{1-\gamma}, & \text{if } \gamma \geq 0; \gamma \neq 1, \\ \log x, & \text{if } \gamma = 1, \end{cases} \tag{2.28}$$

so that $\upsilon'(x) = x^{-\gamma}$ for $\gamma \neq 1$. Then the pricing kernel becomes $e^{-rT}x^{-\gamma}$, and the risk-adjusted density is:

$$f^{\mathbb{P}}_{S_T}(x) = \frac{e^{-rT}f^{\mathbb{Q}}_{S_T}(x)}{M_T} = \frac{x^{\gamma}f^{\mathbb{Q}}_{S_T}(x)}{\int_0^\infty y^{\gamma}f^{\mathbb{Q}}_{S_T}(y)\,\mathrm{d}y}, \tag{2.29}$$

where the denominator in (2.29) ensures that $f^{\mathbb{P}}_{S_T}(x)$ is a valid density. In addition, Bliss and Panigirtzoglou (2004) reported the distribution of γ obtained from the S&P500 and FTSE 100 for the power and exponential-utility functions. In particular Bliss and Panigirtzoglou (2004) found $\gamma = 4.08$ as the optimal relative risk aversion parameter for the power utility function. Following Cuesdeanu and Jackwerth (2018a) and Cristóstomo (2021), we will apply a power utility function with CRRA coefficients of 0, 2 and 4 in this thesis. Furthermore, Thomas (2016) found that the power utility family is the most commonly used utility function in practical applications. The risk aversion parameter is a particularly important economic indicator as it reflects the subjective feelings of the investor (see, Thomas, 2016). Therefore, methods for extracting the CRRA estimate from the forward-looking moments instead of using historical data can improve the predictive power of the forecast density (see e.g., Bakshi and Madan, 2006; Cristóstomo, 2021). However, this is left for further research.

The results above demonstrate that knowing any two of the following three uniquely determine the third: (i) the physical density, (ii) the risk-neutral density; and (iii) the pricing kernel.

In the next section, we consider an extension of the risk-premium transformation presented in this section. This approach incorporates investor sentiment into the transformation from risk-neutral probabilities to real-world probabilities in the aim of obtaining a better estimate of the real-world density.

2.3.6 A Behavioural Approach

Both Barone-Adesi et al. (2017) and Cristóstomo (2021) studied behavioural approaches in estimating the forward-looking real-world return density by incorporating an investor sentiment function into the transformation. The investor sentiment function aims to summarise investor biases in particular regions of the return density. Typically, the stochastic discount factor (pricing kernel) incorporates the investor beliefs and preferences about the asset returns. Therefore, in this section, the stochastic discount factor, presented in Section 2.3.5 is expanded to incorporate sentiment effects as outlined in Cristóstomo (2021). In particular, Barone-Adesi et al. (2017) and Cristóstomo (2021) defined three sentiment-induced factors, namely, excessive optimism (which generates biases in the mean returns), overconfidence (which impacts volatility predictions) and tail sentiment (which is related to non-rational tail expectations).

In what follows next is a study of these factors, which follows from Cristóstomo (2021). In particular, sentiment proxies will be obtained from the observed market activity, which then will be incorporated into the stochastic discount factor, given in (2.23), when transforming the risk-neutral density to a real-world density.

2.3.6.1 A Measure of Investor Optimism and Overconfidence

Following Cristóstomo (2021), the behavioural transformation due to investor optimism and overconfidence can be obtained by the linear transformation of the original asset return forecast, X_T, to an adjusted forecast $\hat{X}_T$ as follows:

$$\hat{X}_T = \theta_{1,t} + X_T\theta_{2,t} + (1-\theta_{2,t})\mu, \tag{2.30}$$

where θ_1 and θ_2 denote the location and scale shift parameters, respectively and μ the mean forecast of X_T. More specifically, this linear transformation shifts the mean and variance of the original forecast to $\hat{\mu} = \mu + \theta_1$ and $\hat{\sigma}^2 = \theta_2^2\sigma^2$, respectively, where σ^2 represents the variation in the original return forecast X_T. In essence, the linear transformation in (2.30) behaves in the following way (see, Cristóstomo, 2021):

- When pessimistic investors are driving market prices, the risk-neutral densities tend to reflect downward bias in average return. Therefore, a behavioural adjustment shift of $\theta_1 > 0$ is carried out, which shifts the probability mass to higher returns. Alternatively, when optimistic investors are driving market prices, the original forecast is adjusted downward by the behavioural shift, $\theta_1 < 0$, which shifts the probability mass to lower returns.
- When under-confident investors are driving market prices, excessive dispersion is typically embedded in the market forecast, and the behavioural correction for $\theta_2 < 1$ reduces the volatility in the distribution. Alternatively, when over-confident investors are driving market prices, the market forecast is adjusted by the behavioural correction, $\theta_2 > 1$, which increases the volatility in the market-implied forecast.

Now that we know how investor optimism and overconfidence affects the risk-neutral return density, we next estimate the investor optimism and overconfidence parameters, namely, θ_1 and θ_2, respectively from market data.

2.3.6.2 Estimation of the Investor Optimism and Overconfidence Parameters

Cristóstomo (2021) proposed using a proxy for investor optimism that is derived directly from the implied volatility extracted from market option prices. Typically when investors are pessimistic about future market returns, they tend to increase their demand for hedging securities, which raises implied volatility levels (see, e.g., Smales, 2017). Furthermore, Cristóstomo (2021) proposed using the implied volatility change, defined as the difference between the current implied volatility and the average implied volatility over the three previous months, rather than the implied volatility itself, as the change in implied volatility often yields superior results in explaining future returns. That is,

$$\Delta\sigma_t^{(\mathrm{IV})} = \sigma_t^{(\mathrm{IV})} - \bar{\sigma}_{t-1}^{(\mathrm{IV})}, \tag{2.31}$$

where

$$\bar{\sigma}_{t-1}^{(\mathrm{IV})} = \sum_{s=1}^{3} \sigma_{t-s}^{(\mathrm{IV})}/3. \tag{2.32}$$

Thereafter, the Gaussian kernel is used to transform the set of historical implied volatility changes until date t, $\left\{\Delta\sigma_1^{(\mathrm{IV})},\ldots,\Delta\sigma_{t-1}^{(\mathrm{IV})}\right\}$, into a continuous distribution. Then the α_t^{IV}th percentile of the new $\Delta\sigma_t^{(\mathrm{IV})}$ observation is found from the corresponding empirical distribution. Observations falling below the 5th percentile and above the 95th percentile are associated with excessive optimism and excessive pessimism, respectively. The behavioural mean shift, θ_1, is then calculated as follows (see, Cristóstomo, 2021):

$$\theta_{1,t} = \begin{cases} (1-e^{r_t\tau})k_1\frac{0.05-\alpha_t^{IV}}{0.05}, & \text{for } 0 \leq \alpha_t^{IV} < 0.05, \\ 0, & \text{for } 0.5 \leq \alpha_t^{IV} \leq 0.95, \\ -(1-e^{r_t\tau})k_1\frac{\alpha_t^{IV}-0.95}{0.05}, & \text{for } 0.95 < \alpha_t^{IV} \leq 1, \end{cases} \tag{2.33}$$

where k_1 represents a calibration impact factor, r_t the risk-free interest rate and τ the forecast duration. Cristóstomo (2021) suggested $k_1 = 1$ for a low calibration impact factor, which represents that the investor optimism may lead to a mean shift of up to one times the risk-free rate, whereas $k_1 = 2$ represents a high impact factor, which represents a behavioural shift of up to two times the risk-free rate.

Similar to the above, Statman et al. (2006) proposed that investor overconfidence can be measured by trading volumes. In particular, overconfident investors tend to typically overstate

their valuation, understating future volatility, and thereby trading more. Cristóstomo (2021), thereby uses the change in trading volume as a measure of overconfidence as it is often a preferable measure to the trading volume itself (see also, Statman et al., 2006). Therefore, at each time-t, Cristóstomo (2021) calculates the change in trading volume as the ratio of the last months trading volume and the average volume over the previous three months. That is,

$$\Delta TV_t = \frac{TV_t}{\bar{TV}_{t-1}}, \tag{2.34}$$

where

$$\bar{TV}_{t-1} = \sum_{s=1}^{3} TV_{t-s}/3. \tag{2.35}$$

Thereafter, the Gaussian kernel is used to transform the set of historical changes in trading volume, $\{\Delta TV_1, \ldots, \Delta TV_{t-1}\}$ into a continuous distribution. Under-confidence is associated with ΔTV_t observations that fall below the 5th percentile and overconfidence the values that fall above the 95th percentile. The volatility adjustment factor is then obtained as follows (see, Cristóstomo, 2021):

$$\theta_{2,t} = \begin{cases} k_2^{\frac{\alpha_t^{TV} - 0.05}{0.05}}, & \text{for } 0 \leq \alpha_t^{TV} < 0.05, \\ 1, & \text{for } 0.5 \leq \alpha_t^{TV} \leq 0.95, \\ k_2^{\frac{\alpha_t^{TV} - 0.95}{0.05}}, & \text{for } 0.95 < \alpha_t^{TV} \leq 1, \end{cases} \tag{2.36}$$

where α_t^{TV} represents the percentile of the new ΔTV_t observation in the empirical distribution, and k_2 an impact factor. Furthermore, Cristóstomo (2021) proposed a low impact scenario when $k_2 = 1.2$, which yields $\theta_2 \in [0.83, 1.2]$ and a high impact scenario when $k_2 = 1.5$, which yields $\theta_2 \in [0.67, 1.5]$.

After the linear transformation, Cristóstomo (2021) defined the mean-variance pricing kernel as:

$$M_T^{mv} = \frac{f_{t,X_T}^{\mathbb{P}}(x_T)}{f_{t,\hat{X}_T}^{mv}(x_T)}, \tag{2.37}$$

where $f_{t,X_T}^{\mathbb{P}}(x_T)$ represents a risk-adjusted density (see, e.g., (2.29)), and $f_{t,\hat{X}_T}^{mv}(x_T)$ represents the behavioural density obtained through the mean-variance transformation given in (2.30).

Next, we outline the behavioural shift represented by tail sentiment.

2.3.6.3 Tail Sentiment

Similar to the above, Cristóstomo (2021) proposes a simple tail adjustment to account for the tail bias, which shifts the probability mass from the one tail to the other tail. By denoting $q(\alpha)$ as the quantile that defines the left tail and $q(1-\alpha)$ the quantile that defines the right tail, the tail shift pricing kernel is given as follows (see, Cristóstomo, 2021):

$$M_T^{ts} = \begin{cases} e^{\theta_3(q(\alpha)-x_T)}, & \text{for } x_T \in (-\infty, q(\alpha)) \\ 1, & \text{for } x_T \in [q(\alpha), q(1-\alpha)] \\ e^{-\theta_3(x_T-q(1-\alpha))}, & \text{for } x_T \in (q(1-\alpha), \infty), \end{cases} \tag{2.38}$$

where θ_3 assigns the direction and intensity of the tail shift. In particular, when investors overstate the left-tail events, the behavioural transformation $\theta_3 > 0$, shifts probability from the left tail to the right tail. Alternatively, when investors overstate right-tail events, the behavioural transformation $\theta_3 < 0$, shift probability from the right tail to the left tail.

The tail sentiment parameter can then be estimated from the skewness of the risk-neutral distribution.

2.3.6.4 Estimation of the Tail Sentiment Parameter

The inferred amount of tail risk in the market can be obtained from the risk-neutral distribution. Typically the risk-neutral distribution becomes more negatively skewed when investors perceive downside risk. Alternately, the risk-neutral skewness becomes positive when investors perceive higher returns. Therefore, on each time-t, the risk-neutral skewness is estimated through the model-free method (see, e.g., Bakshi et al., 2003; Christoffersen et al., 2013):

$$Skew_t = \frac{e^{r\tau}W(t,T) - 3\mu(t,T)e^{r\tau}V(t,T) + 2\mu(t,T)^3}{(e^{r\tau}V(t,T) - \mu(t,T)^2)^{3/2}}, \tag{2.39}$$

where

$$\mu(t,T) = e^{r\tau}\left(1 - e^{-r\tau} - \frac{1}{2}V(t,T) - \frac{1}{6}W(t,T) - \frac{1}{24}X(t,T)\right), \tag{2.40}$$

and $V(t,T), W(t,T)$ and $X(t,T)$ represent hypothetical contracts with a quadratic, cubic and quartic payoff respectively and τ defined as the time to maturity. That is,

$$V(t,T) = \int_{S_t}^{\infty} \frac{2\left(1-\log\left(\frac{K}{S_t}\right)\right)}{K^2}\mathrm{C}(t,T,K)\mathrm{d}K + \int_0^{S_t} \frac{2\left(1+\log\left(\frac{S_t}{K}\right)\right)}{K^2}\mathrm{P}(t,T,K)\mathrm{d}K,$$

$$W(t,T) = \int_{S_t}^{\infty} \frac{6\log\left(\frac{K}{S_t}\right) - 3\left[\log\left(\frac{K}{S_t}\right)\right]^2}{K^2}\mathrm{C}(t,T,K)\mathrm{d}K$$
$$- \int_0^{S_t} \frac{6\log\left(\frac{S_t}{K}\right) + 3\left[\log\left(\frac{S_t}{K}\right)\right]^2}{K^2}\mathrm{P}(t,T,K)\mathrm{d}K,$$

and

$$X(t,T) = \int_{S_t}^{\infty} \frac{12\left[\log\left(\frac{K}{S_t}\right)\right]^2 - 4\left[\log\left(\frac{K}{S_t}\right)\right]^3}{K^2}\mathrm{C}(t,T,K)\mathrm{d}K$$
$$+ \int_0^{S_t} \frac{12\left[\log\left(\frac{K}{S_t}\right)\right]^2 + 4\left[\log\left(\frac{S_t}{K}\right)\right]^2}{K^2}\mathrm{P}(t,T,K)\mathrm{d}K,$$

where $\mathrm{P}(t,T,K)$ represents the market-quoted put option price at time t with expiry T. Han (2008) and Chen and Gan (2018) examined high skewness levels associated with biases in the tail expectations. Cristóstomo (2021) proposes that sentiment biases in asset prices occur when the risk-neutral skewness exceeds a ± 1.5 threshold. It is well-known that equity distributions typically tend to exhibit negatively skewed distributions, which is expected to generate more frequent over weightings of the left tail compared to the right tail. The tail shift adjustment is then defined as (see, Cristóstomo, 2021):

$$\theta_{3,t} = \begin{cases} k_3(-Skew_t - 1.5), & \text{for } Skew_t < -1.5 \\ 0, & \text{for } -1.5 \leq Skew_t \leq 1.5 \\ k_3(-Skew_t + 1.5), & \text{for } Skew_t > 1.5, \end{cases} \tag{2.41}$$

where k_3 represents an impact factor. In particular, $k_3 = 1$ will be used in a low impact sentiment calibration, where $k_3 = 2$ will be used in a high impact sentiment calibration.

Next, the behavioural transformation is incorporated into the stochastic discount factor to obtain a more accurate real-world forecast density estimate.

2.3.6.5 Sentiment Function and Real-World Stochastic Discount Factor

The combined behavioural correction to transform the risk-neutral density into a real-world density is given by the sentiment function, $\Psi(x_T)$. More specifically, this stochastic discount

factor represents the combined effects of investor optimism, investor overconfidence and tail sentiment as follows (see, Cristóstomo, 2021):

$$\Psi(x_T) = M_T^{mv} M_T^{ts}. \tag{2.42}$$

Then, the real-world pricing kernel is given as (see, Cristóstomo, 2021):

$$M_T^{rw} = e^{-r\tau} \upsilon'(x_T)\Psi(x_T), \tag{2.43}$$

which reflects the cumulative impact of investor sentiment and risk-preference. The real-world density can then be obtained from (2.26) using (2.43) as follows:

$$f^{\mathbb{P}}_{t,X_T}(x_T) = \frac{f^{\mathbb{Q}}_{t,X_T}(x_T)/\left[\upsilon'(x_T)\Psi(x_T)\right]}{\int_0^\infty f^{\mathbb{Q}}_{t,X_T}(y)/\left[\upsilon'(y)\Psi(y)\right]\mathrm{d}y}, \tag{2.44}$$

where $\int_0^\infty f^{\mathbb{Q}}_{t,X_T}(y)/\left[\upsilon'(y)\Psi(y)\right]\mathrm{d}y$ represents the integration factor used to ensure that (2.44) integrates to one. The real-world forward-looking density obtained using this method will be referred to as the Behavioural real-world density for the remainder of this thesis.

2.3.7 Calibration Transformations

An alternative transformation process to the utility-transformation method is to make use of the calibration of a misspecified density (see, e.g., Bunn, 1984; Fackler and King, 1990; Liu et al., 2007; Shackleton et al., 2010). This approach is particularly different to that of the utility-transformation function approach given in Section 2.3.5, as it does not make use of a representative agent.

In particular, this approach defines the new random variable, $U_{i+1} = F^{\mathbb{Q}}_{i,S_{i+1}}(S_{i+1})$, where $0 \leq i \leq t-1$ and $F^{\mathbb{Q}}_{i,S_{i+1}}(x)$ denotes the time-i risk-neutral cumulative distribution function of the terminal index price S_{i+1}. Thereafter, following Bunn (1984), the calibration function, $C_T(u)$, is defined as the real-world cdf of U_T. For real-world probabilities, the cdf of S_T is then obtained as follows (see, Liu et al., 2007):

$$\begin{aligned} F^{\mathbb{P}}_{t,S_T}(x) &= \mathbb{P}(S_T \leq x) \\ &= \mathbb{P}\left(F^{\mathbb{Q}}_{t,S_T}(S_T) \leq F^{\mathbb{Q}}_{t,S_T}(x)\right) \\ &= \mathbb{P}\left(U_T \leq F^{\mathbb{Q}}_{t,S_T}(x)\right) \\ &= C_T\left(F^{\mathbb{Q}}_{t,S_T}(x)\right). \end{aligned} \tag{2.45}$$

The real-world density is then given by (see, Liu et al., 2007):

$$f^{\mathbb{P}}_{t,S_T}(x) = \frac{\mathrm{d}}{\mathrm{d}x} C_T\left(F^{\mathbb{Q}}_{t,S_T}(x)\right) = f^{\mathbb{Q}}_{t,S_T}(x) c_T\left(F^{\mathbb{Q}}_{t,S_T}(x)\right), \tag{2.46}$$

where $c_T(u)$ represents the real-world density of U_T.

In order to estimate the real-world density in (2.46), the calibration function must first be estimated. Shackleton et al. (2010) illustrated two specifications of the calibration function, namely a parametric and a non-parametric specification. Furthermore, the calibration function is continuously re-estimated at the end of each period t.

2.3.7.1 Parametric Specification

Shackleton et al. (2010) used the cdf of the beta distribution as their parametric specification of the calibration function. That is, the calibration density is then:

$$c_T(u) = \frac{u^{a-1}(1-u)^{b-1}}{B(a,b)}, \tag{2.47}$$

where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, and the two calibration parameters, a and b depend on the forecast horizon T. From (2.46) and (2.47), the real-world density of S_T is then:

$$f^{\mathbb{P}}_{t,S_T}(x) = f^{\mathbb{Q}}_{t,S_T}(x) \frac{F^{\mathbb{Q}}_{t,S_T}(x)^{a-1}\left(1-F^{\mathbb{Q}}_{t,S_T}(x)\right)^{b-1}}{B(a,b)}. \tag{2.48}$$

In particular, the special case of the calibration density is when $a = b = 1$, which defines an uniform distribution in (2.47). This then implies that the real-world density in (2.48) is the same as the risk-neutral density.

The parameters in the calibration density function, a and b, are obtained by maximising the log-likelihood function at time t, where the log-likelihood function is given as (Shackleton et al., 2010):

$$\ell(S_1,\ldots,S_t|a,b) = \sum_{i=0}^{t-1} \log\left(f^{\mathbb{P}}_{i,S_{i+1}}(S_{i+1}|a,b,\Theta_i)\right), \quad t < T, \tag{2.49}$$

where Θ_i represents the risk-neutral parameter set at the end of the i-th day.

In this study, we will use the risk-neutral probability density function obtained by the Heston model with the beta parametric specification to obtain the real-world probability density function. This method will be denoted as the Heston P1 method for the remainder of this thesis.

2.3.7.2 Non-Parametric Specification

The non-parametric specification function is also calibrated from the historical observations of U_{i+1}. Firstly, Shackleton et al. (2010) uses the set of t observed future prices of the underlying asset to define a set of t cumulative risk-neutral probabilities, $U_{i+1} = F^{\mathbb{Q}}_{i,S_{i+1}}(S_{i+1}|\Theta_i)$, where $0 \leq i \leq t-1$, and Θ_i represents the parameter set of the risk-neutral density function at time-i. It is assumed that these observations are i.i.d. with cdf given by the calibration function $C_T(u)$. Thereafter, the set of U observations are transformed into new variables $y_{i+1} = \Phi^{-1}(U_{i+1})$, where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. This is known as the inverse probability transformation of the realisations (see, Rosenblatt, 1952). Then, a non-parametric kernel cdf is fit to the observation set $\{y_1, y_2, \ldots, y_t\}$. More specifically, a normal kernel with bandwidth h is used so to obtain the following kernel density (see, Shackleton et al., 2010):

$$\hat{g}_T(y) = \frac{1}{th}\sum_{j=1}^{t}\phi\left(\frac{y-y_j}{h}\right), \tag{2.50}$$

and cdf,

$$\hat{G}_T(y) = \frac{1}{t}\sum_{j=1}^{t}\Phi\left(\frac{y-y_j}{h}\right), \tag{2.51}$$

where the Silverman (1986, Chapter 2) standard optimal bandwidth parameter, $h = 0.9\sigma_y/t^{0.2}$, is used, where σ_y represents the standard deviation of the set $\{y_1, y_2, \ldots, y_t\}$. The empirical calibration function is then defined as (see, Shackleton et al., 2010):

$$\hat{C}_T(u) = \hat{G}_T\left(\Phi^{-1}(u)\right). \tag{2.52}$$

Using (2.45), the real-world cdf for the next future price is (see, Shackleton et al., 2010):

$$F^{\mathbb{P}}_{t,S_T}(x) = \hat{C}_T\left(F^{\mathbb{Q}}_{t,S_T}(x)\right) = \hat{G}_T\left[\Phi^{-1}\left(F^{\mathbb{Q}}_{t,S_T}(x)\right)\right], \tag{2.53}$$

with real-world density:

$$f^{\mathbb{P}}_{t,S_T}(x) = \frac{d}{dx}\hat{G}_T\left[\Phi^{-1}\left(F^{\mathbb{Q}}_{t,S_T}(x)\right)\right] = \frac{f^{\mathbb{Q}}_{t,S_T}(x)\hat{g}_T\left[\Phi^{-1}\left(F^{\mathbb{Q}}_{t,S_T}(x)\right)\right]}{\phi\left[\Phi^{-1}\left(F^{\mathbb{Q}}_{t,S_T}(x)\right)\right]}. \tag{2.54}$$

Then, the non-parametric calibration density is given as:

$$\hat{c}_T\left(F^{\mathbb{Q}}_{t,S_T}(x)\right) = \frac{\hat{g}_T\left[\Phi^{-1}\left(F^{\mathbb{Q}}_{t,S_T}(x)\right)\right]}{\phi\left[\Phi^{-1}\left(F^{\mathbb{Q}}_{t,S_T}(x)\right)\right]}. \tag{2.55}$$

This transformation method from risk-neutral to real-world probabilities will be referred to as the Heston P2 method for the remainder of this thesis.

2.4 Conclusion

In this chapter, we introduced methods for extracting the forecast distribution. In particular, we focused on extracting the risk-neutral forward-looking probability density function from option prices, and then transforming the risk-neutral probability density function into a real-world probability density function. The methods discussed in this chapter form part of the standard approaches in transforming risk-neutral probabilities into real-world probabilities, which uses historical asset returns to estimate the pricing kernel and calibration function.

Although, the risk-neutral distribution is considered biased, the bias may be negligible. Furthermore, attempting to remove the bias, by transforming the risk-neutral probabilities into real-world probabilities, may also lead to introducing bias into the distribution as this process often makes use of historical prices and assumptions on investor preference (see, Christoffersen et al., 2013).

In contrast, Ross (2015) extracts both the physical and pricing kernel from the risk-neutral density by placing more strict assumptions on the market. This approach will be discussed, in the next chapter.

CHAPTER 3

THE ROSS RECOVERY THEOREM WITH A REGULARISED MULTIVARIATE MARKOV CHAIN

This chapter is mostly aligned with Van Appel and Maré (2018) and forms part of our first research question.

Chapter Synopsis

Recently, Ross (2015) derived a theorem, namely the "Recovery Theorem", that allows for the recovery of the pricing kernel and real-world asset distribution, under particular assumptions, from a forward-looking risk neutral distribution. However, recovering the real-world distribution involves solving two ill-posed problems. In this chapter, we introduce and test the accuracy of a regularised multivariate mixture distribution to recover the real-world distribution. In addition, we show that this method improves the estimation accuracy of the real-world distribution. Furthermore, we carry out an empirical study, using weekly South African Top40 option trade data, to show that the recovered distribution is in line with economic theory.

3.1 Introduction

Asset distributions are vitally important to solve financial problems in risk management, portfolio optimisation and optimal trading strategies. A commonly used approach to forecast returns is to use historical data or opinion polling to estimate asset distributions. However, financial

markets are quite volatile, and using historical distributions for forecasting are not always desirable. An alternative forecasting method is to extract the forward-looking risk-neutral distribution from the option market data. It is well-known that option prices convey some market risk forecast as payoffs extend out in time. Therefore, option prices are, by nature, forward-looking. In a complete market, Black and Scholes (1973) and Merton (1973) proved that the value of an option is independent of the expected return on the underlying asset. This gave rise to the risk-neutral valuation framework, where the only unknown parameter affecting the option price is the asset's underlying volatility, commonly referred to as the implied volatility. Furthermore, Breeden and Litzenberger (1978) showed that the forward-looking risk-neutral distribution can be derived by option prices under the assumption of complete markets. However, the risk-neutral distribution mostly differs from the real-world distribution, which expresses market participants' consensus. In short, under the risk-neutral measure, the expected return of the asset is the risk-free rate, since the risk-neutral measure is the real-measure with the risk-premium removed.

While financial institutions have long used implied volatilities to gauge the market's perception of risk, there has been a theoretical hurdle in converting the risk-neutral distribution into the real-world distribution using option prices alone. Recently, Ross (2015) published a remarkable theorem that recovers the real-world probability distribution and pricing kernel from option prices under a particular set of assumptions. For example, one of the assumptions is that markets are complete. This is rarely true in any exchange traded option dataset, especially in South Africa, where the option price data are sparse and noisy. To satisfy this assumption, it is necessary to extrapolate forward-looking option price data (see, e.g., Aït-Sahalia and Lo, 1998; Audrino et al., 2014; Melick and Thomas, 1997; Gatheral and Jacquier, 2014; Flint and Maré, 2017). More specifically, Gatheral and Jacquier (2014) proposed a deterministic SVI volatility model with a robust fitting algorithm to estimate volatility surfaces, which proved to be a promising method to estimate the forward-looking risk-neutral distribution (see, also, Flint and Maré, 2017).

It is well-known that the risk-neutral probability measure is extensively used in derivative pricing, however, knowledge of the pricing kernel and real-world distribution will be invaluable for investors regrading risk management, portfolio optimisation and investment strategies. In short, the recovery theorem differs from other approaches in that it adjusts the risk-neutral

distribution to a real-world distribution and does not rely on historical returns.

The empirical problem with the recovery theorem is that it is difficult to recover an accurate real-world distribution (see, e.g., Audrino et al., 2014; Backwell, 2015; Spears, 2013), as it involves solving two ill-posed problems. The first ill-posed problem involves finding the risk-neutral distribution by taking the second derivative of the option pricing function and the second involves calculating the transition matrix that captures the dynamics of the state prices. In this chapter, we will focus on the second ill-posed problem by implementing a regularised multivariate Markov chain in an attempt to stabilise the estimation of the real-world transition distribution matrix. In addition, we will conduct a numerical analysis and a robustness check to show the effectiveness of this method. Thereafter, we will apply the recovery theorem to weekly Top40 option trade data, traded on the South African Futures Exchange (SAFEX), to estimate the real-world distribution. In addition, we compare the first four moments of the real-world distribution to the risk-neutral distribution.

3.2 The Recovery Theorem

In this section, we start by reviewing the recovery theorem Ross (2015). For simplicity, we adopt some of the notation and terminology used in Ross (2015). Intuitively, he attempts to recover the real-world transition probabilities of a Markovian state variable S that determines aggregate consumption, using market derivative prices on S. Real-world probabilities differ from risk-neutral probabilities in that investors require a premium that compensates them for carrying risk. The transformation from risk-neutral to real-world densities rely on assumptions (see, e.g., Bliss and Panigirtzoglou, 2004; Shackleton et al., 2010; Ross, 2015; Dillschneider and Maurer, 2019). Moreover, Ross (2015) proposed a model-free method to recover the real-world transition matrix from a Markovian state variable S, under a particular set of assumptions, using market-based derivative prices. These assumptions are: (i) the transition state prices are strictly positive, (ii) the transition state prices are time-homogeneous, and (iii) the pricing kernel is transition independent. Firstly, he used the method proposed by Breeden and Litzenberger (1978) (see also Section 2.3.4) to construct a $n \times m$ state price matrix, S, by taking the second derivative with respect to the strike of a European call option at each tenor, t, i.e.,

$$S(t,K) = e^{rt}\frac{\partial^2 \mathrm{C}(t,K)}{\partial K^2}, \quad t = 1,\ldots,m. \tag{3.1}$$

Consider a time-homogeneous process $\{S_t\}_{t\geq 0}$ on a finite state space with n states. Numerically approximating (3.1) yields the forward-looking state price function for each tenor. This is done by numerically integrating (3.1) over the discrete grid for each of the n states (see, e.g., Chakraborty, 2015). In essence, the discretised $S(t,n)$ represents the price of an Arrow-Debreu security that agrees to pay one unit of currency if state j is reached at time t and zero in all other states. Secondly, he constructs a $n \times n$ one period ahead irreducible time-homogeneous state transition probability matrix: Since, calendar time is irrelevant, the transition probability moving from state i at time t to state j at time $t+1$ is given by:

$$P_{i,j} = \mathbb{P}(S_{t+1} = j | S_t = i), \qquad \forall t = 1, \ldots, m-1, \tag{3.2}$$

where P denotes a $n \times n$, one period ahead, irreducible transition matrix and the elements of P can easily be estimated by solving the following system of equations:

$$S_{t+1}^{\top} = S_t^{\top} P, \quad t = 1, 2, \ldots, m-1. \tag{3.3}$$

Intuitively, $P_{i,j}$ denotes the contingent price of a security that pays out one unit of currency if the security moves from state i to state j in one period, which is known as the contingent forward prices of a security. That is, P represents the richer set of probabilities of moving from all hypothetical initial states to all hypothetical future states, where (3.1) only represents the probabilities of moving from the single known current state to all future states.

If the rows of P sum to one, then we say P is a stochastic matrix; however, for the recovery theorem, P is sub-stochastic as it captures the dynamics of the discounted risk-neutral distribution, i.e., state prices. Therefore, the elements, $P_{i,j}$, of the transition matrix denote the value of an Arrow-Debreu security contract that pays one unit of the numeraire if a particular state is reached in the next time step and zero otherwise. But, by normalising the rows of P to sum to unity, we define a $n \times n$ transition risk-neutral probability matrix $\mathbb{Q}$, with elements:

$$q_{i,j} = \frac{P_{i,j}}{\sum_{k=1}^{n} P_{i,k}}, \qquad \forall i, j = 1, 2, \ldots, n. \tag{3.4}$$

The transition kernel, ψ, in Ross's framework is defined as the ratio price per unit of probability, i.e.,

$$\psi_{i,j} = \frac{P_{i,j}}{f_{i,j}}, \tag{3.5}$$

where $f_{i,j}$ is the real-world transition probabilities. Intuitively, one needs to solve two unknown quantities in (3.5) in order to recover the real-world probabilities. In order to do this, Ross

(2015) assumes that the kernel is transition independent. This assumption allows us to write the pricing kernel as

$$\psi_{ij} = \delta \frac{h(S_j)}{h(S_i)}, \tag{3.6}$$

where h is a positive function of states and δ a positive discount factor. Substituting (3.5) in (3.6) yields

$$P_{i,j} = \delta \frac{h(S_j)}{h(S_i)} f_{ij}. \tag{3.7}$$

Rewriting the state equations (3.7) in matrix form, we have

$$P = \delta D^{-1} F D, \tag{3.8}$$

where P is the $n \times n$ transition probability matrix, F is the $n \times n$ real-world transition matrix, and D is the $n \times n$ diagonal matrix with the undiscounted kernel, i.e.,

$$D = \text{diag}(h(S_1), h(S_2), \ldots, h(S_n)). \tag{3.9}$$

Solving for F in (3.8) yields

$$F = \frac{1}{\delta} D P D^{-1}. \tag{3.10}$$

Since F is a matrix whose rows are transition probabilities, i.e., a stochastic matrix, we have $F\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is a vector of ones. Using this condition, with (3.10), we have

$$P D^{-1} \mathbf{1} = \delta D^{-1} \mathbf{1}. \tag{3.11}$$

If we define the vector $\mathbf{z} \equiv D^{-1}\mathbf{1}$, we obtain

$$P\mathbf{z} = \delta \mathbf{z}. \tag{3.12}$$

If one assumes no arbitrage, then P is a non-negative matrix. Note that, if P is a positive matrix, then by definition, P is irreducible. However, if P is non-negative and all states are attainable from all other states in k steps, then P is also irreducible. Then from the Perron-Frobenius theorem (see, e.g., Meyer, 2000) there exists a unique positive eigenvector $\mathbf{z}$ and an associated maximum eigenvalue δ. Intuitively, Ross (2015) solves all three unknowns in (3.8) using the Perron-Frobenius Theorem. The following theorem guarantees a unique solution of this problem.

Theorem 3.1 (Recovery Theorem, Ross (2015)) *Assuming no arbitrage, irreducibility of the pricing matrix P, and that the pricing matrix is generated by a transition independent kernel, then given any set of state prices there exists a unique positive solution pair: the pricing kernel and real-world measure.*

In short, the recovery theorem allows us to uniquely find F from P. Knowledge of the real-world distribution will be of great benefit to financial practitioners. Although, many of the assumptions in the recovery theorem are violated in real life, Audrino et al. (2014), Kiriu and Hibiki (2019) and Flint and Maré (2017) showed by empirical studies that the real-world distribution obtained from the recovery theorem added economic value.

3.3 Implementation of the Ross Recovery Theorem

In this section, we describe the three step procedure, outlined in Spears (2013), for implementing the recovery theorem.

Step 1: Use the method proposed by Breeden and Litzenberger (1978) to construct a $n \times m$ state price matrix, S, by taking the second derivative with respect to the strike of a European call option at each tenor, i.e.,

$$S(t,K) = \frac{\partial^2 \mathrm{C}(t,K)}{\partial K^2}, \tag{3.13}$$

where $\mathrm{C}(t,K)$ is the current price of an European call option with strike, K, and tenor, t. Numerically approximating (3.13) yields the forward-looking state price function. In reality, a continuum of traded strikes is not directly observed in the markets. This is the first ill-posed problem. However, a wide range of state price estimation techniques can be found in the literature (see, e.g., Aït-Sahalia and Lo, 1998; Flint and Maré, 2017; Melick and Thomas, 1997). More specifically, Flint and Maré (2017) used the stochastic volatility inspired (SVI) model to produce the implied volatility surface, and thus, the state price surface. Furthermore, they showed that the deterministic SVI model is a promising candidate for modelling implied volatility surfaces and ultimately estimating the underlying risk-neutral distribution. The SVI model was first introduced by Gatheral (2004) and is

given by

$$\sigma^2(x,t) = a + b\left(\rho(x-m) + \sqrt{(x-m)^2 + s^2}\right), \tag{3.14}$$

where $x = \ln\left(\frac{K}{F}\right)$ is the log-forward-moneyness, and the coefficients a, b, ρ, s, and m depend on the expiration and have an intuitive geometric interpretation. Furthermore, the parametrisation of the SVI model makes it relative easy to eliminate calendar spread arbitrage, making the SVI model desirable Gatheral and Jacquier (2014). In Figure 3.1, we display an example of the implied volatility surface obtained by using the SVI model, where we can see that the SVI model provides a good interpolation of implied volatility.

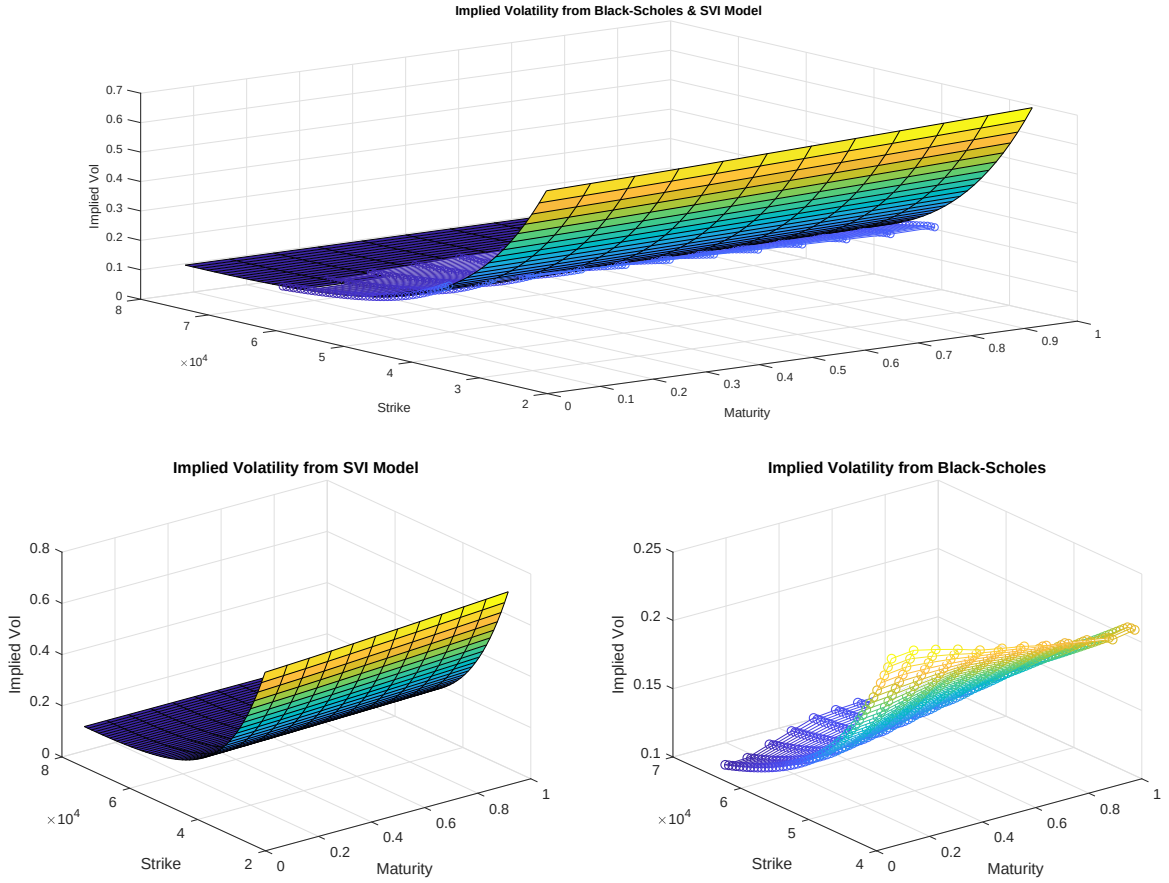


Figure 3.1: Implied volatility for the South African Top40 index on 15 January 2018: The mesh (bottom right) represent the quoted implied volatilities across maturity and strikes and the surface (bottom left) represents the implied volatilities across maturities and strikes using the SVI model. The top figure represents the overlay of the quoted and fitted implied volatilities.

After the implied volatility skews are calibrated, we can calculate the call option prices, using the Black-Scholes formula, across the full strike range for each term of the extrapolated implied volatility skews. Thereafter, using (3.13) we estimate the forward state price matrix. In Figure 3.2, we give an example of the forward state price matrix, using the extrapolated implied volatilities.

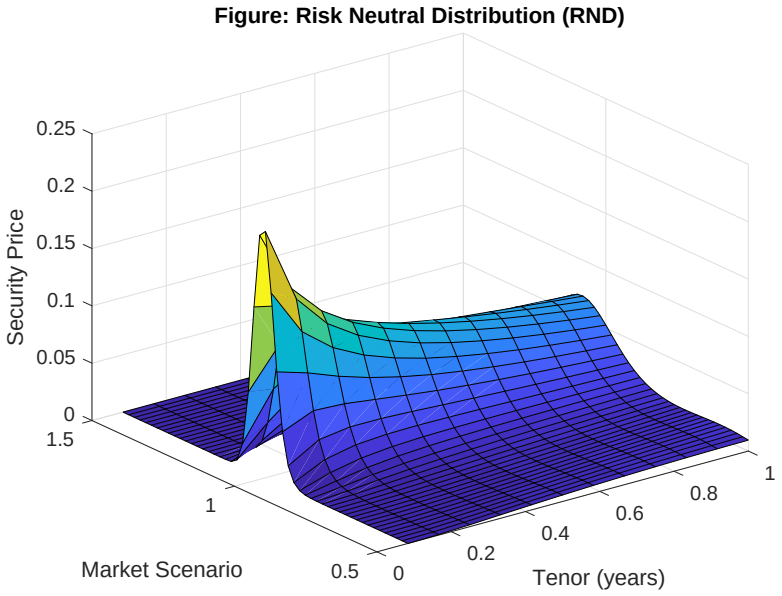


Figure 3.2: State prices for the South African Top40 index on 15 January 2018.

Step 2: Construct a $n \times n$ state transition probability matrix, P. Unfortunately, P is not directly observed, since a rich forward market for options does not exist. However, Ross (2015) shows that if $m \geq n$, we can estimate P, since it specifies a time-homogeneous transition from one maturity to the next, as follows:

$$S_{t+1}^{\top} = S_t^{\top} P, \quad t = 1, 2, \ldots, m-1. \tag{3.15}$$

If one denotes $A = S_{i,j}^{\top}$ and $B = S_{i,j+1}^{\top}$, where $1 \leq i \leq n$ and $1 \leq j \leq m-1$, then (3.15) can be rewritten as an ordinary least squares (OLS) problem, as follows:

$$P = \arg\min_{P} \|AP - B\|_2^2 \tag{3.16}$$

$$\text{subject to} \quad S_1 = P_{i_0}^{\top} \tag{3.17}$$

$$p_{i,j} \geq 0 \quad (i, j = 1, \ldots, n), \tag{3.18}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Since S_1 is the one period ahead state price and P is a one period state transition matrix, we have by definition a constraint (3.17), where i_0 is the current state (normally defined at the centre row of the transition matrix P, i.e., $i_0 = (n+1)/2$). In theory, equation (3.16) can easily be solved with standard optimisation techniques. Therefore, we numerically implement the OLS problem to derive the transition pricing matrix P.

Step 3: Using the Perron-Frobenius theorem, i.e., (3.12), we can extract a unique positive eigenvector, $\mathbf{z}$, and eigenvalue, δ. Thereafter, the elements of F can be calculated using (3.10).

The accuracy of the estimation of the real-world distribution, using the recovery theorem, largely depends on how accurately the transition matrix, P, is estimated. In the literature, it has proven to be difficult to accurately estimate (3.16) and furthermore to replicate the results indicated in Ross (2015). The reason for this, is that it involves solving the second ill-posed problem, where A is ill-conditioned (i.e., a small change in one of the coefficient values in A, results in a large relative change in the solution values), which renders active-set optimisation methods that are dependent on $A^\top A$ infeasible, as in this case. This can be seen in Audrino et al. (2014), Kiriu and Hibiki (2019), and Spears (2013) suggesting that Ross (2015) placed significant constraints on the structure of the transition matrix. In an attempt to replicate the results in Ross (2015), Spears (2013) implemented nine optimisation methods for solving (3.16). Furthermore, Sanford (2018) proposed a mixture transition distribution, where the proposed states depend on the current state price and its option implied volatilities to stabilise the estimation of P. More specifically, Sanford (2018) simplifies the original specification of the multivariate model by assuming that contingent state prices are solely defined by the state levels, but conditioned on the volatility. That is,

$$S_{t+1}^\top = S_t^\top P + \sigma_t^{(\mathrm{IV})}\beta, \qquad t = 1,2,\ldots,m-1, \tag{3.19}$$

where $\sigma_t^{(\mathrm{IV})}$ is the implied volatility state at time t as it is the best representation of the market's future volatility state and β is the volatility state regression coefficient. Furthermore, Sanford (2018) shows that the multivariate method had a significant improvement on the univariate recovery theorem as the volatility acts as a proxy for economical uncertainty. Similarly, equation

(3.19) can be reduced to the following general optimisation problem:

$$P = \arg\min_{P,\beta} \left\| AP + \sigma^{(\mathrm{IV})}\beta - B \right\|_2^2, \tag{3.20}$$

$$\text{subject to} \quad (3.17), (3.18) \text{ and } \beta \geq 0. \tag{3.21}$$

In theory, the multivariate model gives a third dimension in the Markov chain. Intuitively, more variables could be added to the regression model. However, this will come at a computational cost and the more variables added to the regression equation, will result in too few degrees of freedom to consider the resulting state price matrix, P, reliable.

An alternative method of stabilising the estimation of P is by adding a regularisation parameter to the estimation process. This has proven to be a successful method in the studies conducted by Audrino et al. (2014) and Kiriu and Hibiki (2019). Therefore, this chapter contributes in two ways. Firstly, we compare the multivariate method with the regularised methods (this has not been done to our knowledge) and secondly, due to the success of the regularised methods in the literature, we extend the multivariate method by adding a regularisation term.

3.3.1 Ridge Regularisation Methods

An effective method in stabilising the estimation of the transition matrix, P, is to introduce a regularisation term. The use of a regularisation term to solve ill-posed problems was first introduced by Tikhonov (1963). The Tikhonov method is a standard regularisation method used in the literature to solve ill-posed problems.

3.3.1.1 Tikhonov Regularisation without Prior Information

In this section, we review two regularisation methods to estimate P, found in the literature, and extend the multivariate method by adding a regularisation term. Audrino et al. (2014) first introduced the implementation of the Tikhonov regularisation (ridge regression) method in estimating P in the recovery theorem, by the following constrained optimisation problem:

$$P = \arg\min_{P} \|AP - B\|_2^2 + \zeta \|P\|_2^2 \tag{3.22}$$

$$\text{subject to} \quad (3.17) \text{ and } (3.18), \tag{3.23}$$

where ζ is a regularisation parameter that controls the trade-off between fitting and stability. The selection method of ζ is paramount in finding an accurate solution. Therefore, Audrino et al. (2014) proposed that an optimal ζ can be determined by minimising the discrepancy between the observable state price matrix (S^O) and the unrolled state price matrix (S^P) implied by matrix P, i.e.,

$$S_t^{P^\top} = \iota_{i_0}^\top P^t, \qquad t = 1, 2, \ldots, m, \tag{3.24}$$

where ι_{i_0} denotes a vector with 1 in the i_0^{th} position and zeros elsewhere, and P^t denotes the t-steps ahead state approximation. Furthermore, they use the Kullback-Leibler (KL) divergence as a measure of discrepancy between the two matrices, by solving ζ that minimises:

$$\arg\min_{\zeta} D_{KL}\left(S^O || S^P\right), \tag{3.25}$$

where

$$D_{KL}\left(S^O || S^P\right) = \sum_{i=1}^{n}\sum_{t=1}^{m} S_{i,t}^O \ln\left(\frac{S_{i,t}^O}{S_{i,t}^P}\right) - \sum_{i=1}^{n}\sum_{t=1}^{m} S_{i,t}^O + \sum_{i=1}^{n}\sum_{t=1}^{m} S_{i,t}^P, \tag{3.26}$$

and the optimal ζ is derived iteratively.

Note that equation (3.22) can be we rewritten as a constraint OLS problem as follows Audrino et al. (2014):

$$P = \arg\min_{P \geq 0} \left\| \begin{bmatrix} A \\ \sqrt{\zeta} I \end{bmatrix} P - \begin{bmatrix} B \\ O \end{bmatrix} \right\|_2^2, \tag{3.27}$$

where I denotes an identity matrix and O is a vector of zeros. In an empirical study using daily closing prices of out-of-the-money call and put options on the S&P 500 for each Wednesday between 5 January 2000 and 26 December 2012, Audrino et al. (2014) showed that the Tikhonov regularisation drastically improved the stability of the estimation of the transition matrix and showed that there is economic value in the recovered distributions.

In the next section, Kiriu and Hibiki (2019) extended the estimation of P by using the Tikhonov regularisation method with prior information.

3.3.1.2 Tikhonov Regularisation with Prior Information

The second regularisation method we review in this study was introduced by Kiriu and Hibiki (2019), where they extended the regularisation term above to consider prior information. For

the prior information, $\bar{P}$, they suggest that $p_{i,j}$ should be similar to $p_{i+k,j+k}$ for all $k \leq \min(n-i, n-j)$. Furthermore, they estimated P, using a problem specific error function in an attempt to balance the relative gain in the objective function from each term in the regularised optimisation problem, as follows:

$$P = \underset{P \geq 0}{\arg\min} \|AP - B\|_2^2 + \zeta \|P - \bar{P}\|_2^2 \tag{3.28}$$

$$= \underset{P \geq 0}{\arg\min} y^{\text{fit}}(\zeta) + \zeta y^{\text{reg}}(\zeta) \tag{3.29}$$

$$\text{subject to} \quad (3.17) \text{ and } (3.18), \tag{3.30}$$

where

$$\bar{P} = \begin{bmatrix} \sum_{k=1}^{i_0} s_{k,1} & s_{i_0+1,1} & \cdots & s_{n-1,1} & s_{n,1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \sum_{k=1}^{2} s_{k,1} & s_{3,1} & \cdots & s_{i_0,1} & s_{i_0+1,1} & s_{i_0+2,1} & \cdots & s_{n,1} & 0 \\ s_{1,1} & s_{2,1} & \cdots & s_{i_0-1,1} & s_{i_0,1} & s_{i_0+1,1} & \cdots & s_{n-1,1} & s_{n,1} \\ 0 & s_{1,1} & \cdots & s_{i_0-2,1} & s_{i_0-1,1} & s_{i_0,1} & \cdots & s_{n-2,1} & \sum_{k=n-1}^{n} s_{k,1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & s_{1,1} & s_{2,1} & \cdots & s_{i_0-1,1} & \sum_{k=i_0}^{n} s_{k,1} \end{bmatrix}, \tag{3.31}$$

$y^{\text{fit}}(\zeta)$ represents the fitting error and $y^{\text{reg}}(\zeta)$ represents the deviation between P and $\bar{P}$. Furthermore, Kiriu and Hibiki (2019) showed that as ζ increases, y^{fit} decreases and y^{reg} increases monotonically. Therefore, they selected ζ by minimising the problem specific function:

$$h(\zeta) = \frac{y^{\text{fit}}(\zeta) - y^{\text{fit}}(0)}{y^{\text{fit}}(\infty) - y^{\text{fit}}(0)} + \frac{y^{\text{reg}}(\zeta) - y^{\text{reg}}(\infty)}{y^{\text{reg}}(0) - y^{\text{reg}}(\infty)}, \tag{3.32}$$

where the denominators represents the maximum spread in each term and the numerator represents the spread for a specified ζ value.

In addition, Kiriu and Hibiki (2019) compared the effectiveness of this selection method with (3.26), where they found that (3.32) yielded better results. Therefore, for the remainder of this study, we will use (3.32) as the selection method for ζ. Equation (3.28) can also be formulated as an OLS problem, as follows Kiriu and Hibiki (2019):

$$P = \underset{P \geq 0}{\arg\min} \left\| \begin{bmatrix} A \\ \sqrt{\zeta} I \end{bmatrix} P - \begin{bmatrix} B \\ \sqrt{\zeta} \bar{P} \end{bmatrix} \right\|_2^2. \tag{3.33}$$

In a simulated study, Kiriu and Hibiki (2019) showed that their method estimated the real-world distribution more accurately than the Tikhonov method proposed by Audrino et al. (2014). Furthermore, in a similar empirical study to Audrino et al. (2014), Flint and Maré (2017) implemented the regularisation method with prior information to extract the real-world distribution on a history of implied volatility surfaces for the South African Top40 index, where they showed that the recovered real-world moments are in line with economic rationale and showed promising results when used in a simple asset allocation framework.

Since the regularisation methods have proven to be a powerful method in estimating the real-world distribution in the recovery theorem, we extend the multivariate method to a regularised multivariate method in the next section.

3.3.1.3 The Multivariate Model with a Tikhonov Regularisation

Later, we will show that the addition of the regularisation term in the estimation procedure improves the estimation of P and ultimately F (see also, e.g., Audrino et al., 2014; Kiriu and Hibiki, 2019). Therefore, we extend the multivariate Markov chain proposed by Sanford (2018) to a regularised multivariate Markov chain by adding the regularisation parameter as follows:

$$P = \underset{P,\beta}{\arg\min} \left\| AP + \sigma^{(\mathrm{IV})}\beta - B \right\|_2^2 + \zeta \|P\|_2^2, \qquad \forall t = 1,2,\ldots,m-1 \tag{3.34}$$

$$\text{subject to} \quad (3.17), (3.18) \text{ and } \beta \geq 0. \tag{3.35}$$

Furthermore, we also extend the optimisation problem above, with the regularisation of prior information, as such,

$$P = \underset{P,\beta}{\arg\min} \left\| AP + \sigma^{(\mathrm{IV})}\beta - B \right\|_2^2 + \zeta \|P - \bar{P}\|_2^2, \qquad \forall t = 1,2,\ldots,m-1 \tag{3.36}$$

$$\text{subject to} \quad (3.17), (3.18) \text{ and } \beta \geq 0, \tag{3.37}$$

where $\bar{P}$ is given in (3.31). We found that the regularised method with prior information performed better than the regularised method without prior information. Therefore, we will only consider (3.36) in the remainder of this chapter.

3.3.2 Elastic Net Regularisation Method

Elastic net is a regression regularisation method used in statistics, that linearly combines the L_1 and L_2 penalties of the lasso and ridge methods. The (L_1) penalty achieves sparsity in the model by setting the irrelevant regression coefficient equal to zero and the (L_2) penalty achieves robustness in the model. Therefore the optimisation problem becomes:

$$P = \underset{P \geq 0}{\arg\min} \, \|AP - B\|_2^2 + \zeta \, \|P\|_2^2 + \lambda \, \|P\|_1 \tag{3.38}$$

$$\text{subject to} \quad (3.17) \text{ and } (3.18), \tag{3.39}$$

where the estimation is carried out in a two-stage procedure as follows: for each fixed ζ, it finds the ridge regression coefficients and then does a lasso shrinkage along the lasso coefficient path Zou and Hastie (2005). Furthermore, Zou and Hastie (2005) refer to this as the naïve elastic net criterion, since it appears to amount to double shrinkage, where it was found that the naïve elastic net regularisation method does not perform well, unless it is close to ridge or lasso. In our study, we find that λ is small indicating that it is close to ridge. However, to improve the prediction performance, Zou and Hastie (2005) rescale the coefficients of the naïve version of elastic net by multiplying the estimated coefficients by $(1 + \lambda_2)$. Next, we add the prior information to (3.38), yielding

$$P = \underset{P \geq 0}{\arg\min} \, \|AP - B\|_2^2 + \zeta \, \|P - \bar{P}\|_2^2 + \lambda \, \|P\|_1 \tag{3.40}$$

$$\text{subject to} \quad (3.17) \text{ and } (3.18). \tag{3.41}$$

The elastic net with prior information yielded better results than without the prior information. Therefore, for the remainder of this study, we will only show the results for the elastic nets with prior information.

In the next section, we compare the estimation methods discussed above by estimating the real-world distribution, where we will show that the regularised multivariate method gives a better estimate than the methods reviewed by conducting a similar simulation study to Kiriu and Hibiki (2019).

3.4 Comparison of Methods

In this section, we compare the accuracy of the estimation of P, using the methods discussed in Section 3. We will follow the same estimation accuracy procedure and robust check outlined in Kiriu and Hibiki (2019) as follows:

1. Firstly, a hypothetical real-world matrix (F^H) is obtained from the historical daily S&P 500 index price data. More specifically, we set 11 returns (states) in total, placed every 6% symmetrically around 0%. F^H is generated by setting a reference date and calculating 12 returns every 30 calendar days, where the S&P 500 logarithmic returns are calculated as follows:

$$\text{Return} = \log\left(\frac{S_T}{S_0}\right) \times 100\%.$$

A matrix is generated by calculating the number of state transitions of the return in one period. This is repeated daily by changing the reference date from 02 January 1986 to 30 December 2016. Thereafter, all matrices are summed up and divided by the summed matrix row total, giving an 11×11 probability matrix (see Figure 3.3). Secondly, the

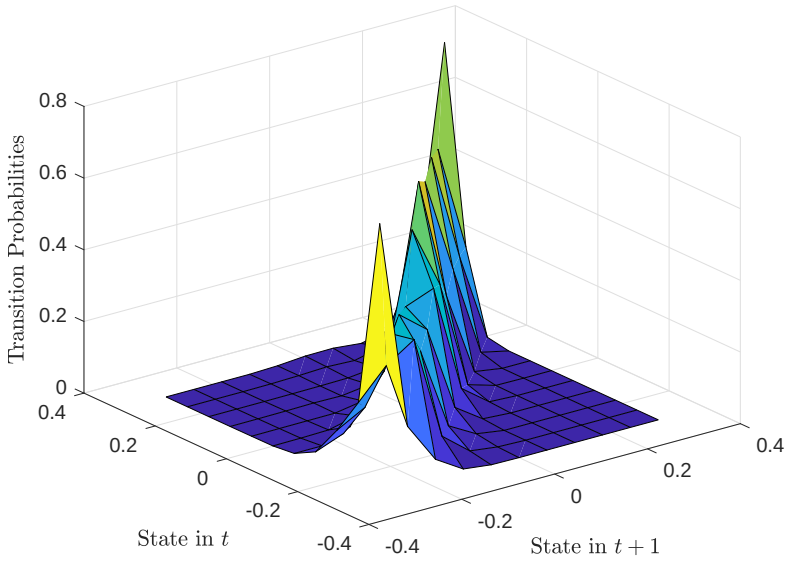


Figure 3.3: Hypothetical real-world distribution, F^H.

pricing kernel matrix (Φ^H) is obtained by assuming that the investor has a CRRA utility

function, $U(c) = c^{1-\gamma_R}/(1-\gamma_R)$, with relative risk aversion γ_R, i.e.,

$$\phi_{i,j} = \delta \left(\frac{1+r_j}{1+r_i} \right)^{-\gamma_R} \quad \forall i, j = 1, \ldots, n, \tag{3.42}$$

where $\gamma_R = 3$ and $\delta = 0.999$. These parameters were chosen to be consistent with the parameters reported in Bliss and Panigirtzoglou (2004), where they estimated the risk aversion parameter, γ, implied in the S&P 500 option data and historical option price data from 1993 to 2010, to have a minimum risk aversion parameter value to be 3.37 and a maximum value of 9.52. The maximum parameter value will be used in the robustness check in Section 4.2.

2. A hypothetical transition state price matrix P^H (see Figure 3.4) is calculated backward from the matrices F^H and Φ^H, i.e.,

$$P^H = \Phi^H F^H. \tag{3.43}$$

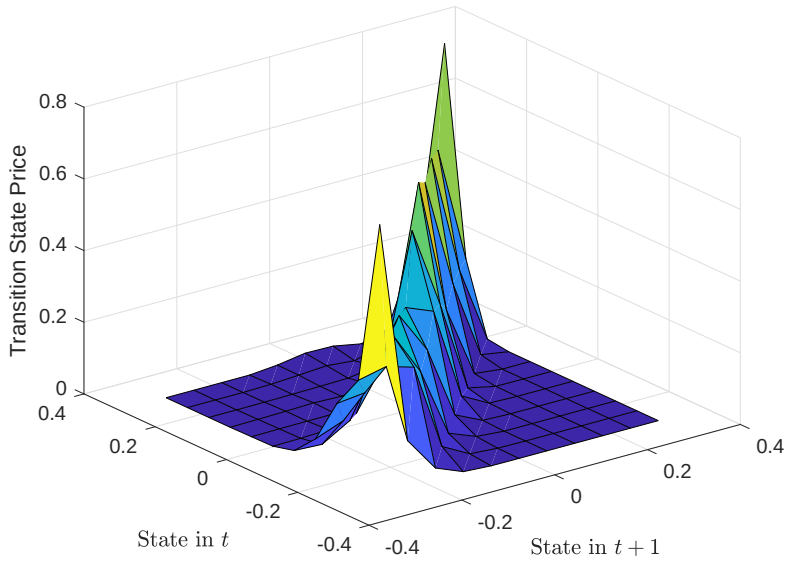


Figure 3.4: Hypothetical transition state price matrix, P^H.

3. A hypothetical current state price matrix S^H (see Figure 3.5) is calculated backward from the matrix P^H, i.e.,

$$S^H_{j+1} = \left(S^H_j\right)^\top P^H, \quad \forall j = 1, \ldots, m, \tag{3.44}$$

where $S^H_1 = \left(P^H_{i_0}\right)^\top$.

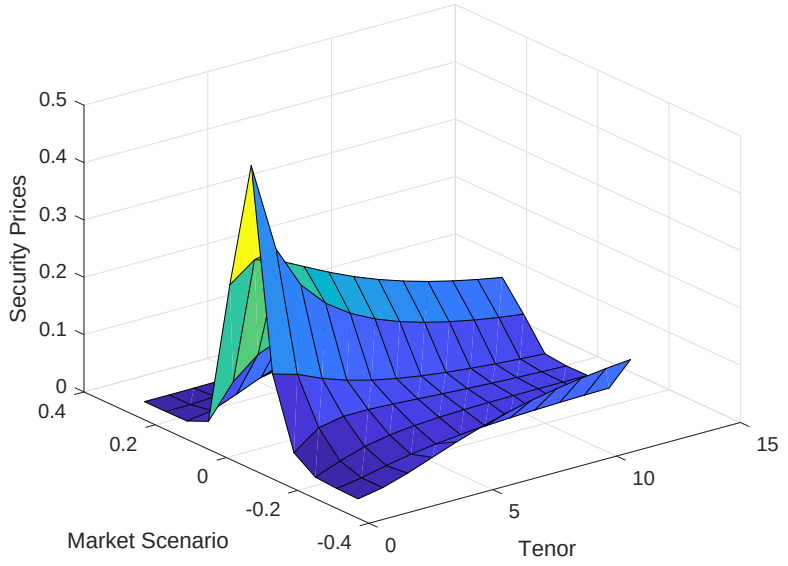


Figure 3.5: Hypothetical current state price matrix, S^H.

4. White noise is added to S^H to obtain S^N, as follows:

$$S^N_{i,j} = S^H_{i,j}(1 + e_{i,j}), \quad \forall i, j = 1, 2, \ldots, n, \tag{3.45}$$

where $e_{i,j} \sim N(0, \sigma)$.

5. Estimate P^N from S^N, using (3.16), (3.20), (3.27), (3.33), (3.36) and (3.40). In the case of the multivariate estimation methods, we will use a flat implied volatility, $\sigma^{(\mathrm{IV})}$, of 10%. We note that more accurate results could be achieved by modelling the behaviour of volatility and incorporating a forward-looking volatility structure than only looking at a flat or current volatility.

6. F^N is derived by applying the recovery theorem for each of the estimated matrices P^N.

7. The closer the estimated real-world distribution matrix F^N is to F^H, the more accurate the estimation process is.

Next, in order to measure how close the two distributions are we use the Kullback-Leibler Divergence test.

3.4.1 Kullback-Leibler Divergence

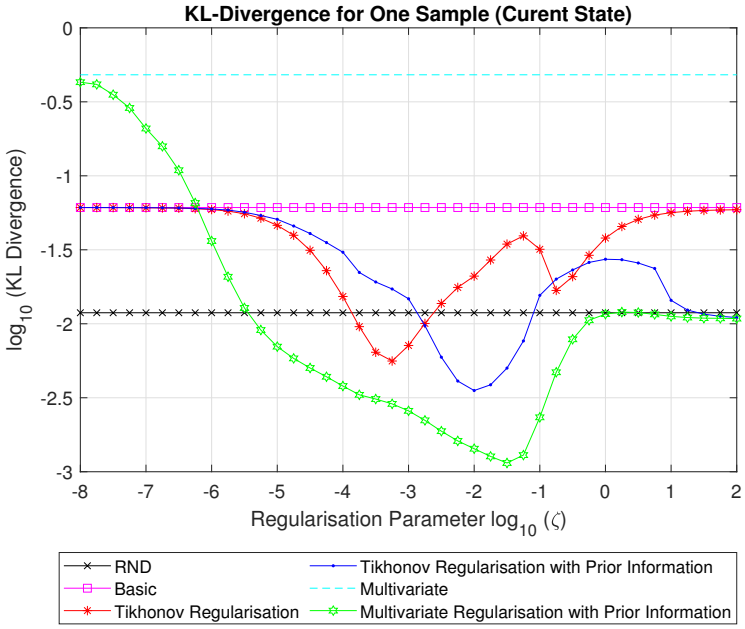
Intuitively, we would like to measure how close we can get back to F^H using S^N. Therefore, we will follow the same estimation accuracy method outlined in Kiriu and Hibiki (2019), namely, the Kullback-Leibler (KL) divergence test. The KL divergence test measures the difference between two distributions and is given as follows:

$$D_{KL}\left(F^N|F^H\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{i,j}^N \log\left(\frac{f_{i,j}^N}{f_{i,j}^H}\right). \tag{3.46}$$

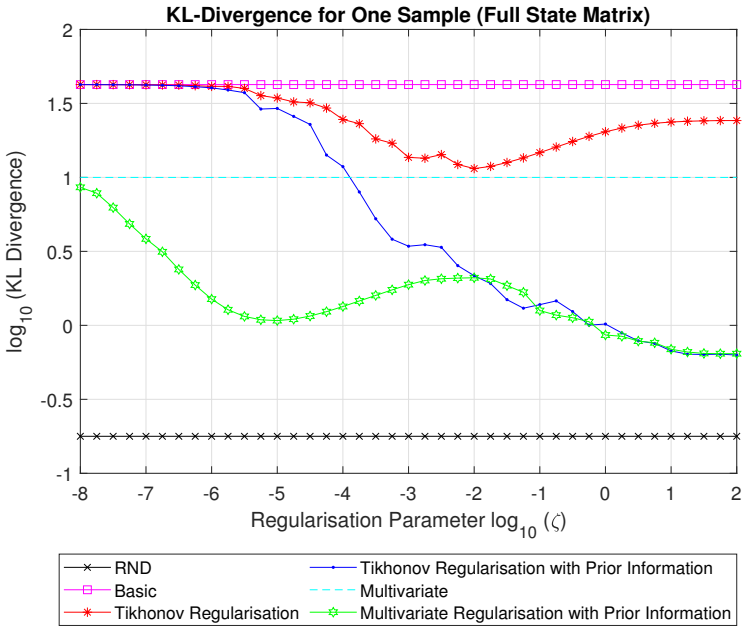
Obviously, when the estimated distribution and true distribution are exactly the same, the D_{KL} will equal zero.

In Figure 3.6, we show the log-log plots of the KL divergence at current state and at full state for five estimation methods discussed in this study for the regularisation parameter, $\zeta = 10^{-8}, 10^{-0.75}, \ldots, 10^{1.75}, 10^{2}$ and $\sigma = 5\%$. In addition, we also show the KL divergence for the risk-neutral distribution (RND). The RND, $\mathbb{Q}$, is the distribution obtained when using P^H in (3.4). Note that this is the best possible estimate for the RND as P^H is used. Therefore, obtaining a KL divergence less than the KL divergence for the RND will indicate that the estimation of the real-world distribution is more beneficial than the RND. For the current state (see Figure 3.6a), we found that both the basic method and the multivariate method provided a worse estimate of the real-world distribution than the RND. However, this is not the case for the estimation methods with the regularisation term (see Figure 3.6a).

The regularisation methods clearly outperform the non-regularised methods, where the multivariate regularisation method, proposed in this chapter, yielded the smallest KL divergence at current state. Similarly, the two regularisation methods with prior information clearly yield a lower KL divergence at full state compared with the basic, multivariate, and Tikhonov regularisation methods without prior information. However, the RND yielded the lowest KL divergence at full state (see Figure 3.6b). In Figure 3.7, we show that $h(\zeta)$ is a smooth and continuous function, where a minimum value can easily be estimated, making it an appealing selection function.



(a) KL divergence at current state.



(b) KL divergence at full state.

Figure 3.6: KL divergence of the real-world transition matrix.

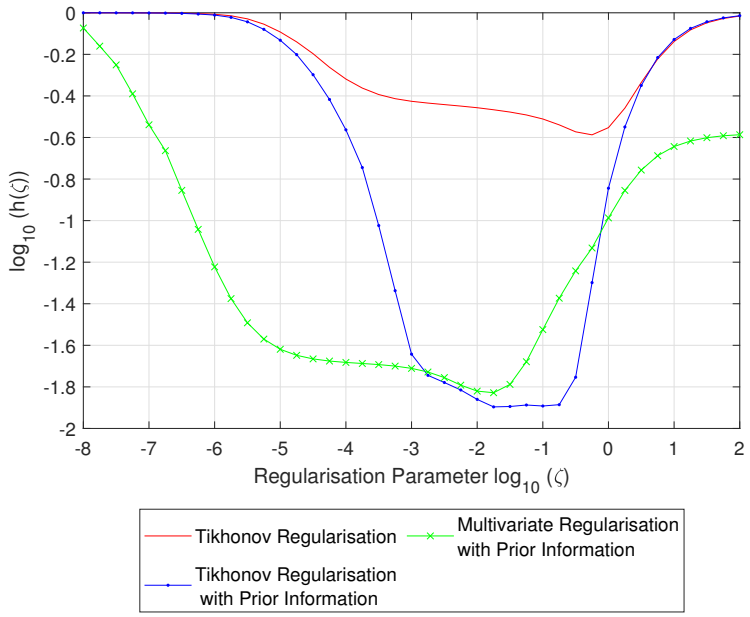


Figure 3.7: Optimisation of h_k.

Next, we examine the effectiveness of the estimation methods and the selection criteria when the regularisation term is added by carrying out 1000 Monte Carlo simulations. In Table 3.1, we show that the expected KL divergence and standard error for the 1000 Monte Carlo simulations for the current state, i.e., the i_0^{th} row vector of matrix F. More specifically, $\mathbb{E}(KL)$ represents the expected KL divergence, $\mathbb{E}(KL_{\min h_k})$ represents the expected KL divergence, where $h(\zeta)$ is a minimum, and $\mathbb{E}(\min KL)$ represents the minimum KL divergence across all ζ. We can see that the RND provides a better estimation, with a lower KL divergence, than the basic and multivariate estimation methods (as seen in Figure 3.6a). This is a direct consequence of the ill-posed problem, when solving (3.16).

The regularised methods clearly outperformed the RND, basic, and multivariate methods, indicating the strength of adding the regularising term when solving ill-posed problems. More specifically, the multivariate regularised method and the elastic net method, proposed in this chapter, yielded the best results with the lowest expected KL divergence. In all cases the standard errors are small indicating the estimation methods provide stable estimates. However, it must be noted that the elastic net method is significantly more computationally expensive than the other methods discussed in this study.

Table 3.1: KL divergence at current state matrix.

Method		$\sigma = 5\%$	$\sigma = 10\%$
RND	$\mathbb{E}(KL)$	0.0119	0.0119
Basic (3.16)	$\mathbb{E}(KL)$	0.1931	0.2155
	SE	0.0067	0.0073
Multivariate (3.20)	$\mathbb{E}(KL)$	0.5908	0.7702
	SE	0.0038	0.0168
Regularised* (3.27)	$\mathbb{E}(KL_{\min h_k})$	0.0335 (0.5378)	0.0344 (0.4674)
	SE	0.0002 (0.0049)	0.0004 (0.0046)
	$\mathbb{E}(\min KL)$	0.0101 (0.0256)	0.0165 (0.0640)
	SE	0.0002 (0.0020)	0.0003 (0.0028)
Regularised with Prior Information* (3.33)	$\mathbb{E}(KL_{\min h_k})$	0.0124 (0.0179)	0.0494 (0.0154)
	SE	0.0003 (0.0007)	0.0017 (0.0006)
	$\mathbb{E}(\min KL)$	0.0061 (5.9265)	0.0111 (41.3819)
	SE	0.0001 (0.7453)	0.0001 (1.5536)
Multivariate Regularised with Prior Information* (3.36)	$\mathbb{E}(KL_{\min h_k})$	0.0082 (0.0156)	0.0532 (0.0253)
	SE	0.0005 (0.0002)	0.0051 (0.0005)
	$\mathbb{E}(\min KL)$	0.0034 (0.0747)	0.0072 (6.7491)
	SE	0.0001 (0.0020)	0.0002 (0.7843)
Elastic Net* (3.38)	$\mathbb{E}(\min KL)$	0.0031 (0.0178)	0.0062 (0.0562)
	SE	0.0001 (0.0000)	0.0001 (0.0000)

*The $\mathbb{E}(\zeta)$ and SE_ζ are displayed in parenthesis

Similarly, in Table 3.2 we show that the expected KL divergence and standard error for the entire F matrix. We see that the multivariate method yields a smaller KL divergence than the basic and regularised method proposed by Audrino et al. (2014). However, the methods that are regularised with prior information still yielded the lowest KL divergence, with the multivariate regularised method yielding the lowest expected KL divergence, where $h(\zeta)$ is a minimum. Furthermore, the elastic net method yielded the lowest KL divergence across all regularisation parameters. However, the elastic net method is substantially more computationally expensive than the multivariate regularised method and, therefore, will not be studied any further in this thesis.

Table 3.2: KL divergence at full state matrix.

Method		$\sigma = 5\%$		$\sigma = 10\%$	
RND	$\mathbb{E}(KL)$	0.1779		0.1779	
Basic (3.16)	$\mathbb{E}(KL)$	43.7850		58.3776	
	SE	0.2718		0.3417	
Multivariate (3.20)	$\mathbb{E}(KL)$	10.3142		10.8718	
	SE	0.0136		0.0580	
Regularised* (3.27)	$\mathbb{E}(KL_{\min h_k})$	18.9698	(0.5378)	18.6056	(0.4674)
	SE	0.0229	(0.0049)	0.0268	(0.0046)
	$\mathbb{E}(\min KL)$	11.3012	(0.0055)	12.5754	(0.0162)
	SE	0.0270	(0.0002)	0.0308	(0.0005)
Regularised with Prior Information* (3.33)	$\mathbb{E}(KL_{\min h_k})$	2.3659	(0.0179)	3.7873	(0.0154)
	SE	0.0236	(0.0007)	0.0582	(0.0006)
	$\mathbb{E}(\min KL)$	0.6828	(75.5576)	0.7064	(65.1735)
	SE	0.0010	(1.1127)	0.0022	(1.2155)
Multivariate Regularised with Prior Information* (3.36)	$\mathbb{E}(KL_{\min h_k})$	2.2282	(0.0156)	3.6187	(0.0253)
	SE	0.0190	(0.0002)	0.0564	(0.0005)
	$\mathbb{E}(\min KL)$	0.6853	(87.3217)	0.7130	(87.2990)
	SE	0.0010	(0.6783)	0.0021	(0.6926)
Elastic Net* (3.38)	$\mathbb{E}(\min KL)$	0.6567	(49.7875)	0.6758	(54.0514)
	SE	0.0010	(0.9809)	0.0020	(1.2007)

*The $\mathbb{E}(\zeta)$ and SE_ζ are displayed in parenthesis

It is evident from the above that the multivariate regularised method introduced in this chapter improved the estimation of the real-world distribution. It must also be noted that the further the row, in the state transition matrix, is from the current state's row (i.e., normally defined as the middle row), the more difficult it is to determine, but also the less influential it is on the real-world distribution (see, Backwell, 2015). Therefore, the transition from the current state is of greater interest in this study as we are mostly interested in how the asset would change over one period given today's state. In the next section, we conduct a robust check.

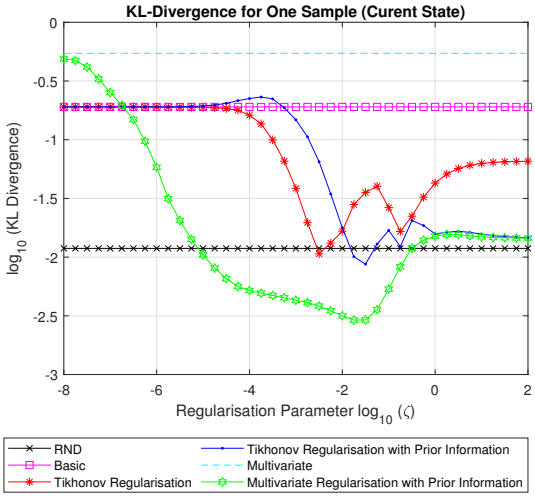
3.4.2 Robustness Check

In this section, we conduct a robustness check by using different hypothetical data obtained from the real-world distribution used above Kiriu and Hibiki (2019). More specifically, Figures 3.8a-3.8b shows the KL divergence where $\delta = 0.995$, Figures 3.8c-3.8d shows the results for

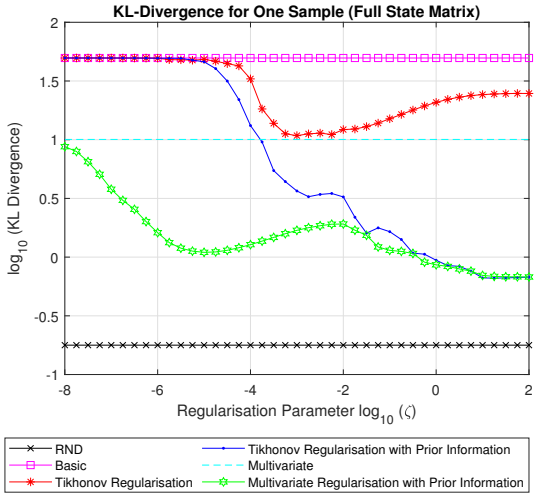
a large risk aversion parameter, namely, $\gamma = 10$ and lastly Figures 3.8e-3.8f shows the KL divergence using the CARA utility function, i.e.,

$$\phi_{i,j} = \delta e^{-\gamma(r_j - r_i)}, \quad i, j = 1, \ldots, n \tag{3.47}$$

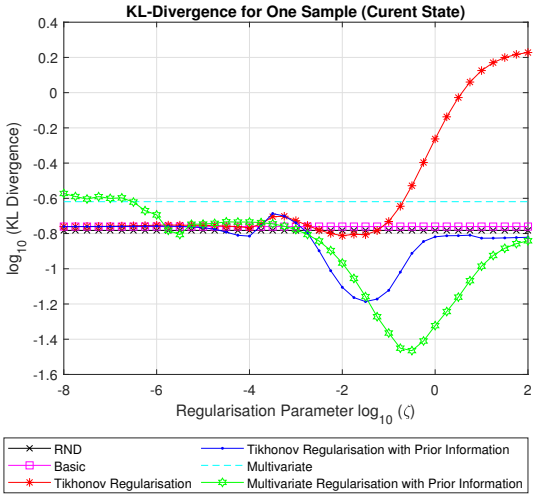
with $\gamma = 3$, instead of CRRA utility function. The results obtained in Figure 3.8 shows that the multivariate regularised method, proposed in this chapter, yields a robust estimate of the real-world distribution. Furthermore, we carry out the robust check on the South African Top40 index, where we obtained similar results (see Figures 3.8g-3.8h).



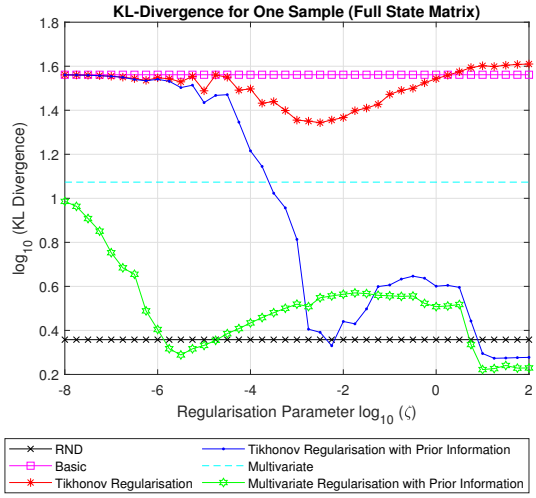
(a) current state: $\delta = 0.995$



(b) full state: $\delta = 0.995$

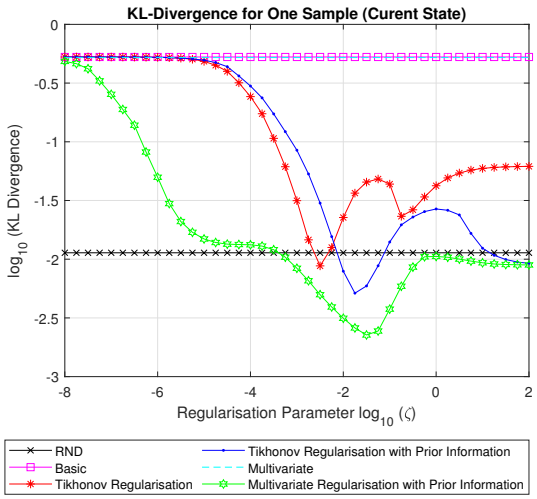


(c) current state: $\gamma_R = 10$

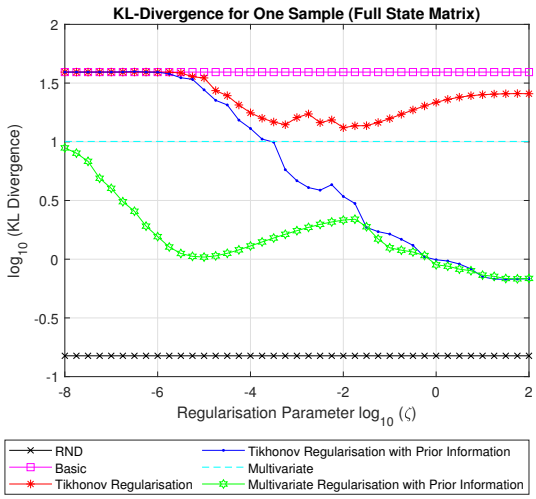


(d) full state: $\gamma_R = 10$

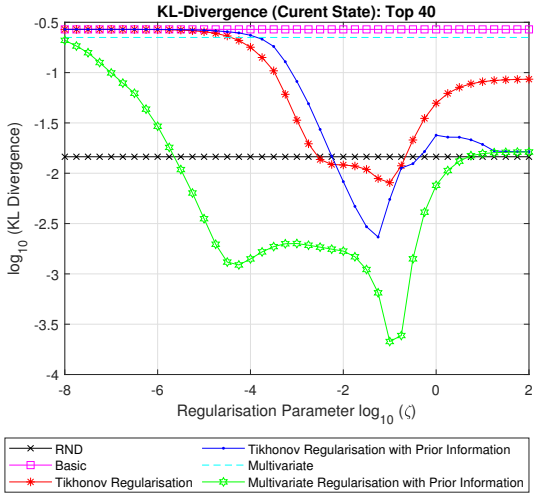
Figure 3.8: KL divergence: robust test.



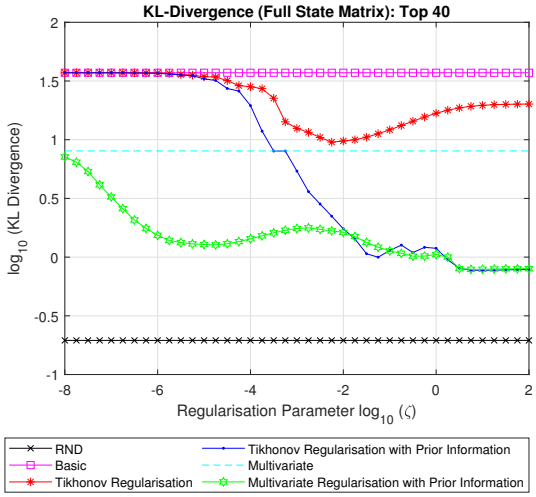
(e) current state: CARA utility



(f) full state: CARA utility



(g) current state: Top40



(h) full state: Top40

Figure 3.8: KL divergence: robust test (cont.).

We note that other norms, such as, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ could be used to estimate P more accurately. Chvátal (1983) asserts that when estimating linear functions, $\|\cdot\|_1$ gives the most robust answer, $\|\cdot\|_\infty$, avoids gross discrepancies with the data, and if the errors are known to be normally distributed then $\|\cdot\|_2$ is the best choice. However, in our analysis, we found that the Euclidean norm yielded the most accurate and stable results. In the next section, we conduct an empirical study.

3.5 Empirical Results

In this section, we compare some distributional properties of the risk-neutral and real-world distributions by using the weekly Top40 option trade data, traded on the South African Futures Exchange (SAFEX). We start by using weekly arbitrage-free implied volatility surfaces to estimate the risk-neutral distribution over the period 5 September 2005 - 15 January 2018. Furthermore, we used the SVI model to interpolate over the fixed domain $\psi \in [0.5, 1.5]$, where ψ is defined as the spot moneyness (i.e., $\psi = K/S_0$) and $T \in [1, 12]$ as outlined in Flint and Maré (2017). The evolution of the weekly one-month percentiles and mean of the risk-neutral distribution is shown in Figure 3.9.

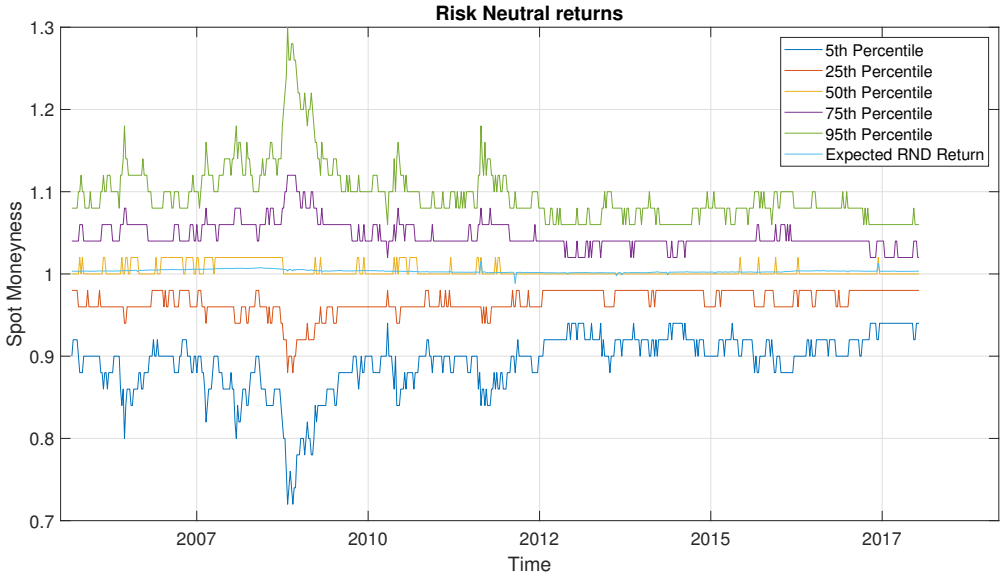
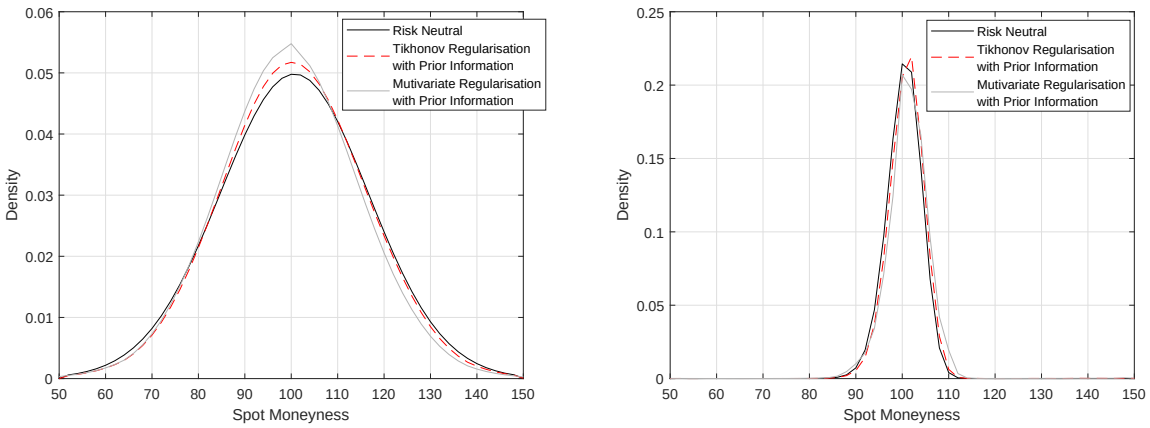


Figure 3.9: Weekly one-month percentiles of the risk-neutral Top40 distributions, 05 Sep 2005 - 15 Jan 2018.

As expected the risk-neutral distribution widened over the global financial crisis (2008-2009) and has since narrowed considerably. Next, we estimated the transition probability matrix, P, using the methods proposed by Kiriu and Hibiki (2019) and the regularised multivariate method with prior information. Thereafter, we applied the recovery theorem. In Figure 3.10, we show how the risk-neutral and recovered real-world distributions widened during the financial crises (03 November 2008) compared to the risk-neutral and real-world distributions after the financial crisis (15 January 2018).

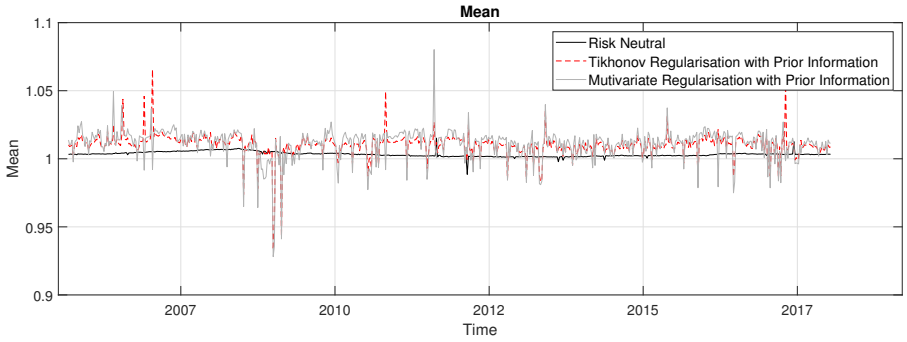


(a) 03 Nov. 2008 (b) 15 Jan. 2018

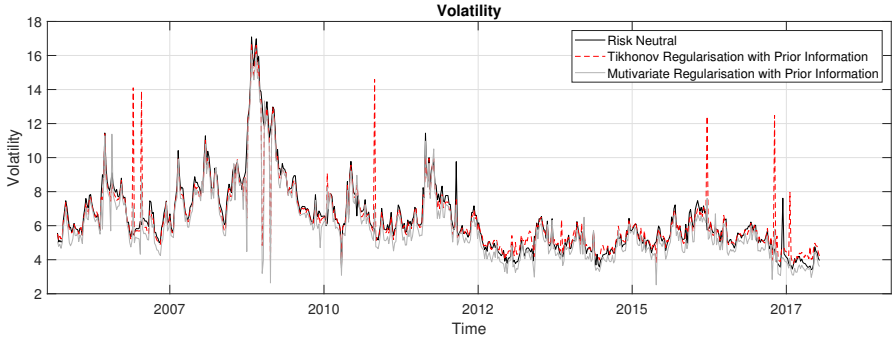
Figure 3.10: Risk-neutral and real-world distributions.

In Figure 3.11, we show the evolution over time of the weekly one-month first four moments. We see that the expected returns of the two real-world distributions are mostly above the risk-neutral distributions expected returns, except during the financial crisis. The volatility has steadily decreased since the global financial crisis (a peak of approximately 17% down to 4%). In addition, the real-world distribution obtained by using the regularised multivariate Markov chain with prior information showed a lower volatility than the distributions obtained using the univariate regularised method with prior information and risk-neutral (which showed similar volatility). This is somewhat expected, since controlling the volatility in the multivariate regression model provided us with a better sense of future economical uncertainty (see, e.g., Sanford, 2018). The skewness for the risk-neutral distribution became less negative during the financial crisis along with a drop in kurtosis. The skewness has since reverted to a skewness around -0.5 along with an increase in kurtosis. In addition, the weekly skewness coefficients for the real-world distributions showed sharp spikes (became positively skewed) in 2012 and 2016.

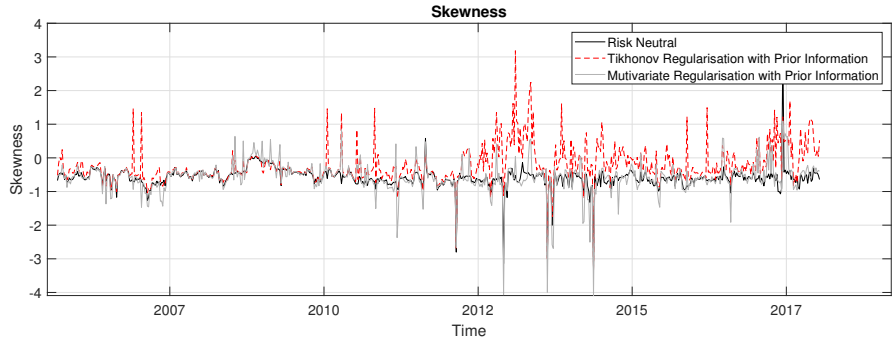
In Table 3.3, we show the mean and volatility for the Top40 index with the first four moments of the risk-neutral and real-world distributions. The recovered moments estimated from option prices clearly provides insight above the risk-neutral moments. Furthermore, we found that the recovered kurtosis of the real-world distribution using the Tikhonov regularisation method with prior information was considerably more volatile over time than the multivariate regularisation method with prior information and the RND.



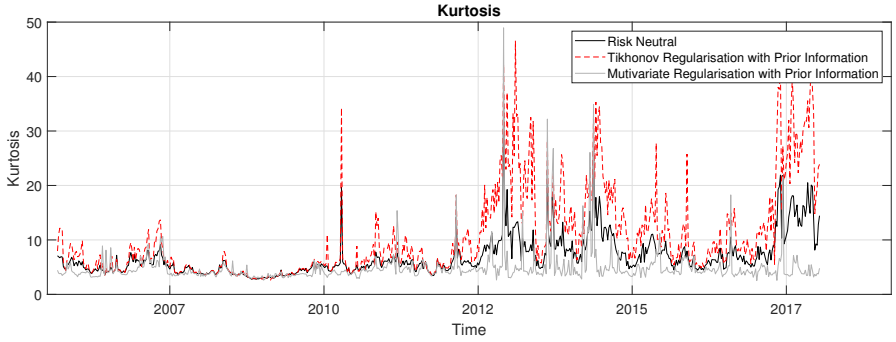
(a) Mean Return



(b) Volatility



(c) Skewness



(d) Kurtosis

Figure 3.11: Top40 weekly one-month moments.

Table 3.3: Top40 weekly one-month moments.

	Mean	**Volatility**
Top40 Returns*	28.87%	20.20%
Risk-Neutral Distribution		
Expected Return*	4.07%	2.20%
Volatility on Return*	22.38%	7.34%
Skewness	-0.59	0.32
Kurtosis	6.89	3.67
Real-World Distribution: Tikhonov Regularisation with Prior Information		
Expected Return*	13.17%	11.09%
Volatility on Return*	22.32%	6.98%
Skewness	-0.23	0.54
Kurtosis	10.49	8.08
Real-World Distribution: Multivariate Regularisation with Prior Information		
Expected Return*	15.05%	14.19%
Volatility on Return*	20.62%	7.18%
Skewness	-0.58	0.46
Kurtosis	4.87	3.32

*Values are annualised

The predictive information obtained using the recovery theorem along with real-world data surely yielded some insight into the markets subjective probabilities. However, the true practicality and usefulness of the model remains elusive in the literature.

3.6 Conclusion

The recovery theorem is a remarkable theorem that allows us to estimate the real-world distribution from the risk-neutral distribution. However, the implementation of the recovery theorem requires the solution of two ill-posed problems.

The first is estimating the state price matrix by calculating the second partial derivative of the option price with respect to the strike. This is especially problematic in noisy and sparse markets. Flint and Maré (2017) proposed an algorithm for this first ill-posed problem.

The second entails the estimation of the transition price matrix that captures the state price dynamics. Audrino et al. (2014) and Kiriu and Hibiki (2019) used a regularisation technique to

obtain a stable transition matrix. In addition, Audrino et al. (2014) and Flint and Maré (2017) showed by empirical work that there is information contained in the recovered distributions.

In this study, we investigated several estimation methods to accurately estimate the transition price matrix. The accuracy of the estimated transition matrix has a significant impact on the estimation of the real-world distribution implied from option prices using the Ross recovery theorem. In addition, we presented a regularised multivariate Markov chain with prior information to estimate the transition matrix. This is a first attempt to regularise the multivariate Markov chain for the recovery theorem. In our analysis, we found that the regularised multivariate Markov chain method improved upon the estimation of the real-world distribution. Furthermore, we conducted an empirical study using weekly South African Top40 option trade data to estimate the risk-neutral and real-world distributions.

CHAPTER 4

FORWARD-LOOKING DISTRIBUTIONS WITH APPLICATION TO RISK MANAGEMENT

This chapter is adapted from Van Appel and Maré (2020b) and presents various extensions that addresses the second research question.

Chapter Synopsis

The forward-looking nature of option prices provides an appealing way to extract risk measures. In this chapter, we extract forecast distributions, using the methods studied in Chapters 2 and 3, from market-quoted option prices that we then use to forecast risk measures. In addition, we backtest and compare the predictive power of the real-world return distribution forecasts with the risk-neutral return distribution forecasts, implied from option prices, and the historical return distribution forecasts. In an empirical study, using the South African FTSE/JSE Top 40 (Top40) index, we show that the extracted real-world distribution forecasts, especially the method using the recovery theorem and the method using the non-parametric calibration transformation function, yield satisfying forecasts of risk-measures. Furthermore, in order to improve the forecast accuracy, we introduce the mixing of historical and option-implied distributions.

4.1 Introduction

John Maynard Keynes said that "*successful investing is anticipating the anticipations of others*". In essence, financial derivative securities are forward-looking and essentially embed information about investors' beliefs about the distribution of asset returns (see, e.g., Christoffersen et al., 2013; Hollstein et al., 2019; Dillschneider and Maurer, 2019). Investors, policy makers, and risk

managers therefore look at market variables (or derivations thereof) aiming to gauge forecasts of economic variables or sentiment (or changes thereof) (see, Bliss and Panigirtzoglou, 2004).

Financial derivative securities are frequently used to infer information. The prime example is the VIX index, which is derived from the prices of equity index options traded on the Chicago Board of Options Exchange (CBOE). This index reflects the market's view of 30-day index volatility and is used as a risk-sentiment gauge by investors. Bollerslev et al. (2009) showed that the difference between the VIX and the realised volatility on the S&P500 index carries significant explanatory power for future equity returns. Moreover, the VIX index is calculated using a model-free approach, illustrating the effectiveness of model-free approaches in the literature (see, e.g., CBOE, 2009; Christoffersen et al., 2013). A more recent innovation by the CBOE is the SKEW index, which reflects the index option market's perceptions of so-called tail risk (see, e.g., CBOE, 2011; Christoffersen et al., 2013).

The ability to accurately forecast future asset prices is an important and frequently studied problem in financial economics (see, e.g., Bollerslev et al., 2009; Crisóstomo and Couso, 2018). The recent global financial crisis highlighted this problem, where many conventional financial theories were unable to realistically forecast risk measures. Recent studies have shown that option-implied moments, such as the VIX and SKEW, have predictive power for the realised variance (see, e.g., Hollstein et al., 2019).

Forecasts of the option-implied return density can provide risk managers with more information than forecasts of the moments alone (see, Barone-Adesi, 2016). Such measures of risk include Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), which are two popular measures of risk used by financial practitioners and regulators, which is related to a quantile of the return distribution. More specifically, VaR is a single value summarising the potential loss of a financial asset (or portfolio). In percentage terms this corresponds to the α-th percentile of the asset return distribution and CVaR is a measure of tail-risk, which measures "how bad things could get" (see, e.g., McNeil et al., 2005). That is,

$$VaR_{(1-\alpha)} = F^{-1}(\alpha) \tag{4.1}$$

and CVaR for the discrete case is defined as

$$\text{CVaR}_{(1-\alpha)} = \mathbb{E}[R|R \leq \text{VaR}_{(1-\alpha)}], \tag{4.2}$$

where $1-\alpha$ is the confidence level, $F(R)$ the return forecasted cumulative distribution function, and $R = S_T/S_0$ the random variable representing the asset return from time zero to time T.

Many traditional strategies of measuring VaR rely on a parametric return density, such as the normal density, and past (historical) data to make market assumptions (see, e.g., McNeil et al., 2005). In practice, financial returns exhibit skewness and kurtosis that are not captured in the normal assumption framework (see, e.g., Cont, 2001). Consequently, this has rekindled great interest in fat-tail distributions (see, e.g., Hull and White, 1998b). In contrast to using historical data, one can also make use of market-quoted option prices to extract a forward-looking risk-neutral return density forecast (see, e.g., Barone-Adesi, 2016; Breeden and Litzenberger, 1978). The purported forward-looking nature of option prices makes it conceptually better suited for forecasting than a historical scheme, especially during stressed economic environments.

In particular, historical simulation and risk-neutral methods are the most widely used methods in financial risk management, where most financial institutions prefer to use historical simulations to manage risk (see, Pérignon and Smith, 2010). However, Christoffersen et al. (2013) and Crisóstomo and Couso (2018) found that methods based on option-implied information generally outperformed historical-based estimates. Similarly, Shackleton et al. (2010) compared the real-world option-implied densities to that of historical densities, where they found that the real-world option-implied forecasts for two- and four-week horizons were superior to that of the historical forecasts. The transformation from risk-neutral to real-world return densities have been studied in several papers (see, e.g., Bakshi et al., 2003; Bliss and Panigirtzoglou, 2004; Shackleton et al., 2010; Ross, 2015). More specifically, the recovery theorem, proposed by Ross (2015), is a model-free method that extracts a real-world return density forecast from option prices.

The aim of this chapter is to (i) extract, backtest, and compare the real-world return distribution forecast models to the historical and risk-neutral distribution forecast models studied in this thesis; and (ii) backtest the tail of the return distribution forecast models for risk management purposes and show that one can extract reliable risk measures using option-implied data and the recovery theorem. Moreover, research into the forecasting ability of real-world return distributions are scarce in the literature. Furthermore, it is likely that the entire return distribution forecast may be misspecified, but performs better in certain regions of the distribution, such as the tail, which is often more valuable to risk managers (see, Berkowitz, 2001).

The remainder of this chapter is structured as follows. Section 4.2 studies some commonly used backtesting approaches found in the literature. Section 4.3 analyses the forecasts models studied in in this thesis by carrying out an empirical study using the South African FTSE/JSE Top 40 (Top40) index. In addition, the performance of five forward-looking real-world return distribution forecast methods will be backtested and benchmarked, using the Top40 index, against two historical simulated return distribution forecast methods and three risk-neutral return distribution forecast models, with a specific focus on risk management. Furthermore, in Section 4.4, we study the mixture of historical distributions and forward-looking distributions to obtain a more accurate forecast model. Lastly, in Section 4.5 we extract the risk-premium, which provides useful information about the market sentiment that can be used in portfolio management.

4.2 Verification of the Distribution Forecasts

In principle, a good forecast distributions should coincide with the true return distribution of the asset or portfolio under study (see, Knüppel, 2015). The aim of this section is to introduce methods to verify the accuracy of the return distribution forecast models introduced in Chapters 2 and 3. In practice, it is highly unlikely that an optimal model will exist as the true distribution may be too complicated to be represented by a simple mathematical model, or might not be adequately represented over all economic periods. Therefore, each model can only be considered an approximation of the truth. In order to assess whether (i) the real-world return distribution forecast models approximate the truth better than the simple historical simulation or risk-neutral models; and (ii) under which circumstances it can approximate the truth better, we introduce some commonly used forecast evaluation tests found in the literature, with a specific application to risk management.

Interval forecasts such as VaR are based on the inverse distribution function,

$$\bar{y}_t = F^{-1}(\alpha), \tag{4.3}$$

where, for example, $\alpha = 0.05$ for the $\text{VaR}_{(0.95)}$. Christoffersen (1998) asserted that the interval should be exceeded $\alpha\%$ of the time and the violations should be uncorrelated across time.

Combining these properties, the hit function

$$\mathbb{I}_t = \begin{cases} 1 & \text{if violation occurs,} \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

should be an independent and identically distributed (i.i.d.) Bernoulli sequence with parameter α. In a VaR setting, the Bernoulli variable rarely takes on the value 1, requiring a large number of sample observations to test the distribution forecast. In contrast, Rosenblatt (1952) proposed a transformation of the observed realisations into a series of i.i.d. random variables as follows:

$$x_t = \int_{-\infty}^{y_t} \hat{f}_t(u)\,\mathrm{d}u = \hat{F}_t(y_t), \tag{4.5}$$

where y_t is the *ex-post* return realisation and $\hat{f}(\cdot)$ is the *ex-ante* return distribution forecast[1]. More specifically, he showed that x_t is i.i.d. uniform on $(0,1)$. This procedure, also commonly known as the probability integral transform (PIT), allows for a wider variety of tests to be conducted. Furthermore, this result is valid irrespective of the underlying distribution of returns, y_t, and remains valid even when the forecast model, $\hat{F}(\cdot)$, changes over time. A series of forecast evaluation tests using graphical displays were proposed by Diebold and Mariano (1995). In contrast, Berkowitz (2001) proposed a series of likelihood ratio (*LR*) tests for model evaluation by generating a sequence $z_t = \Phi^{-1}(x_t)$ for the given model, where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative standard normal distribution function. Then, by definition, z_t should be independent across variables with standard normal distribution. This second transformation allows for convenient tests that are associated with the Gaussian likelihood function. In particular, Berkowitz (2001) jointly assesses the mean (μ), variance (σ^2), and serial correlation (ρ) by testing the null hypothesis that z_t are i.i.d. $N(0,1)$ distributed against the following first-order autoregressive model with mean and variance other than $(0,1)$:

$$z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t. \tag{4.6}$$

The log-likelihood function of (4.6) is often seen in statistics and is reproduced below for convenience (see, Berkowitz, 2001):

$$\begin{aligned} \ell(\mu, \sigma^2, \rho | z) = &-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{(z_1 - \mu/(1-\rho))^2}{2\sigma^2/(1-\rho^2)} \\ &- \frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \sum_{t=2}^{T}\left[\frac{(z_t - \mu - \rho z_{t-1})^2}{2\sigma^2}\right], \end{aligned} \tag{4.7}$$

[1] Since the RND, RWD and RWD-M are recovered on a discrete grid, where the future realised returns are not likely to fall on one of the state grid points, we linearly interpolate the recovered cdf to obtain x_t (see, Jackwerth and Menner, 2020).

where σ^2 is the variance of ε_t.

Firstly, Berkowitz (2001) uses (4.7) to test for independence by considering the following *LR* test statistic:

$$LR_{\text{ind}} = -2\left[\ell\left(\hat{\mu},\hat{\sigma}^2,0\right) - \ell\left(\hat{\mu},\hat{\sigma}^2,\hat{\rho}\right)\right]. \tag{4.8}$$

Under the null hypothesis, (4.8) is distributed $\chi^2(1)$. More specifically, this test statistic is a measure of the degree to which the data support a non-zero persistent parameter. Secondly, he also tests the null hypothesis that not only are the observations independent, but also have mean and variance equal to 0 and 1 respectively, using the following *LR* test statistic:

$$LR = -2\left[\ell(0,1,0) - \ell\left(\hat{\mu},\hat{\sigma}^2,\hat{\rho}\right)\right]. \tag{4.9}$$

Under the null hypothesis, (4.9) is distributed $\chi^2(3)$. For multi-step-ahead forecasts, practitioners use the following test statistic instead (see, Knüppel, 2015):

$$LR_{\text{MS}} = -2\left[\ell\left(0,\sqrt{1-\hat{\rho}^2},0\right) - \ell\left(\hat{\mu},\hat{\sigma}^2,\hat{\rho}\right)\right], \tag{4.10}$$

which is distributed $\chi^2(2)$. More specifically, multi-step-ahead forecasts are complicated with serial correlation of the outcomes with respect to the density forecast. That is, for example, if the true return turns out to be higher than our one-month forecast from today, then it is likely that the true one-month return for the next week's forecast will also be higher than the forecasted return. Therefore, (4.10) is particularly useful for density forecast evaluation for practitioners.

It must be noted that a density forecast model may be falsely rejected as it does not forecast well for particular regions of the distribution. It is possible that a forecast model performs poorly in forecasting expected returns, but performs better in predicting a certain region of the distribution, such as the tail of the distribution. Thirdly, cognisant of this, Berkowitz (2001) introduced a *LR* test that intentionally ignores model failures in the interior of the distribution and compares the lower tail of the foretasted density with the observed density by truncating any observed values that fall outside the tail area. Let this cut-off point be $\text{VaR} = \Phi^{-1}(\alpha)$. The new variable of interest, z_t^*, is then defined as:

$$z_t^* = \begin{cases} \text{VaR} & \text{if} \quad z_t \geq \text{VaR} \\ z_t & \text{if} \quad z_t < \text{VaR}. \end{cases} \tag{4.11}$$

The log-likelihood function for the lower tail is given as (see, Berkowitz, 2001):

$$\ell(\mu,\sigma|z^*) = \sum_{z_t^*<\text{VaR}} \left(-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(z_t^*-\mu)^2\right) + \sum_{z_t^*=\text{VaR}} \log\left(1-\Phi\left(\frac{\text{VaR}-\mu}{\sigma}\right)\right), \tag{4.12}$$

where the first two terms represent the usual Gaussian likelihood of losses and the third term is a normalisation factor arising from the truncation. As before a *LR* test is constructed with null hypothesis $\mu = 0$ and $\sigma^2 = 1$ against the unrestricted alternative with mean and variance other than 0 and 1 respectively, i.e.,

$$LR_{\text{tail}} = -2\left[\ell(0,1) - \ell\left(\hat{\mu},\hat{\sigma}^2\right)\right]. \tag{4.13}$$

Under the null hypothesis that the model is correct, the test statistic is distributed $\chi^2(2)$. In addition, Berkowitz (2001) showed, by using a Monte Carlo simulation, that these *LR* tests are powerful, even for samples containing only as few as 100 observations.

In summary, a well-specified model should simultaneously pass as many statistical backtests as possible. Therefore, in Appendix A, we briefly list additional backtests, which form part of the MATLAB Risk Management Toolbox (2018).

4.3 Application

In this section, we used weekly Top40 option trade data and the Top40 index prices to construct weekly one-month return distribution forecasts for the Top40 index over the period 05 September 2005 to 15 January 2018, giving us a total of 646 weekly one-month return distribution forecasts. The Top40 index is particularly useful as the underlying risky asset, as it is a key risk factor in the economy and is amongst the most liquid traded derivative contracts in the South African market. In particular, Carr and Madan (2000) showed that a major financial market index, such as the Top40 index, could be used as a proxy to price options on individual stocks that are illiquid. This makes the Top40 index an important market factor to illustrate the accuracy of forecast models. The number of weeks that a return distribution forecast was made for each subset of the time series considered in this study is shown in Table 4.1. The extracted distribution forecast models in this study are: (i) historical simulation, (ii) historical

Table 4.1: Market data.

Panel A: Monthly one-Month returns

Time Period	Label	Number of weeks (N)
Sep 2005 - Jan 2018	Full-period	161

Panel B: Weekly one-Month returns

Time Period	Label	Number of weeks (N)
Sep 2005 - Dec 2007	Pre-crisis	122
Jan 2008 - Dec 2009	Crisis	104
Jan 2010 - Jan 2018	Post-crisis	420
Sep 2005 - Jan 2018	Full-period	646

simulation with volatility updating, (iii) model-free risk-neutral density extracted from option prices (RND), (iv) risk-neutral density extracted using the Heston model, (v) risk-neutral density extracted using the Bates model, (vi) real-world density extracted using the parametric calibration specification (Heston P1), (vii) real-world density extracted using the non-parametric calibration specification (Heston P2), (viii) real-world density extracted using the behavioural approach (Behavioural), (ix) real-world distribution extracted using the recovery theorem that is regularised with prior information (see, (3.28)) (RWD), and (x) real-world distribution using the recovery theorem extracted using the regularised multivariate model with prior information (see, (3.36)) (RWD-M).

For the historical simulation methods we used a five-year historical period and for the historical-HW approach we used the EWMA model with $\alpha = 0.94$ for the volatility updating process (see, Hull and White, 1998a). For the forecast densities we extracted a 50-150% moneyness range. Similar to Cristóstomo (2021), we found a power utility function with a constant risk aversion parameter of $\gamma = 2$ and high sentiment effects delivered the best results for the behavioural distribution's forecasts, and therefore will be reported in this study. For the real-world forecast distribution using the recovery theorem, we constructed a 51×51 one-month ahead transition probability matrix, P, spanning a 50-150% moneyness range, which is placed every 2% symmetrically around the moneyness of 100%. Recall, the one-month ahead forecast from today's state will correspond to the centre row of the real-world transition probability matrix, F. The performance of these distribution forecasts are evaluated using the verification tests discussed in Section 4.2.

In testing the consistency between the *ex-ante* return distribution scheme and the observed

return realisation, we firstly, used Rosenblatt's PIT to transform the realisation of returns to a series of i.i.d. uniform random variables. Thereafter, we made the second transformation, proposed by Berkowitz (2001), to a realisation of i.i.d. standard normal random variables. The empirical cdf versus the standard normal cdf for each method during the global financial crisis is shown in Figure 4.1, where it can be seen that the random variable z_t deviates from the standard normal distribution for both historical simulation methods.

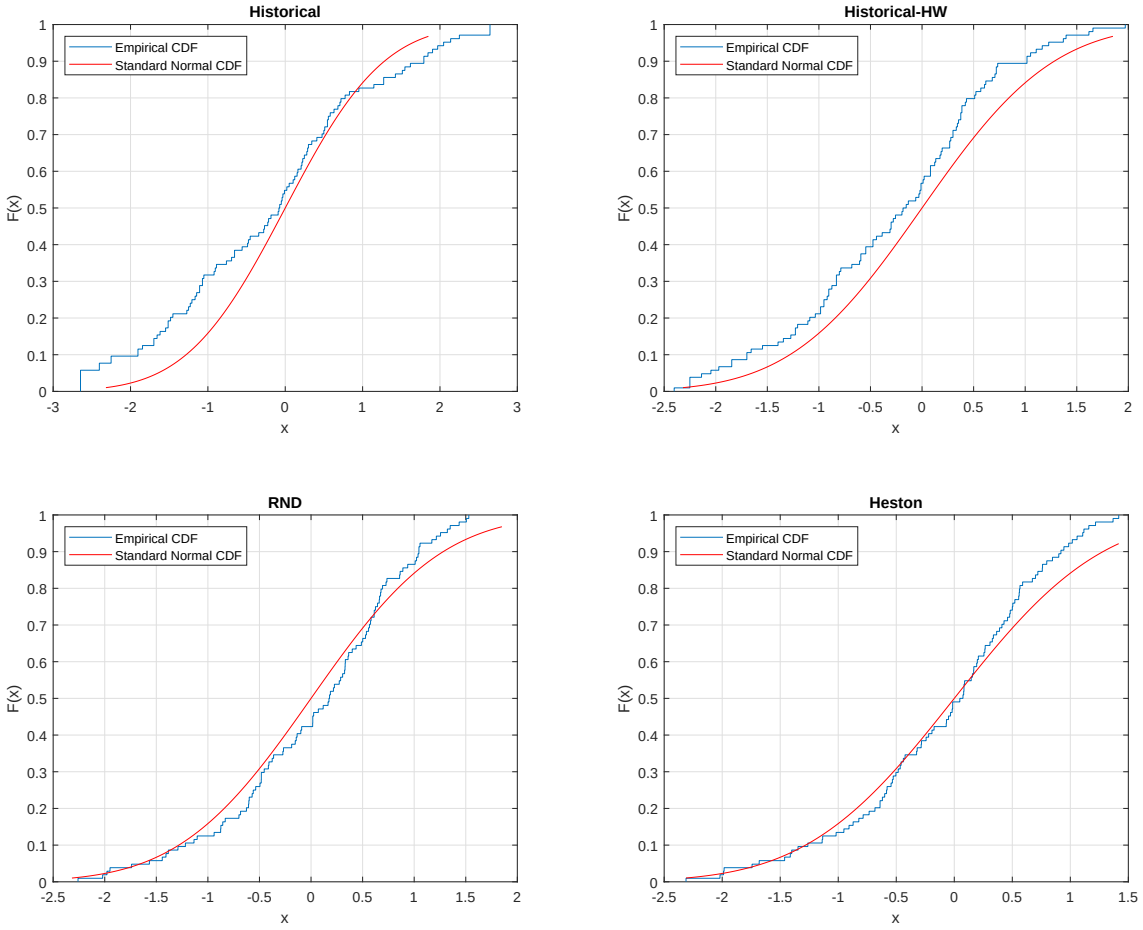


Figure 4.1: Crisis period (Jan 2008 - Dec 2009): empirical cdf and normal cdf (continued on next page).

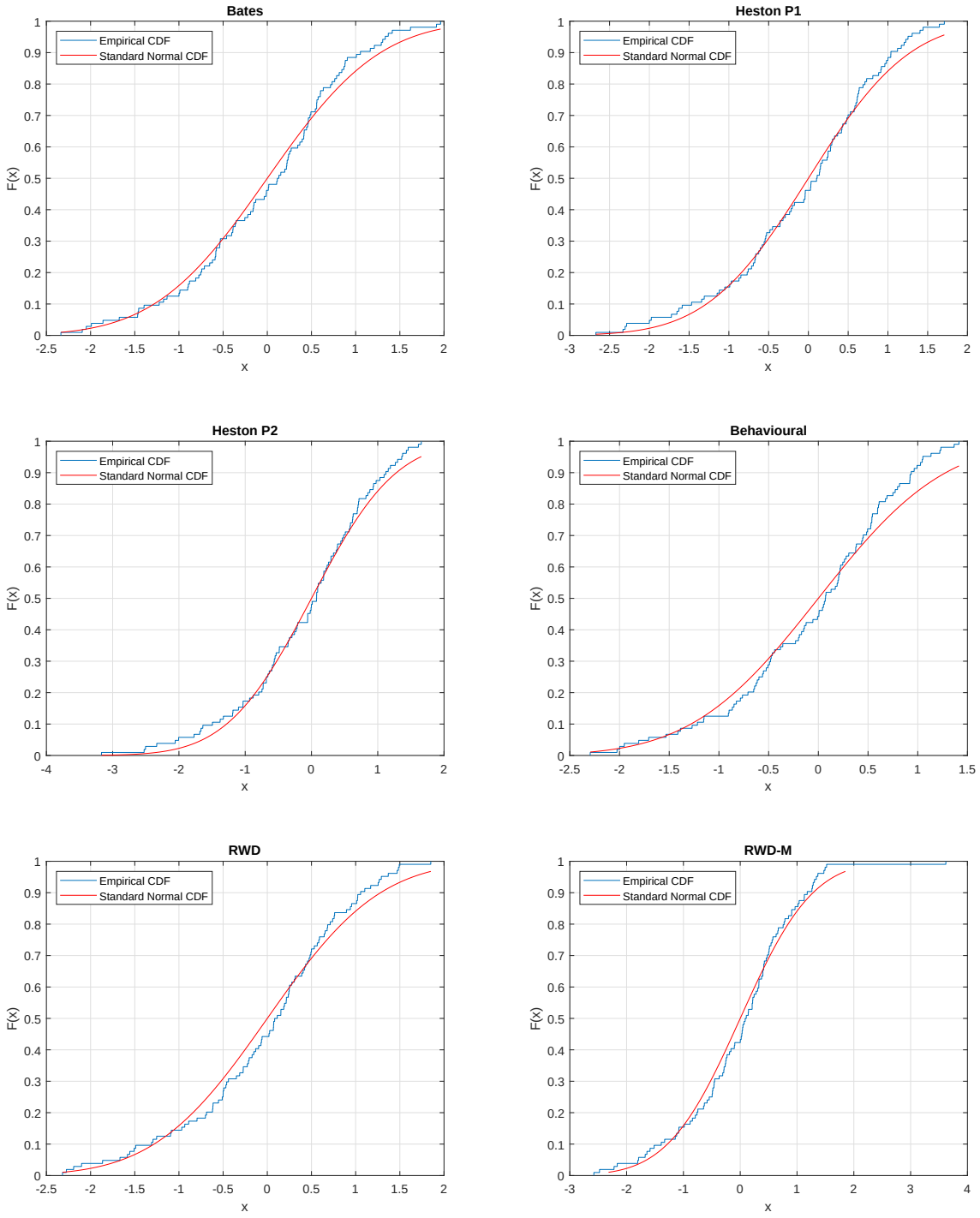


Figure 4.1: Crisis period (Jan 2008 - Dec 2009): empirical cdf and normal cdf (continued from previous page).

In addition, the Kolmogorov-Smirnov (KS) and Jacque-Bera (JB) normality tests are carried out to test for normality of z_t for each of the time periods considered in this study. The results of these tests are shown in Table 4.2. More specifically, Panel A shows the normality test results for the monthly one-month return distribution forecast and Panel B shows the normality test results for the weekly one-month return distribution forecast. The JB test assesses whether the random variable, z_t, has skewness and kurtosis matching the normal distribution, which is not assessed in Berkowitz's likelihood ratio tests. Considering the distribution forecasting methods, it is only the Historical-HW method and the RWD-M method that passed the KS and JB normality tests, at a 5% significance level, for all time periods considered in this study.

Table 4.2: Goodness-of-fit: normality tests.

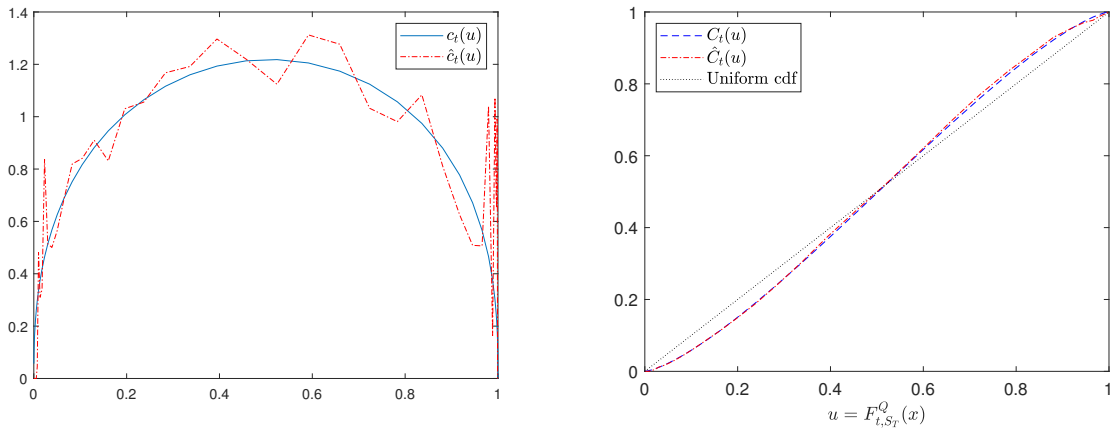
Kolmogorov-Smirnov (KS) normality test

	Panel A: Monthly	**Panel B: Weekly**			
		p-values shown			
Method	**Full-period**	**Pre-crises**	**Crises**	**Post-crises**	**Full-period**
Historical	0.685	0.365	0.004	0.136	0.323
Historical HW	0.853	0.119	0.050	0.680	0.667
RND	0.132	0.000	0.420	0.000	0.000
Heston	0.243	0.001	0.609	0.000	0.000
Bates	0.456	0.030	0.709	0.001	0.000
Heston P1	0.586	0.070	0.861	0.382	0.130
Heston P2	0.985	0.120	0.904	0.937	0.410
Behavioural	0.648	0.005	0.250	0.016	0.000
RWD	0.817	0.011	0.732	0.074	0.002
RWD-M	0.897	0.071	0.534	0.358	0.058

Jarque-Bera (JB) normality test

	Panel A: Monthly	**Panel B: Weekly**			
		p-values shown			
Method	**Full-period**	**Pre-crises**	**Crises**	**Post-crises**	**Full-period**
Historical	0.500	0.172	0.372	0.397	0.500
Historical HW	0.500	0.107	0.426	0.286	0.496
RND	0.222	0.053	0.056	0.500	0.336
Heston	0.232	0.002	0.075	0.500	0.002
Bates	0.456	0.135	0.247	0.500	0.139
Heston P1	0.500	0.420	0.049	0.153	0.500
Heston P2	0.500	0.100	0.029	0.278	0.440
Behavioural	0.001	0.287	0.042	0.001	0.001
RWD	0.226	0.069	0.062	0.470	0.081
RWD-M	0.239	0.202	0.050	0.357	0.082

In Figure 4.2, we illustrate the estimated calibration functions on 15 January 2018 using the parametric beta distribution, specified in (2.47), and the non-parametric calibration function, specified in (2.55), using historical data from 01 August 1996 to 08 January 2018. The estimated parametric and non-parametric calibration densities obtained were similar, except near the end points.[2] Furthermore, Figure 4.2b shows that the risk-neutral distribution over estimates unfavourable returns and underestimates the favourable returns.



(a) Calibration pdf

(b) Calibration cdf

Figure 4.2: The estimated parametric calibration functions with parameters $a = 1.40$ and $b = 1.38$; and non-parametric calibration functions for the Top40 index on 15 January 2018.

For the forecast methods that make use of historical data in the transformation of the estimated risk-neutral distribution into a real-world distribution (such as the parametric calibration transformation method) to be practically desirable, we would ideally want the systematic difference between the risk-neutral and real-world distributions to be stable over time and converge to some constant value as the sample of price out-turns increases (see, de Vincent-Humphreys and Noss, 2012). In particular, since the adjustment from risk-neutral to real-world would require parameter estimates based on past calibrations, it would therefore be desirable to have stable parameters over time. In Figure 4.3, we show the evolution of the beta parameters for the duration of this study. It is evident that the beta parameters typically vary over time, where a was mostly larger than b.

[2]Similar results were reported in Shackleton et al. (2010).

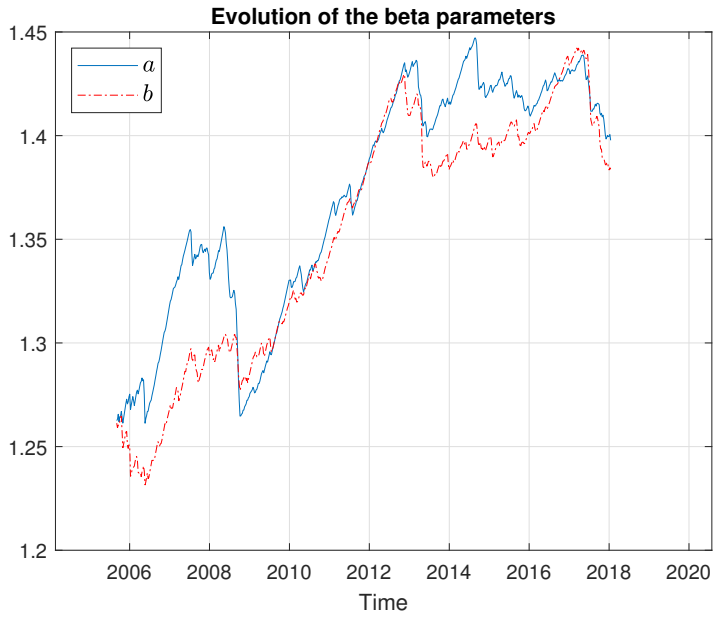


Figure 4.3: Evolution of the beta(a,b) parameters for the Top40 index.

In Figure 4.4, we show when the behavioural sentiment parameters, $(\theta_{1,t},\theta_{2,t},\theta_{3,t})$, were activated for the Top40 index over the period 05 September 2005 to 15 January 2018 with their associated adjustment values. From Figure 4.4 we observed a clustering of the investor optimism sentiment (θ_1) effect from 2006 to around 2012 which corresponds to the period of the global financial crises. In particular, this period covers most of the optimism corrections in our study period. Similar to Cristóstomo (2021), we observed positive and negative optimism adjustments during this time, signalling investor confusion. Regarding investor tail fear (θ_3), we found two clusters. The first around the year 2007 and the second over the period 2010 to 2016, which indicated the excessive fear the market had in obtaining substantial losses. There is no clear clustering of investor confidence (θ_2) over the sample period which varied between under-and-over-confident sentiment adjustments. Furthermore, we notice that the optimism and confidence sentiment adjustments alternate frequently between negative and positive sentiment adjustments. Since the sentiment adjustment factors are obtained from forward-looking information and weekly updated trading volumes, the behavioural adjustments corresponds to the weekly changes in investor sentiment and can accommodate market changes.

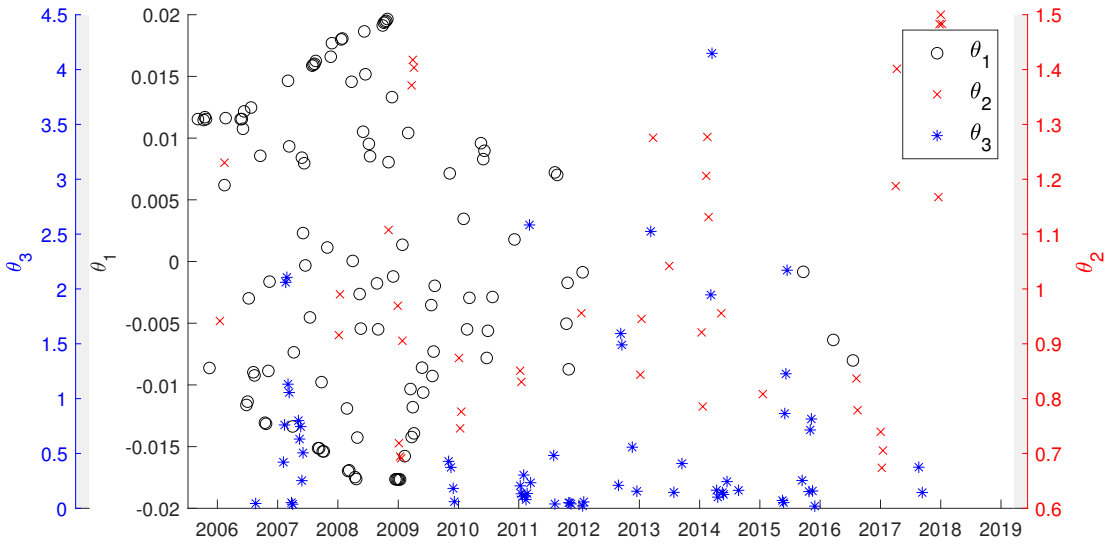


Figure 4.4: Sentiment parameters, $(\theta_{1,t}, \theta_{2,t}, \theta_{3,t})$, for the Top40 index.

The results for Berkowitz's tests for the entire distribution is shown in Table 4.3. Since we are evaluating our distribution forecasts for more than one period ahead in Panel B, the evaluation is complicated by serial correlation of the outcomes with respect to the distribution forecast. Therefore, the distribution forecast evaluation in Panel B will be more distorted by the serial correlation of the outcomes than the distribution forecasts in Panel A. Due to the serial correlation, the LR_{MS} yielded the most accurate test results, where the historical-HW obtained acceptable distribution forecasts, at a 5% level of significance, for all time periods considered in this study. Interestingly the option-implied models (namely, RND, Heston, Bates, Heston P1, Heston P2, Behavioural, RWD, and RWD-M) provided superior distribution forecasts during the global financial crisis, at a 5% level of significance, to the ordinary historical simulation method, which is a direct consequence of using forward-looking information rather than a historical database.

Table 4.3: Goodness-of-fit: Berkowitz forecast density test.

Panel A: Monthly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR^*_{ind}	LR^*	LR^*_{MS}
Monthly (Sep 2005 - Jan 2018)						
Historical	-0.09	1.13	-0.19	4.79 (0.0286)	7.26 (0.0639)	2.94 (0.2290)
Historical HW	-0.04	1.16	-0.20	5.19 (0.0227)	7.58 (0.0556)	2.56 (0.2774)
RND	0.17	0.71	-0.23	7.09 (0.0078)	14.44 (0.0024)	9.26 (0.0097)
Heston	0.06	0.62	-0.22	6.30 (0.0121)	16.93 (0.0007)	10.95 (0.0042)
Bates	0.07	0.70	-0.25	8.26 (0.0041)	13.61 (0.0035)	5.90 (0.0524)
Heston P1	0.09	0.89	-0.24	7.75 (0.0053)	8.32 (0.0398)	1.22 (0.5428)
Heston P2	0.00	0.89	-0.25	8.75 (0.0031)	8.61 (0.0350)	0.14 (0.9318)
Behavioural	0.08	0.79	-0.22	6.33 (0.0119)	8.95 (0.0300)	3.12 (0.2097)
RWD	0.02	0.72	-0.23	7.16 (0.0075)	11.28 (0.0103)	4.32 (0.1155)
RWD-M	-0.01	0.76	-0.24	7.48 (0.0062)	10.05 (0.0181)	2.76 (0.2517)

Panel B: Weekly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR^*_{ind}	LR^*	LR^*_{MS}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical	0.01	0.53	0.68	103.02 (0.0000)	74.63 (0.0000)	0.05 (0.9764)
Historical HW	0.03	0.61	0.65	88.99 (0.0000)	67.35 (0.0000)	0.37 (0.8316)
RND	0.11	0.39	0.60	75.35 (0.0000)	75.52 (0.0000)	15.31 (0.0005)
Heston	0.08	0.45	0.59	66.79 (0.0000)	63.25 (0.0000)	8.85 (0.0120)
Bates	0.08	0.61	0.59	68.57 (0.0000)	57.72 (0.0000)	1.53 (0.4644)
Heston P1	0.06	0.52	0.60	68.65 (0.0000)	58.76 (0.0000)	3.28 (0.1938)
Heston P2	0.05	0.52	0.61	69.48 (0.0000)	58.33 (0.0000)	2.73 (0.2548)
Behavioural	0.06	0.36	0.60	71.26 (0.0000)	73.55 (0.0000)	16.86 (0.0002)
RWD	0.05	0.40	0.59	66.25 (0.0000)	66.56 (0.0000)	13.13 (0.0014)
RWD-M	0.04	0.42	0.59	64.52 (0.0000)	62.50 (0.0000)	10.40 (0.0055)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical	-0.08	1.00	0.66	83.53 (0.0000)	85.95 (0.0000)	21.76 (0.0000)
Historical HW	-0.08	0.43	0.73	130.43 (0.0000)	88.20 (0.0000)	1.96 (0.3738)
RND	0.00	0.30	0.76	146.88 (0.0000)	95.57 (0.0000)	4.98 (0.0827)
Heston	-0.01	0.33	0.75	140.88 (0.0000)	91.34 (0.0000)	3.17 (0.2049)
Bates	-0.01	0.37	0.73	124.36 (0.0000)	83.49 (0.0000)	2.45 (0.2939)
Heston P1	-0.02	0.36	0.77	158.83 (0.0000)	96.61 (0.0000)	0.82 (0.6652)
Heston P2	-0.03	0.39	0.77	155.45 (0.0000)	95.05 (0.0000)	0.31 (0.8567)
Behavioural	-0.02	0.28	0.77	151.85 (0.0000)	100.10 (0.0000)	7.42 (0.0245)
RWD	-0.01	0.40	0.71	105.54 (0.0000)	74.60 (0.0000)	2.32 (0.3123)

Continued on next page

Table 4.3 – *Continued from previous page*

RWD-M	-0.01	0.57	0.64	73.09 (0.0000)	55.15 (0.0000)	0.05 (0.9735)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical	-0.02	0.42	0.68	364.77 (0.0000)	273.44 (0.0000)	10.54 (0.0051)
Historical HW	0.00	0.59	0.67	339.99 (0.0000)	249.78 (0.0000)	1.02 (0.5992)
RND	0.05	0.34	0.71	436.95 (0.0000)	332.16 (0.0000)	31.33 (0.0000)
Heston	0.01	0.24	0.70	406.65 (0.0000)	375.84 (0.0000)	91.26 (0.0000)
Bates	0.01	0.27	0.70	399.90 (0.0000)	349.15 (0.0000)	68.70 (0.0000)
Heston P1	0.02	0.47	0.71	420.13 (0.0000)	292.32 (0.0000)	0.78 (0.6775)
Heston P2	0.00	0.46	0.71	412.39 (0.0000)	288.40 (0.0000)	1.32 (0.5168)
Behavioural	0.02	0.38	0.67	340.17 (0.0000)	274.90 (0.0000)	25.40 (0.0000)
RWD	0.01	0.36	0.69	371.23 (0.0000)	293.54 (0.0000)	27.16 (0.0000)
RWD-M	0.00	0.44	0.66	320.34 (0.0000)	250.96 (0.0000)	12.78 (0.0017)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical	-0.02	0.54	0.68	552.41 (0.0000)	400.04 (0.0000)	0.43 (0.8084)
Historical HW	-0.01	0.57	0.68	553.86 (0.0000)	400.98 (0.0000)	1.11 (0.5738)
RND	0.05	0.34	0.70	650.03 (0.0000)	499.45 (0.0000)	47.43 (0.0000)
Heston	0.02	0.30	0.69	592.50 (0.0000)	506.51 (0.0000)	85.62 (0.0000)
Bates	0.02	0.36	0.68	534.42 (0.0000)	437.88 (0.0000)	49.42 (0.0000)
Heston P1	0.02	0.47	0.70	627.63 (0.0000)	441.73 (0.0000)	2.55 (0.2796)
Heston P2	0.00	0.46	0.70	622.01 (0.0000)	437.96 (0.0000)	2.39 (0.3031)
Behavioural	0.02	0.36	0.68	545.54 (0.0000)	441.79 (0.0000)	45.72 (0.0000)
RWD	0.02	0.38	0.68	542.34 (0.0000)	433.59 (0.0000)	39.15 (0.0000)
RWD-M	0.01	0.46	0.65	460.81 (0.0000)	365.88 (0.0000)	17.76 (0.0001)

Models, which do not perform well in forecasting the entire return distribution, may perform better in forecasting the tail of the return distribution. Since risk managers are often more concerned about protection against extreme losses (i.e., the lower tail of the return distribution), the Berkowitz tail test is carried out for the $\text{VaR}_{(0.95)}$ and $\text{VaR}_{(0.90)}$. In particular, we failed to use the Berkowitz tail test for the $\text{VaR}_{(0.99)}$ as we obtained no realised barrier hits for the option-implied distributions over the sample period. This may be a direct consequence that options are often used for protection against large losses that may cause the option-implied distribution to have a longer lower tail than what is normally expected by the spot market. The Berkowitz tail verification test results for the $\text{VaR}_{(0.95)}$ and $\text{VaR}_{(0.90)}$ forecasts are shown in Table 4.4.

Table 4.4: Goodness-of-fit: Berkowitz tail test.

Panel A: Monthly one-month returns: Berkowitz tail

p-values shown in parenthesis

Method	**95% VaR**			**90% VaR**		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR^*_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR^*_{tail}
Monthly (Sep 2005 - Jan 2018)						
Historical	-0.73	0.38	2.74 (0.2547)	-0.06	1.20	1.94 (0.3791)
Historical HW	-0.31	0.75	0.46 (0.7963)	0.14	1.42	1.37 (0.5047)
RND	-1.23	0.04	8.21 (0.0165)	-0.30	0.41	4.64 (0.0984)
Heston	-0.77	0.24	2.93 (0.2309)	0.13	0.84	2.59 (0.2733)
Bates	-1.02	0.13	5.26 (0.0719)	0.00	0.69	2.99 (0.2239)
Heston P1	-0.69	0.36	1.81 (0.4045)	0.15	1.03	0.73 (0.6944)
Heston P2	-0.49	0.52	0.83 (0.6615)	0.46	1.40	1.48 (0.4777)
Behavioural	-1.21	0.06	7.87 (0.0195)	0.08	0.73	4.03 (0.1330)
RWD	-0.81	0.24	3.16 (0.2064)	-0.20	0.62	1.27 (0.5298)
RWD-M	-0.65	0.38	1.78 (0.4099)	0.15	1.02	0.77 (0.6815)

Panel B: Weekly one-month returns: Berkowitz tail

p-values shown in parenthesis

Method	**95% VaR**			**90% VaR**		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR^*_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR^*_{tail}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical	-0.81	0.23	1.36 (0.5064)	0.14	1.31	0.53 (0.7687)
Historical HW	0.15	1.19	1.34 (0.5115)	0.40	1.70	1.33 (0.5132)
RND	-0.46	0.36	3.15 (0.2075)	0.45	1.00	5.97 (0.0506)
Heston	0.48	1.50	0.41 (0.8152)	1.56	2.99	5.75 (0.0564)
Bates	0.59	1.65	0.57 (0.7520)	1.60	3.06	5.97 (0.0505)
Heston P1	-0.17	0.73	0.42 (0.8118)	0.29	1.09	1.81 (0.4045)
Heston P2	0.33	1.29	0.26 (0.8768)	0.60	1.56	2.03 (0.3623)
Behavioural	-0.74	0.21	4.09 (0.1292)	0.31	0.84	6.15 (0.0432)
RWD	0.10	0.79	2.14 (0.3427)	0.96	1.68	5.79 (0.0553)
RWD-M	-0.42	0.44	1.81 (0.4045)	0.83	1.48	5.79 (0.0553)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical	-0.56	1.15	18.38 (0.0001)	-0.31	1.54	18.38 (0.0001)
Historical HW	-0.81	0.49	7.73 (0.0210)	-0.16	1.15	2.92 (0.2317)
RND	-0.38	0.58	0.42 (0.8089)	-0.15	0.75	0.29 (0.8653)
Heston	-0.22	0.83	0.19 (0.9102)	0.34	1.44	0.48 (0.7854)
Bates	-0.43	0.60	0.50 (0.7773)	0.00	0.99	0.01 (0.9956)
Heston P1	-0.23	1.00	1.51 (0.4694)	0.01	1.27	1.31 (0.5195)
Heston P2	-0.04	1.39	3.34 (0.1884)	0.25	1.78	3.18 (0.2035)
Behavioural	-0.55	0.48	0.88 (0.6427)	0.10	1.05	0.16 (0.9191)
RWD	-0.20	0.85	0.18 (0.9163)	-0.15	0.91	0.30 (0.8600)

Continued on next page

Table 4.4 – *Continued from previous page*

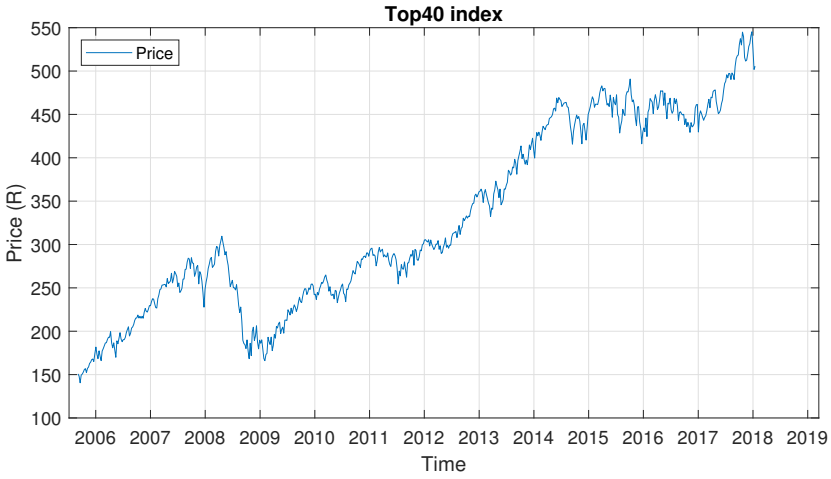
RWD-M	-0.16	0.99	0.65 (0.7216)	0.04	1.23	0.62 (0.7312)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical	-0.54	0.37	6.78 (0.0337)	0.25	0.58	4.85 (0.0883)
Historical HW	0.06	0.98	0.50 (0.7795)	-0.19	0.76	1.09 (0.5793)
RND	-0.41	0.26	27.28 (0.0000)	-0.07	0.41	37.31 (0.0000)
Heston	-0.68	0.16	28.39 (0.0000)	0.07	0.45	44.81 (0.0000)
Bates	-0.47	0.23	27.41 (0.0000)	-0.31	0.29	33.97 (0.0000)
Heston P1	-0.25	0.56	5.56 (0.0622)	-0.21	0.59	5.50 (0.0640)
Heston P2	-0.38	0.51	3.92 (0.1407)	-0.24	0.60	3.83 (0.1471)
Behavioural	-0.08	0.45	22.85 (0.0000)	-0.35	0.30	27.58 (0.0000)
RWD	-0.54	0.27	17.29 (0.0002)	-0.28	0.41	16.22 (0.0003)
RWD-M	-0.33	0.53	4.37 (0.1123)	-0.08	0.72	3.97 (0.1376)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical	-0.17	0.92	1.81 (0.4040)	0.01	1.12	1.40 (0.4954)
Historical HW	-0.21	0.85	1.52 (0.4674)	-0.07	1.00	1.04 (0.5943)
RND	-0.15	0.50	22.26 (0.0000)	0.16	0.70	33.66 (0.0000)
Heston	0.44	1.07	13.71 (0.0011)	0.94	1.55	36.09 (0.0000)
Bates	0.39	1.02	13.90 (0.0010)	0.08	0.81	11.51 (0.0032)
Heston P1	-0.12	0.76	2.21 (0.3317)	-0.03	0.84	2.66 (0.2646)
Heston P2	-0.02	0.93	0.26 (0.8761)	0.04	0.99	0.61 (0.7385)
Behavioural	-0.26	0.46	18.96 (0.0001)	-0.03	0.59	25.95 (0.0000)
RWD	-0.09	0.62	12.44 (0.0020)	-0.02	0.70	12.98 (0.0015)
RWD-M	-0.23	0.66	3.24 (0.1984)	0.09	0.93	4.22 (0.1212)

The Heston P1, Heston P2, and RWD-M model provided an acceptable fit, at a 5% level of significance, for the $\text{VaR}_{(0.95)}$ and the $\text{VaR}_{(0.90)}$ forecasts for all time periods considered in this study, where the historical-HW model provided acceptable VaR forecasts for all time periods, with exception to the $\text{VaR}_{(0.95)}$ forecast during the global financial crisis. The RND, Heston, Bates and Behavioural models all performed poorly during the post crisis and full period. Furthermore, the historical-HW, Heston P1, Heston P2, and RWD-M are the more stable preforming models in this study, outperforming the ordinary historical simulation, RND, Heston, Bates, Behavioural and RWD models for the Berkowitz tail test. In addition, the results for several backtests using the MATLAB Risk Management Toolbox (2018) are shown, for some of the extracted forecast distributions, in Appendix A (see Tables B.1 and B.2).

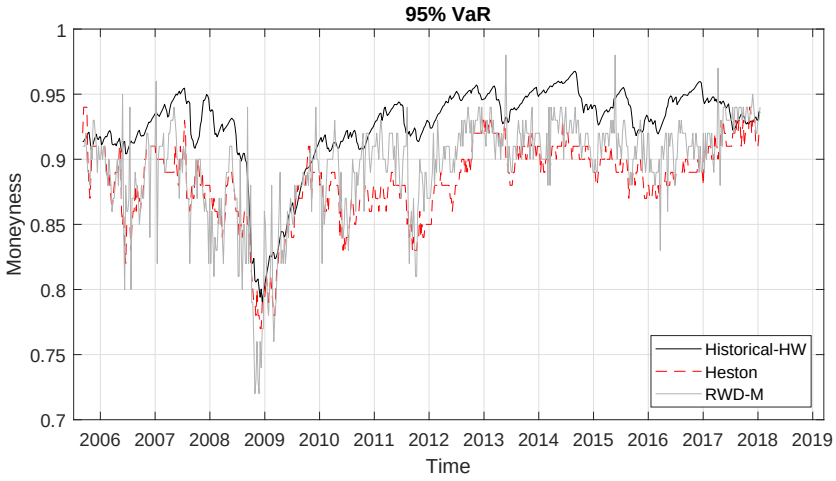
In Figure 4.5a, the weekly historical Top40 index prices is shown and in Figure 4.5b the weekly one-month $\text{VaR}_{(0.95)}$ forecasts calculated for the Historical-HW, Heston, and RWD-M models are shown. It can be seen that during the financial crisis, option-implied methods were quicker to react to market shifts than historical methods; thus making the option-implied methods more favourable during stressed (or uncertain) economic times.

For comparison of shorter term VaR estimates, we also applied the commonly used square-root scaling principle to the option-implied one-month VaR to obtain a one-week VaR forecast. This is done by multiplying the option-implied one-month VaR forecast by $1/\sqrt{4}$ (see, e.g., McNeil et al., 2005). Furthermore, we also obtained the weekly one-week VaR forecast using the two historical simulation methods. We have chosen to use the one-week VaR measures as we used weekly option prices in our dataset, making it easy to compare. The backtest results for the scaled weekly one-week $\text{VaR}_{(0.95)}$ for the option-implied models, and the $\text{VaR}_{(0.95)}$ for the historical models using a weekly return database is shown in Appendix A (see Table B.3). The results obtained are similar to that of the one-month VaR results where the return distribution forecasts obtained using option prices yielded better results than the historical simulation methods during the global financial crisis. In addition, the results for the weekly one-week $\text{VaR}_{(0.90)}$ is shown in Appendix A (see Table B.4).

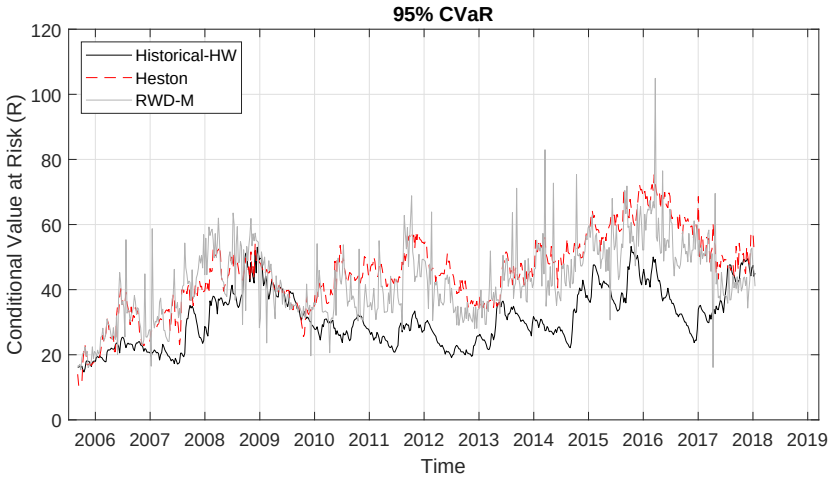
In Figure 4.5c, the weekly one-month $\text{CVaR}_{(0.95)}$ forecasts for the Historical-HW, Heston and RWD-M methods are shown. The option-implied CVaR forecasts were mostly above the historical CVaR forecasts. This indicates that the option-implied densities allocate more probability to significant losses than the historical densities. In addition, the option-implied CVaR estimates showed a significant increase over the global financial crisis period, whereas the historical methods lagged behind. Furthermore, the option-implied CVaR estimates also displayed an increase in CVaR during the period 2015-2017 when the index plateaued, indicating higher market uncertainty during the period, which was not captured by the historical simulation methods.



(a) Top40 index price



(b) $\text{VaR}_{(0.95)}$



(c) $\text{CVaR}_{(0.95)}$

Figure 4.5: Comparison of the weekly Top40 index price with the forecasted weekly one-month $\text{VaR}_{(0.95)}$, and $\text{CVaR}_{(0.95)}$.

A challenging task for risk managers is to put in place the appropriate level of capital to cover unexpected losses. Unexpected loss is a measure of operational risk and is defined to be the difference between VaR and expected loss. In short, this is the required capital that a financial institution should have to cover unexpected losses corresponding to a desired confidence level. Figure 4.6 shows the evolution of the weekly one-month forecast of unexpected losses per Top40 index share for a 95% confidence level. Similar to the CVaR forecast, the option-implied models yielded larger unexpected loss forecasts than the historical methods. This will require financial institutions to carry more capital to cover unexpected losses under the option-implied models.

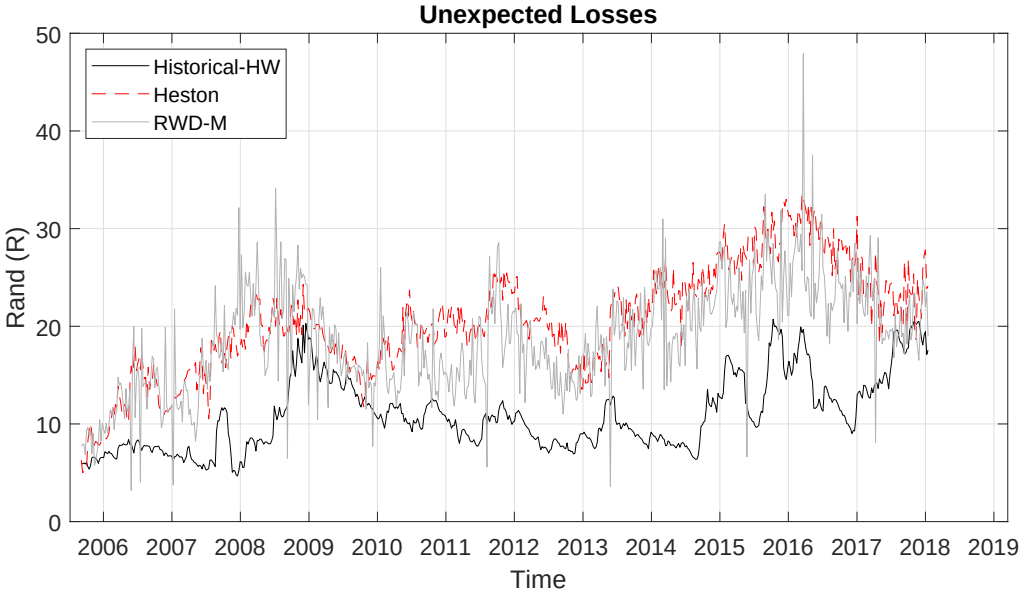


Figure 4.6: Weekly one-month forecasts of unexpected losses.

In Table 4.5 the mean Sharpe ratio[3] and volatility is shown for each time period considered in this study. We notice that the forward-looking real-world Sharpe ratio is more sensitive and showed a considerable drop during the financial crisis period, where the other methods did not.

[3]The Sharpe ratio is calculated as the ratio of excess asset return above the risk-free rate to the standard deviation of the returns. The Sharpe ratio is a measure of risk-adjusted return and indicates how well the return of an asset compensates the investor for the risk taken.

Table 4.5: Additional risk measures.

Sharpe Ratio

	Panel A: Monthly	Panel B: Weekly			
Method	**Full-period**	**Pre-crisis**	**Crisis**	**Post-crisis**	**Full-period**
Historical	0.519	0.675	0.678	0.436	0.520
Historical HW	0.529	0.628	1.009	0.389	0.534
RND	-0.147	-0.110	-0.098	-0.168	-0.146
Heston	0.020	-0.007	-0.096	0.071	0.029
Bates	-0.100	-0.160	-0.159	-0.072	-0.103
Heston P1	-0.017	0.098	-0.103	-0.026	-0.015
Heston P2	0.204	0.281	0.019	0.229	0.205
Behavioural	-0.007	0.083	-0.129	0.037	0.008
RWD	0.271	0.420	0.006	0.309	0.281
RWD-M	0.405	0.628	-0.040	0.504	0.440

Volatility

	Panel A: Monthly	Panel B: Weekly			
Method	**Full-period**	**Pre-crisis**	**Crisis**	**Post-crisis**	**Full-period**
Historical	0.183	0.191	0.209	0.175	0.183
Historical HW	0.169	0.183	0.291	0.137	0.170
RND	0.222	0.239	0.328	0.194	0.224
Heston	0.227	0.239	0.332	0.198	0.228
Bates	0.219	0.232	0.324	0.189	0.219
Heston P1	0.183	0.201	0.281	0.154	0.183
Heston P2	0.182	0.199	0.280	0.153	0.182
Behavioural	0.225	0.238	0.330	0.193	0.224
RWD	0.221	0.237	0.311	0.198	0.223
RWD-M	0.203	0.221	0.299	0.179	0.206

As we have seen, forward-looking distributions improved forecasting in certain situations (see also, e.g., Shackleton et al., 2010). Therefore, combining historical and forward-looking information may be an effective approach in obtaining a stable forecast model. The mixture of distributions was first used to estimate the risk-neutral density in Ritchey (1990), Melick and Thomas (1997) and Brigo and Mercurio (2002) as a mixture of normal densities. Furthermore, it was found that the mixture distribution was more flexible in capturing higher moments.

Historical and forward-looking information are both useful for forecasting, and effectively combining the historical information with the forward-looking information may be a desirable approach (see, e.g. Christoffersen et al., 2013).

4.4 The Mixture Distribution

Combinations of density forecast are becoming more popular in improving forecast accuracy (see, e.g., Kapetanios et al., 2015; Shackleton et al., 2010). In the previous section, we extracted forecast distributions for the Top40 index using backward-looking (historical based) methods and forward-looking (option-implied) methods. Intuitively, we are extracting market information from different sources. Both option-implied and historical forecast distributions may contain information about the underlying asset at time t for the forecast at time $t+1$. Combining these sources of information leads to evaluating the mixture of two distributions.

The *ex-ante* mixture distribution for S_{t+1} is then given by (see, e.g., Shackleton et al., 2010):

$$f^{\text{mix}}_{t,S_{t+1}}(x) = \alpha f^{\text{hist}}_{t,S_{t+1}}(x) + (1-\alpha) f^{\text{imp}}_{t,S_{t+1}}(x), \qquad 0 \leq \alpha \leq 1, \tag{4.14}$$

where α represents the mixture parameter, $f^{\text{hist}}_{t,S_{t+1}}(x)$ the historical forecast distribution function and $f^{\text{imp}}_{t,S_{t+1}}(x)$ the implied forecast distribution function.

In some instances, the option-implied distribution will outperform the historical component and in other instances the opposite may happen. Therefore, the mixture distribution may be preferred above the individual components. Finding the 'optimal' mixing parameter, α, specified in (4.14) is therefore vitally important. Shackleton et al. (2010) suggested maximising the likelihood function of (4.14) over past realisations. Therefore, the maximum log-likelihood function of α at time t is given by the number $\hat{\alpha}$ that maximises:

$$\ell(S_1, S_2, \ldots, S_t | \alpha) = \sum_{i=0}^{t-1} \log\left(\alpha f^{\text{hist}}_{i,S_{i+1}}(S_{i+1}) + (1-\alpha) f^{\text{imp}}_{i,S_{i+1}}(S_{i+1})\right). \tag{4.15}$$

Obviously, here we are optimally estimating α over the historical sample period. In Table 4.6, we give the optimal mixture parameter for each period considered in this study obtained by the maximum likelihood estimation. It is observed in Table 4.6 that more weight is given to the forward-looking distribution during economic uncertain time periods.

Table 4.6: Mixing parameter estimate.

Method	$\hat{\alpha}$
Panel A: Monthly one-month returns	
Monthly one-Month returns: Sep 2005 - Jan 2018	
Historical HW-Heston	0.4600
Historical HW-RWD-M	0.5161
Panel B: Weekly one-month returns	
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)	
Historical HW-Heston	0.2670
Historical HW-RWD-M	0.0994
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)	
Historical HW-Heston	0.0000
Historical HW-RWD-M	0.1102
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)	
Historical HW-Heston	0.6222
Historical HW-RWD-M	0.7328
Weekly one-Month returns: Sep 2005 - Jan 2018	
Historical HW-Heston	0.4398
Historical HW-RWD-M	0.5168

On the other hand, letting α optimally vary for each forecast by using forward-looking information to estimate α will be more favourable. In particular, Barone-Adesi et al. (2019) and Crisóstomo and Couso (2018) proposed such alternative ways of optimally mixing the distributions by connecting α to the percentage of deep out the money listed options, or computing α as a function of the VIX index.[4] The aim is therefore for us to choose an α structure in such a way that it can yield stable forecast results. In this study, we will follow a similar approach and relate α to the implied volatility, which is observed directly from option prices. The use of the implied volatility is a desirable choice, as it was used to identify economic uncertainty through the optimism sentiment parameter in Section 2.3.6. In particular, we will consider two simple functions related to the implied-volatility to choose α. That is,

Scenario 1: We use the notion that higher implied volatility rates imply higher economic uncertainty which will favour the use of forward-looking information above backward-looking

[4]Other approaches, such as, the Bayesian model averaging (BMA) are also useful techniques that can be used to combine density forecasts (see, e.g., Timmermann, 2006).

information. Then α is chosen as follows:

$$\alpha = \begin{cases} 0.10, & \text{for } \sigma_t^{(\mathrm{IV})} \geq 20\%, \\ 0.75, & \text{for } \sigma_t^{(\mathrm{IV})} < 20\%. \end{cases} \tag{4.16}$$

Scenario 2: Here we will use the implied volatility change $\Delta\sigma_t^{(\mathrm{IV})}$, defined in (2.31), which represents the difference between the current implied volatility and the average implied volatility over the three previous months as the proxy to determine α. In particular, we let α be the percentile of the new $\Delta\sigma_t^{(\mathrm{IV})}$ observation in the corresponding empirical distribution of $\Delta\sigma^{(\mathrm{IV})}$. That is, α can be represented as a percentile of the historical distribution of the change in implied volatility:

$$\alpha = F_{n,\Delta\sigma^{(\mathrm{IV})}}\left(\Delta\sigma_t^{(\mathrm{IV})}\right), \tag{4.17}$$

where $F_{n,\Delta\sigma^{(\mathrm{IV})}}\left(\Delta\sigma_t^{(\mathrm{IV})}\right)$ represents the empirical distribution of $\Delta\sigma^{(\mathrm{IV})}$.

The Berkowitz goodness-of-fit likelihood ratio tests for the entire mixture-distribution and tail is shown in Table 4.7 and Table 4.8, respectively for scenario 1, and in Table 4.9 and Table 4.10, receptively for scenario 2. In particular, we only considered the mixing of the risk-neutral distribution extracted using the Heston model with the historical simulation with volatility updating (Historical HW) and the mixing of the real-world distribution extracted by the recovery theorem using the multivariate model (RWD-M) with the Historical HW. We observed that in most cases the mixed distributions yielded an improvement over the forward-looking distributions. Therefore, mixing the distributions is a viable way of obtain more stable forecast distributions. Moreover, the mixing of distributions will be left for further research.

Table 4.7: Goodness-of-fit (Scenario 1): The Berkowitz test based on the mixture distribution with mixing parameter based on the implied volatility.

Panel A: Monthly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR_{ind}	LR	LR_{MS}
Monthly (Sep 2005 - Jan 2018)						
Historical HW-Heston	-0.01	0.74	-0.24	7.53 (0.0061)	11.05 (0.0115)	3.72 (0.1560)
Historical HW-RWD-M	-0.04	0.80	-0.23	6.81 (0.0091)	8.55 (0.0359)	1.99 (0.3690)

Panel B: Weekly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR_{ind}	LR	LR_{MS}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical HW-Heston	0.05	0.40	0.60	71.11 (0.0000)	67.80 (0.0000)	11.36 (0.0034)
Historical HW-RWD-M	0.03	0.44	0.60	68.36 (0.0000)	62.25 (0.0000)	7.79 (0.0204)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical HW-Heston	-0.02	0.29	0.76	143.45 (0.0000)	96.51 (0.0000)	7.33 (0.0257)
Historical HW-RWD-M	-0.02	0.41	0.72	113.88 (0.0000)	77.71 (0.0000)	1.27 (0.5302)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical HW-Heston	0.01	0.31	0.70	389.50 (0.0000)	290.15 (0.0000)	14.45 (0.0007)
Historical HW-RWD-M	-0.01	0.43	0.67	364.12 (0.0000)	271.35 (0.0000)	9.05 (0.0108)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical HW-Heston	0.00	0.38	0.69	590.70 (0.0000)	449.02 (0.0000)	29.28 (0.0000)
Historical HW-RWD-M	-0.01	0.43	0.68	545.97 (0.0000)	411.25 (0.0000)	15.41 (0.0004)

Table 4.8: Goodness-of-fit (Scenario 1): The Berkowitz tail test based on the mixture distribution with mixing parameter based on the implied volatility.

Panel A: Monthly one-month returns: Berkowitz tail

p-values shown in parenthesis

Method	**95% VaR**			**90% VaR**		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}
Monthly (Sep 2005 - Jan 2018)						
Historical HW-Heston	-0.88	0.20	3.66 (0.1601)	-0.21	0.62	1.25 (0.5351)
Historical HW-RWD-M	-0.47	0.53	0.75 (0.6869)	0.14	1.08	0.29 (0.8653)

Panel B: Weekly one-month returns: Berkowitz tail

p-values shown in parenthesis

Method	**95% VaR**			**90% VaR**		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical HW-Heston	0.34	1.17	0.83 (0.6619)	1.47	2.51	6.43 (0.0402)
Historical HW-RWD-M	0.73	1.67	1.03 (0.5966)	1.75	3.06	7.06 (0.0293)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical HW-Heston	-0.56	0.48	0.92 (0.6312)	-0.08	0.85	0.11 (0.9479)
Historical HW-RWD-M	-0.44	0.71	1.50 (0.4713)	0.02	1.20	0.59 (0.7439)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical HW-Heston	0.11	0.68	7.64 (0.0220)	-0.12	0.62	8.11 (0.0173)
Historical HW-RWD-M	-0.32	0.54	3.99 (0.1358)	-0.17	0.78	3.33 (0.1892)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical HW-Heston	-0.25	0.58	6.44 (0.0399)	0.08	0.84	9.26 (0.0097)
Historical HW-RWD-M	-0.14	0.76	1.80 (0.4064)	0.08	0.96	2.47 (0.2908)

Table 4.9: Goodness-of-fit (Scenario 2): The Berkowitz test based on the mixture distribution with mixing parameter based on the implied volatility.

Panel A: Monthly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR_{ind}	LR	LR_{MS}
Monthly (Sep 2005 - Jan 2018)						
Historical HW-Heston	-0.05	0.84	-0.23	7.29 (0.0069)	8.13 (0.0433)	1.15 (0.5636)
Historical HW-RWD-M	-0.09	0.90	-0.23	7.06 (0.0079)	7.74 (0.0517)	1.32 (0.5168)

Panel B: Weekly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR_{ind}	LR	LR_{MS}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical HW-Heston	0.03	0.46	0.64	82.03 (0.0000)	67.16 (0.0000)	5.48 (0.0647)
Historical HW-RWD-M	0.02	0.48	0.63	77.70 (0.0000)	63.16 (0.0000)	2.10 (0.3496)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical HW-Heston	-0.06	0.34	0.73	125.39 (0.0000)	87.43 (0.0000)	3.81 (0.1487)
Historical HW-RWD-M	-0.06	0.40	0.73	127.19 (0.0000)	85.49 (0.0000)	2.77 (0.2504)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical HW-Heston	0.00	0.42	0.68	352.58 (0.0000)	268.27 (0.0000)	12.33 (0.0021)
Historical HW-RWD-M	-0.01	0.47	0.67	337.75 (0.0000)	253.68 (0.0000)	5.87 (0.0532)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical HW-Heston	0.00	0.41	0.68	557.28 (0.0000)	421.89 (0.0000)	19.86 (0.0000)
Historical HW-RWD-M	-0.01	0.46	0.68	540.73 (0.0000)	402.79 (0.0000)	9.69 (0.0079)

Table 4.10: Goodness-of-fit (Scenario 2): The Berkowitz tail test based on the mixture distribution with mixing parameter based on the implied volatility.

Panel A: Monthly one-month returns: Berkowitz tail

p-values shown in parenthesis

Method	**95% VaR**			**90% VaR**		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}
Monthly (Sep 2005 - Jan 2018)						
Historical HW-Heston	-0.13	0.87	0.06 (0.9729)	0.06	1.10	0.05 (0.9766)
Historical HW-RWD-M	0.05	1.19	0.45 (0.7969)	0.30	1.53	0.93 (0.6297)

Panel B: Weekly one-month returns: Berkowitz tail

p-values shown in parenthesis

Method	**95% VaR**			**90% VaR**		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical HW-Heston	0.46	1.63	0.52 (0.7709)	1.02	2.34	3.40 (0.1830)
Historical HW-RWD-M	0.55	1.77	0.73 (0.6951)	0.78	2.03	2.35 (0.3081)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical HW-Heston	-0.36	0.68	0.35 (0.8404)	-0.17	0.88	0.35 (0.8393)
Historical HW-RWD-M	-0.59	0.66	3.87 (0.1444)	0.04	1.35	1.64 (0.4403)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical HW-Heston	0.88	1.48	10.97 (0.0041)	-0.13	0.63	6.73 (0.0346)
Historical HW-RWD-M	-0.09	0.76	2.03 (0.3627)	-0.15	0.74	1.65 (0.4374)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical HW-Heston	0.39	1.20	5.58 (0.0613)	0.01	0.85	4.04 (0.1325)
Historical HW-RWD-M	-0.12	0.85	0.28 (0.8684)	0.03	0.99	0.23 (0.8921)

The risk-premium is an important number in finance as it provides an estimate of the markets forward looking expectations (see, e.g., Lewis, 2019; Christoffersen et al., 2013). Next, we outline a simple approach in extracting an estimate of the risk-premium.

4.5 A Measure of the Risk-Premium

The expectations of the risk-neutral and real-world distributions can be used to provide an indication of the risk-premium embedded in the price of a futures contract (see, e.g., de Vincent-Humphreys and Noss, 2012). In essence, the current price of an asset is written as the discounted expected future price, under a different probability measures. In particular, the risk-neutral expectation is discounted by the risk-free rate, where the real-world expectation is discounted by a premium above the risk-free rate. That is (see, e.g., de Vincent-Humphreys and Noss, 2012),

$$S_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}(S_T) = e^{-(r+\mu)T}\mathbb{E}^{\mathbb{P}}(S_T). \tag{4.18}$$

The risk-premium can then be recovered once the expectations are found under the two measures. That is,

$$\mu = \frac{1}{T}\log\left(\frac{\mathbb{E}^{\mathbb{P}}(S_T)}{\mathbb{E}^{\mathbb{Q}}(S_T)}\right). \tag{4.19}$$

The evolution of the risk premium, extracted using the Heston P2 and RWD-M distributions for the real-world expected future price and the Heston model for the risk-neutral expected future price, is shown in Figure 4.7.[5] Furthermore, the mean and volatility of the risk-premium for the duration of this study is shown in Table 4.11.

Table 4.11: Yearly risk-premium rate.

	Mean	Volatility
$\mu_{Heston\,P2}$	3.5%	3.7%
μ_{RWD-M}	6.3%	14.0%

[5] Similar results to Figure 4.7 were achieved for the Heston P1 and Behavioural approach.

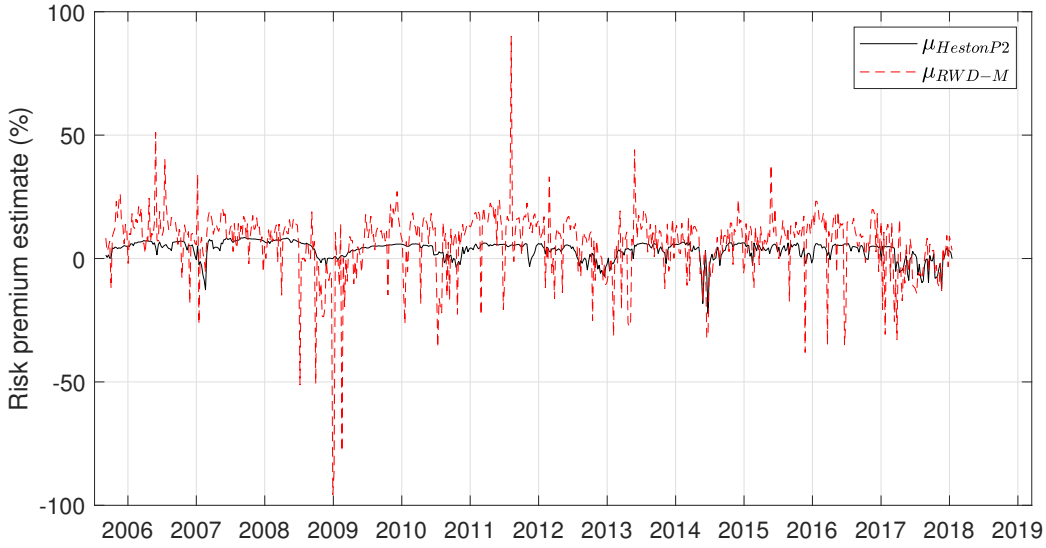


Figure 4.7: Evolution of the risk premium for the Top40 index.

We mostly observed a positive risk premium for the duration of the study. Furthermore, the estimated risk-premium extracted using the RWD-M model is more volatile than the risk-premium extracted using the Heston P2 model. This is likely a result that the RWD-M does not make use of historical data to transform the risk-neutral probabilities to real-world probabilities and is therefore more sensitive to market changes. The estimated equity risk premium is a very important number in finance. Such information can be useful in a tactical asset allocation framework for portfolio managers.

4.6 Conclusion

In this study, we implemented several methods for extracting the return distribution forecasts for the South African Top40 index with application to risk management. More specifically, two of these methods extracted the return distribution forecast using historical simulation, three methods extracted the risk-neutral return distribution forecast from option prices, three methods extracted the real-world distribution from the risk-neutral distribution by making use of transformation methods and two methods used the recovery theorem proposed by Ross (2015) to extract the real-world return distribution forecast.

These methods were backtested and their performances over multiple time-periods were compared. Using a series of likelihood ratio tests, proposed by Berkowitz (2001), we found that no model proved to be reliable in extracting the entire return distribution forecast in all tests, where only the Historical-HW, Heston P1 and Heston P2 models proved to be reliable in extracting the entire return distribution forecast when Berkowitz's test was relaxed for serial correlation. However, it is naïve to expect that one can accurately extract the entire true market return distribution using a simple statistical model. A more realistic expectation is that only a specific region of the return distribution forecast is accurately extracted. Since risk managers are often more concerned with experiencing extreme losses, we used the Berkowitz tail test and other commonly used VaR backtests found in the literature to test whether the tail of the extracted real-world return distribution forecasts provided us with a more reliable VaR forecast than the historical simulation and risk-neutral VaR forecast.

In our study using the Top40 index, we found that the option-implied methods provided information about the potential losses in the Top40 index. More specifically, the extracted distributions using option prices yielded superior VaR measures to the historical methods during the global financial crisis. Although the historical methods are well suited during normal economic periods, the real-world distribution forecasts can be an effective alternative during crisis periods. In addition, the RWD-M yielded more stable VaR forecasts over all time periods considered than the risk-neutral distributions, making the recovery theorem useful in forecasting VaR. Moreover, using the option-implied distributions will lead to overestimating the required risk capital during normal market conditions. Therefore, optimally mixing the information obtained from risk-neutral, real-world and historical methods to obtain better risk forecasts can be valuable.

Part II

Modelling using the Forecast Distribution

CHAPTER 5

QUANTITATIVE GUIDELINES FOR RETIRING (MORE SAFELY) IN SOUTH AFRICA

This chapter is adapted from Van Appel et al. (2021) and addresses the third research question.

Chapter Synopsis

In this chapter, we present guidelines for safe withdrawal rates from a living annuity (income drawdown accounts), periodically, to cover living expenses. In essence, a retiree is faced with the risk management problem of outliving their retirement fund (withdrawing too much) versus living below their means (withdrawing too little). The empirical evidence in the literature advocates for a 'safe' 4% annual withdrawal (or spending) rate. Therefore, the object of this chapter is to examine withdrawal rates for retirees in the South African economy. Furthermore, we carry out a simulation study using historical data while incorporating longevity and fund management fees. Our analysis emphasises the risks associated with different withdrawal rates and asset allocations. We then give an example of how derivative instruments can increase the success rate of a retirement portfolio.

5.1 Introduction

Retirement planning is an important topic for actuaries and financial advisors as it plays a vital role in society, especially in an era of increasing longevity. The demographer, James Vaupel, is reported to have said that "half of the children born in Sweden in 2012 will live to be 104"

(Ennart, 2012). Moreover, reports of increased longevity abound; see for example, Innovation Hub (Purdy, 2015) reports, "the first person to live to 150 has already been born." In Table 5.1, we illustrate how life expectancy (and conditional life expectancy) has increased globally over time (see, World Health Organization, 2020).

Table 5.1: Life expectancy at birth, at age 60, for global and South African citizens of all sexes.

Year	Region	Life expectancy at birth			Life expectancy at age 60		
		Male	Female	Both sexes	Male	Female	Both sexes
2000	Global	64.4	68.7	66.5	17.2	20.2	18.8
	RSA	56.1	62.0	59.0	13.4	17.8	15.7
2005	Global	66.1	70.3	68.2	17.8	20.7	19.3
	RSA	52.6	56.2	54.4	13.4	17.8	15.7
2010	Global	68.0	72.3	70.1	18.4	21.3	19.9
	RSA	55.6	60.4	58.0	13.6	18.1	16.0
2016	Global	69.8	74.2	72	19	21.9	20.5
	RSA	60.2	67.0	63.6	14.0	18.8	16.6

Increased longevity does, however, pose challenges. In a recent study, Allianz (2010) surveyed US adults aged between 44 and 75; of the people surveyed, 61% reported being more afraid of outliving their financial assets than dying! Similar findings were reported in a recent survey in the UK (see, Institute and Faculty of Actuaries, 2019), where the majority of the people surveyed believed they were not currently saving enough for retirement.

To illustrate some of the concerns faced by retirees, we need to look at life expectancy post-retirement. Richman (2017) studied the mortality of those aged 75 and above in the South African population, where he found mortality improvement rates of 0.7% and 0.1% per annum over the period 1985-2011 for males and females respectively. However, mortality rates for the population can be quite different to that of insured (or pensioner) lives. The most recent report focusing on mortality rates of insured (or pensioner) lives in a South African context can be found in CSI (2017) covering the period 2005-2010. Furthermore, Richman and Velcich (2020) studied insured (or pensioner) lives mortality improvements in South Africa, where they found that from their sample, that mortality improvements are slowing down at all ages.

It is evident from most recent reports on pensioner mortality in South Africa (CSI, 2017; Richman, 2017; Richman and Velcich, 2020) that pensioners who reach their retirement years could face 20 more years of life with substantial probability. The retiree, therefore, needs to

understand the balance between periodic inflation adjusted payments to sustain living expenses and avoid running out of capital - all given an inherently volatile investment environment (e.g., Scott et al., 2009).

In studies carried out by Butler and van Zyl (2012a,b), it was found that consumption rates after retirement do not tend to decrease, as popularly believed, and may suggest the need of an upward adjustment of retirement adequacy goals. Furthermore, they conclude that based on their analysis, retirement before the age of 67 is unlikely to be affordable for most households.

Cooley et al. (1998) note,

> "Most investors who plan for retirement eventually confront the question of how much money they should plan to withdraw from their investment portfolio. The dilemma is that if they withdraw too much, they prematurely exhaust the portfolio, but if they withdraw too little, they unnecessarily lower their standard of living."

Their conclusions, the so-called "4% safe withdrawal rate", derived for the US-market are often used as a rule-of-thumb by advisors to guide to "safe" inflation-adjusted spending by retirees (see also, Bengen, 1994).

In the South African context, retirees choose between life annuities and living annuities[1]. In particular, a life annuity is an insurance contract that provides the retiree with a stipulated income for life. Alternately, living annuities allow the retiree the freedom to invest in a wide range of investment product while drawing a monthly amount for pension - this amount is currently limited, by South African law, to between 2.5% and 17.5% per annum. A life annuity will typically leave no benefit to the pensioner's estate, whereas a living annuity could bequeath a substantial amount to their estate.

The 4% safe withdrawal rate studies by Cooley et al. (1998, 1999) were performed for retirees in the United States. In 5.2, we demonstrate differences between asset class returns in the United States and South Africa (over the period 1900 to 2015) using reference returns from Dimson et al. (2016). Real asset-class returns are similar in both countries, however, we note that South African inflation is typically significantly higher (and more volatile) than inflation in

[1]Typically referred to as income drawdown accounts in international markets.

the United States.

Table 5.2: AM/GM (arithmetic/geometric) mean returns (% p.a.).

		Asset class returns					
		US			RSA		
	Asset Class	**AM**	**GM**	**StdDev**	**AM**	**GM**	**StdDev**
Nominal	Equities	11.4%	9.4%	19.9%	14.7%	12.6%	22.9%
	Bonds	5.3%	4.9%	9.0%	7.2%	6.8%	9.6%
	Cash	3.8%	3.8%	2.9%	6.1%	6.0%	5.5%
	Inflation	3.0%	2.9%	4.8%	5.2%	4.9%	7.3%
Real	Equities	8.3%	6.4%	20.1%	9.4%	7.3%	22.1%
	Bonds	2.5%	2.0%	10.4%	2.3%	1.8%	10.5%
	Cash	1.0%	0.8%	4.6%	1.2%	1.0%	6.1%

Given the importance of retirement planning, a number of research papers can be found in the literature (see, for example, Milevsky and Huang, 2011; Butler and van Zyl, 2012b; Waring and Siegel, 2015; Maré, 2016; Rusconi, 2020; Klein and Sapra, 2020). Moreover, Maré (2016) considers safe withdrawal rates in the South African context using different asset allocations between stocks and bonds, and Rusconi (2020) considers regulatory and government policies in the South African context.

Although, there have been numerous studies on safe withdrawal rates for different asset allocations, fewer studies incorporate transactional fees (costs) and longevity into the analysis. Therefore, the aim of our research is to,

(i) extend the research done by Maré (2016) by examining withdrawal rates (read synonymously with spending rates) in living annuities for South African retirees using an extended dataset of historical asset class returns over the period 1900 to 2020 for equity, bonds, cash, and inflation. A basic requirement for any statistical analysis is that some of the statistical properties of the data under study remains stable over time, which corresponds to the stationarity hypothesis (see, Cont, 2001). In the analysis, we also considered reducing the historical dataset to more recent historical returns, where we found no difference in the results.

(ii) incorporate transactional fees (costs) and longevity into the analysis. Longevity is a key component in measuring the success of a portfolio, as a portfolio needs to outlast the indi-

vidual's life expectancy and not a predestined time of, for example, 30 years. Therefore, a typical question would be to ask whether a 5% annual spending rate remains sustainable for South African retirees. Since life expectancy is a vital part of retirement portfolio success, we incorporate the conditional probability of surviving into our simulation.

(iii) consider the impact of hedging some of the downside of the equity market on the portfolio safety.

Furthermore, we concur with Rusconi (2020) that research of this nature is highly relevant given the well-established markets in South Africa. The data and asset classes used for this study are primarily based on South African assets.

An outline of the rest of the chapter is as follows. In Section 5.2 we provide detail of our general simulation methodology, which is based on randomly sampling (with replacement) returns in a Monte Carlo simulation based on historical returns. We provide results in Section 5.3 with details on probabilities of depleting capital for various periods of investment. We provide relevant conclusions and areas for further investigation in Section 5.4.

5.2 Methodology

Cooley et al. (1998, 1999) and Bengen (1994) proposed a bootstrap simulation approach for calculating the end-of-period portfolio values from equities and bond returns for overlapping periods (known as rolling periods). Furthermore, Cooley et al. (2003) also consider a Monte Carlo simulation based on the distributional characteristics of the asset classes. In our study we calculate end-of-period portfolio values from equities, bonds, and cash based on a bootstrap simulation of the historical monthly asset returns.

5.2.1 Portfolio Make-up

The retirement portfolio process is typically structured as follows: an asset allocation is decided based on advice from the retiree's financial advisor, with the portfolio weights updated annually (typically at the start of each year). In the cash account, we place a forecasted amount to be

withdrawn from the portfolio for the year based on the particular withdrawal rate at retirement and the December inflation rate. Thus, the monthly retirement spending is withdrawn from the cash account. The remaining portfolio value is rebalanced between equity, bonds, and cash according to the specified weighting structure.

In practice, institutions often demand a fee, such as advisor fees, for managing a retirement portfolio. These costs could have a significant impact on the success of a retirement portfolio and should be taken into account in retirement planning. Often there are three costs associated in the managing of a retirement portfolio, namely fund management fees, platform fees, and advisor fees. The fund management fees are typically structured as follows:

- Cash: 0.25% p.a.
- Bonds: 0.5% p.a.
- Equity: 0.75% p.a.,

where the platform and advisor fees are 0.5% and 1%, respectively, of the total portfolio value per annum.

5.2.2 The Data

In this study, we will be making use of historical asset class returns as a guide to future returns. The data used comprises historical equity, bonds, cash, and inflation total return performances in South Africa for the period 1900 to 2020[2]. This data consisted of monthly and yearly rates (see Table 5.3). When only yearly rates were available, we approximated the monthly rates by using the 12th root function of the yearly return rate and adjusting this rate with a monthly seasonal adjustment factor based on average corresponding returns so that the monthly returns match the yearly return.

[2]Sourced from Firer and McLeod (1999), Firer and Staunton (2002), and I-Net.

Table 5.3: Data summary.

	Yearly data	Monthly data
Equity	1900-1959	1925-2020
Bonds	1900-1945	1946-2020
Cash	1900-1928	1929-2020
Inflation	1900-1938	1939-2020

The stylised facts present in the data are generally not easy to exhibit by using stochastic processes (see, e.g., Cont, 2001), hence the preferred use of historical returns in this simulation study. Although, historical returns do not necessary reflect future returns, the underlying statistical properties, to some extent, remain stable over time. Figure 5.1a through Figure 5.1c show the historical monthly returns over the period 1900 to 2020 for a portfolio consisting of equity, bonds, and cash, in South Africa. While the individual asset-class returns are summarised in Table 5.2, portfolios typically consist of a range of asset classes. Therefore, in Figure 5.1, we show the historical returns for a typical balanced portfolio asset allocation. The sample correlation matrix, based on total returns, is shown in Table 5.4.

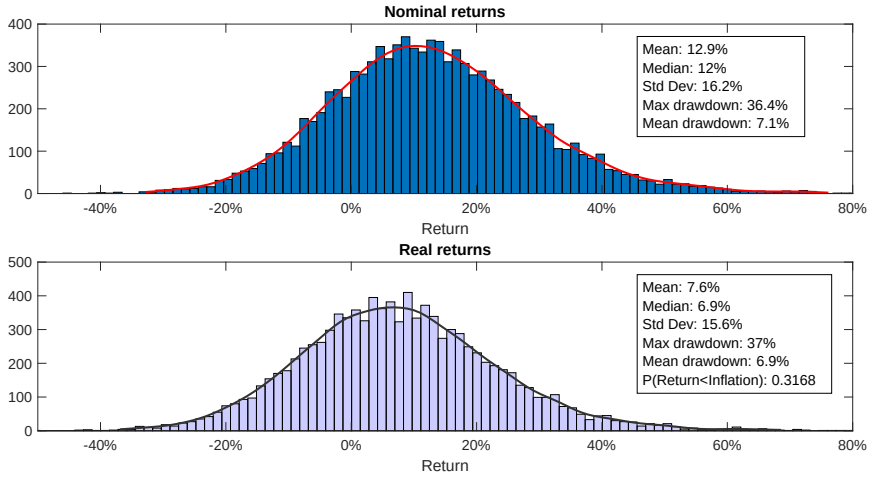
Table 5.4: Correlation matrix.

	Equity	Bonds	Cash	Inflation
Equity	1.0000			
Bonds	0.2434	1.0000		
Cash	0.0455	0.2144	1.0000	
Inflation	0.0342	0.0332	0.4391	1.0000

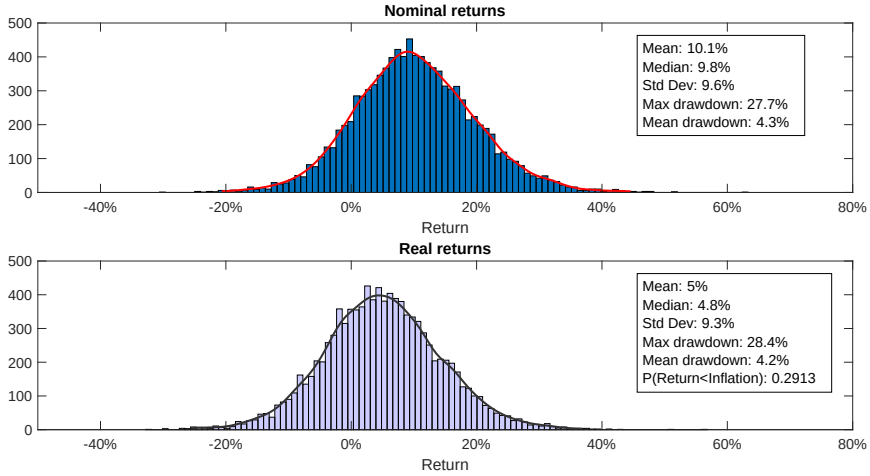
It is worthwhile to note that the South African equity market has witnessed a period of low growth over the last decade (2008-2018)[3], in-line with emerging markets, whereas the USA has had a tremendous equity bull market over the same period. Withdrawals from a retirement fund under a bear market are far more costly than withdrawals under a bull market and could drastically reduce the success of the fund. This is known as sequence risk (see, e.g., Blanchett et al., 2013).

It is important to note that our analysis is based on index returns; this decision is based on data limitations - the longest representative dataset at our avail is limited to equity, bond,

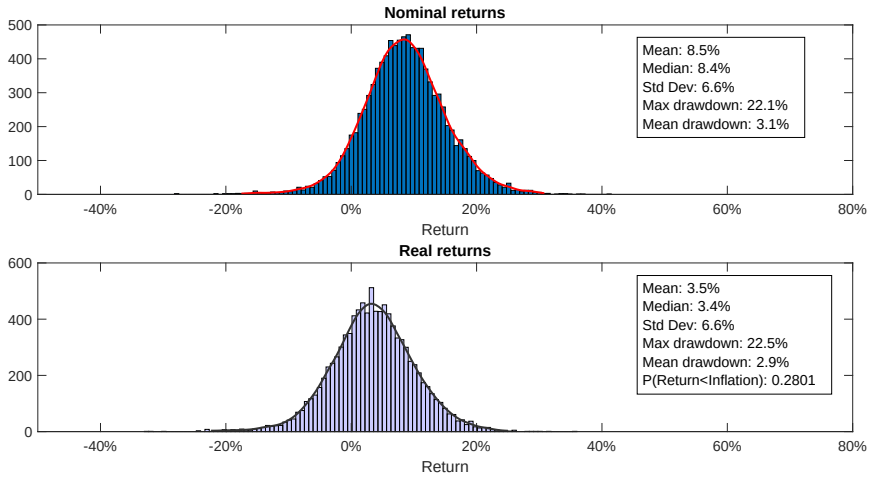
[3]See, for example, Hugo (2017).



(a) Portfolio (90% equity and 10% cash) return



(b) Portfolio (50% equity, 40% bonds and 10% cash) return



(c) Portfolio (25% equity, 65% bonds and 10% cash) return

Figure 5.1: Annual returns for each asset class from 1900 to 2020.

and cash returns. Furthermore, growth assets here are represented by equities, while we would typically add some property exposure in practice as well. In the fixed income investment space one would typically choose between a variety of bonds, which could alter risk/return ratios relative to the index-based returns contained in the data.

5.2.3 Simulation by Random Sampling

Our methodology is based on a Monte Carlo simulation using the historical return data for our asset classes (see Figure 5.1). Our data consist of monthly South African total-returns for each asset class, i, namely cash, bonds, and equity (i.e., we assume full reinvestment of interest, dividend proceeds and incidental accruals) over the period January 1900 to April 2020, i.e., $X_1^{(i)}, X_2^{(i)}, \ldots, X_{1444}^{(i)}$. From these historical returns we assume each X_j has equal probability of being selected. Furthermore, the jth sampled return is then used across all asset classes to keep the correlation structure intact. Therefore, each path constitutes a random scenario based on the monthly bootstrapped returns.

We consider an investment portfolio with a yearly rebalanced fixed asset allocation between equities, bonds, and cash, and draw an income from the cash account on a monthly basis (income is adjusted monthly for inflation, i.e., we look at real spending rates). Firstly, we consider the Monte Carlo simulation over fixed investment periods of 15, 20, 25 and 30 years, assuming no mortality, indicating the intended holding period of the portfolio for the retiree. It is important to consider success rates for a fixed investment period, as it gives a retiree a better understanding for retirement planning as the retiree's longevity is unknown. Secondly, we consider the simulation incorporating longevity using the South African pensioner mortality tables (CSI, 2017) covering mortality in the years 2005-2010, and thirdly, we incorporate portfolio costs into the simulation. The monthly portfolio value is, therefore, a function of the simulated investment returns less the inflation adjusted amounts withdrawn (inclusive of levied costs) by the retiree.

We consider a portfolio to be successful if it has capital left at the end of the specific investment period considered. We report the portfolio success rates, i.e., the percentage of portfolio values that are non-negative at the end of an investment period based on the simulations described above (see, Bengen, 1994; Cooley et al., 1998, 1999). It is important to note that

our analysis does not account for taxes; although, in principle, these can easily be taken into account.[4]

5.3 Results

Practically, safe withdrawal rates are heavily dependent on the retiree's longevity. In Table 5.5, we demonstrate portfolio success rates as a function of spending rates versus varying asset allocations over a period of 15, 20, 25, and 30 years, ignoring longevity. More specifically, the full year's withdrawal amount is placed into a cash account at the beginning of each year, and the remaining portfolio value is allocated between equity and bonds.

It is evident from the table that higher spending levels result in portfolios that will fail the retiree within the total investment period, independent of the chosen asset allocation. It is, however, interesting to note from the results that portfolios with more growth assets, such as equities, have a higher propensity for success than portfolios that are more fixed-income oriented.

To measure the extent of ruin, it is also instructive to consider the conditional expected time to failure of the portfolio - we call this the fugit[5]. More specifically, in this context, the fugit is defined to be the expected life of the portfolio given that the portfolio was unsuccessful before the intended holding period, e.g., 30 years. The fugit is expressed in monthly periods; 200 months would, for example, mean the average failed portfolio lasts 200 months out of the full intended investment period of 360 months. In Table 5.6, we demonstrate the portfolio fugit along with the standard deviation in the time to ruin as a function of spending rates versus varying asset allocations over a period of 30 years.

A key conclusion from Table 5.6 pertains to the spending rate - to sustain higher spending rates, a retiree needs to be willing to allocate more to risk-bearing assets. This can be seen by the shaded values in Table 5.6 which represents the largest expected time to ruin for each withdrawal rate. However, this strategy comes with an increase in the variation in the time to

[4]Moreover, South African taxation rules on retirement portfolios can be found in Butler et al. (2013), Van Zyl and Van Zyl (2016) and the Income Tax Act (South African Government, 1962 as amended)

[5]The term fugit was first introduced by Garman (1989) and was used to represent the optimal date to exercise an American option.

ruin. This needs to be clearly understood and forms a key consideration in the financial advisor discussion process. Asset allocation appears to be of lesser concern when spending is low.

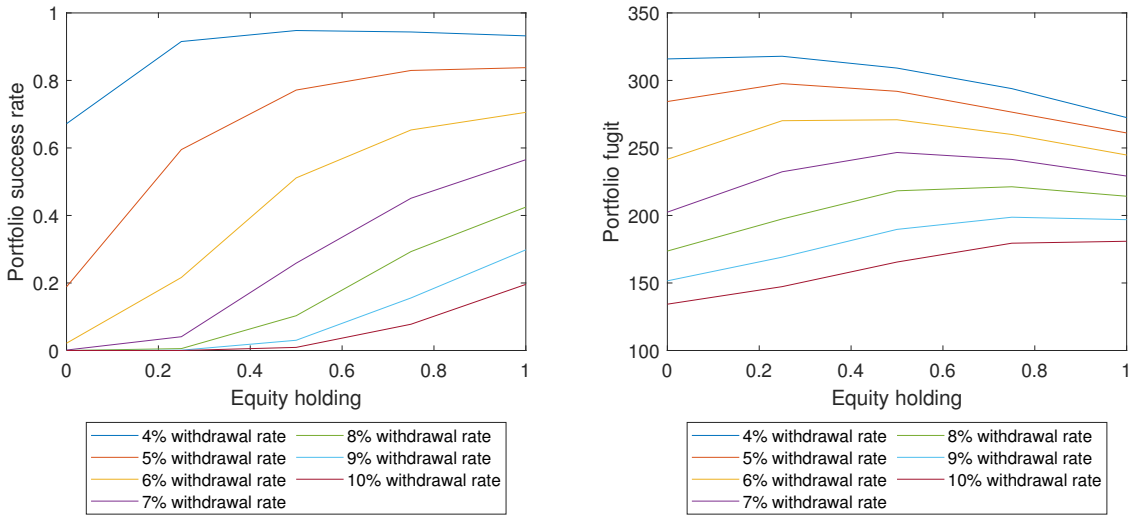
Table 5.5: Portfolio success rate given withdrawal rate vs. asset allocation (equity, bonds & cash) over 30-year period.

Payout Period	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
100% Equity							
15 Years	100%	99%	96%	88%	80%	67%	54%
20 Years	98%	94%	87%	75%	62%	47%	33%
25 Years	96%	89%	78%	64%	49%	36%	24%
30 Years	94%	84%	70%	56%	42%	29%	19%
75% Equity / 25% Bonds							
15 Years	100%	99%	97%	91%	79%	63%	43%
20 Years	99%	96%	86%	71%	53%	36%	21%
25 Years	97%	89%	75%	56%	38%	22%	12%
30 Years	94%	83%	66%	46%	28%	16%	8%
50% Equity / 50% Bonds							
15 Years	100%	100%	98%	92%	75%	51%	29%
20 Years	100%	97%	85%	64%	38%	19%	7%
25 Years	98%	88%	66%	40%	19%	7%	2%
30 Years	95%	78%	51%	26%	10%	4%	1%
25% Equity / 75% Bonds							
15 Years	100%	100%	99%	90%	63%	30%	9%
20 Years	100%	97%	77%	42%	13%	3%	0%
25 Years	98%	81%	43%	13%	2%	0%	0%
30 Years	91%	59%	22%	4%	1%	0%	0%
100% Bonds							
15 Years	100%	100%	97%	74%	34%	8%	1%
20 Years	99%	89%	48%	12%	2%	0%	0%
25 Years	90%	48%	10%	1%	0%	0%	0%
30 Years	66%	20%	2%	0%	0%	0%	0%

Table 5.6: Fugit given withdrawal rate and asset allocation (equity, bonds & cash) over 30-year period (shaded values represent the largest expected time to ruin for each withdrawal rate).

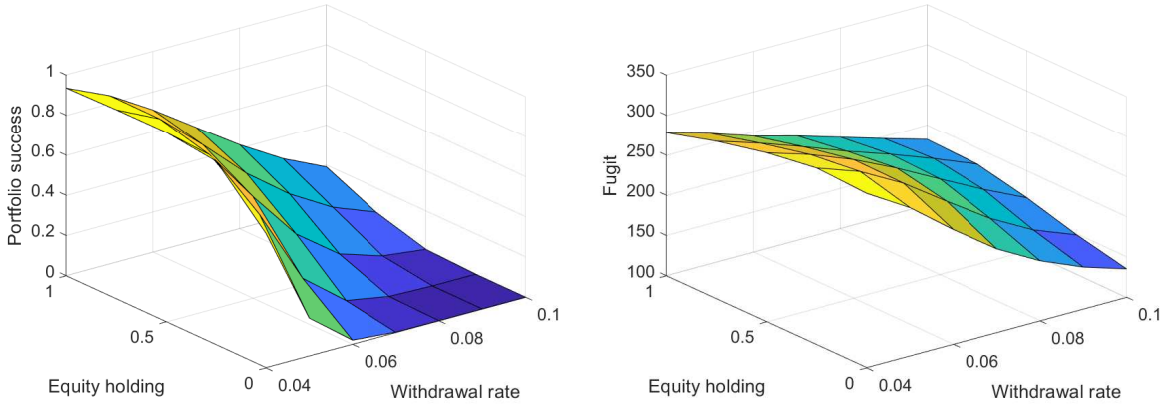
	Withdrawal rate						
Asset allocation	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
100% Stocks							
Mean	272.8	262.4	247.1	230.4	212.7	198.1	181.2
Standard deviation	53.6	56.9	61.1	63.6	64.6	64.5	63.2
75% Stocks/25% Bonds							
Mean	291.7	275.0	259.2	241.8	219.7	200.1	178.8
Standard deviation	45.7	50.8	55.7	59.1	59.9	59.7	56.2
50% Stocks/50% Bonds							
Mean	306.4	293.3	272.0	245.5	218.4	191.2	165.8
Standard deviation	37.1	44.9	50.1	53.4	53.2	50.0	42.4
25% Stocks/75% Bonds							
Mean	318.7	299.5	269.4	232.7	197.2	168.8	147.7
Standard deviation	30.8	39.2	45.2	44.9	38.0	29.0	23.4
100% Bonds							
Mean	315.7	283.9	241.9	203.0	173.3	151.5	134.2
Standard deviation	30.3	39.4	39.6	31.9	25.2	20.0	16.2

In Figure 5.2 and Figure 5.3, we show the portfolio success rates across different withdrawal rates and equity holdings, thus confirming that portfolios with higher success and longevity have larger equity allocation. In Figure 5.4, we show some descriptive measures calculated from the 10 000 Monte Carlo simulations with a 75% equity ratio and a 5% initial portfolio withdrawal. Although, we expect, the portfolio to have a positive balance at $T = 30$ years, it is also evident from Figure 5.4a that there is a 5% chance that the portfolio will run out of money within 245 months (i.e., approximately 20 years). This is obviously not a desirable outcome for retirees, even with low probability. In Figure 5.4b, we show the conditional probability of portfolio success at time, t, given success at time, $t-1$, where $t \in [0,360]$ months. We note that the conditional probability of success drastically increases from around 200 months, indicating the importance of effective early portfolio management.



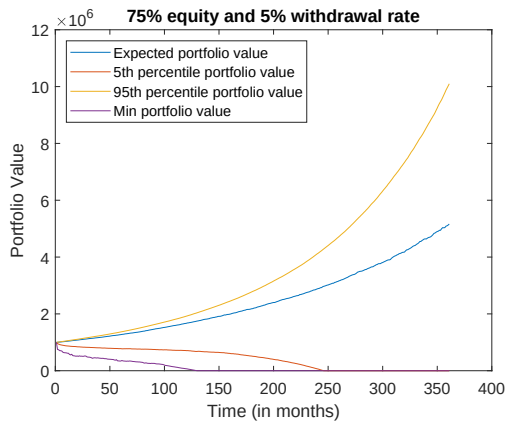
(a) Portfolio success (b) Portfolio fugit

Figure 5.2: Portfolio success and fugit over a 30 year period.

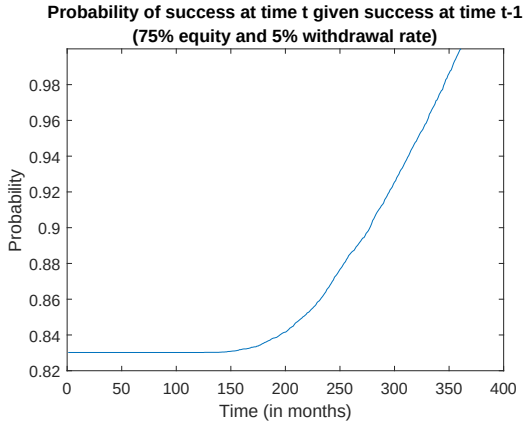


(a) Portfolio success (b) Portfolio fugit

Figure 5.3: Portfolio success and fugit over a 30 year period.



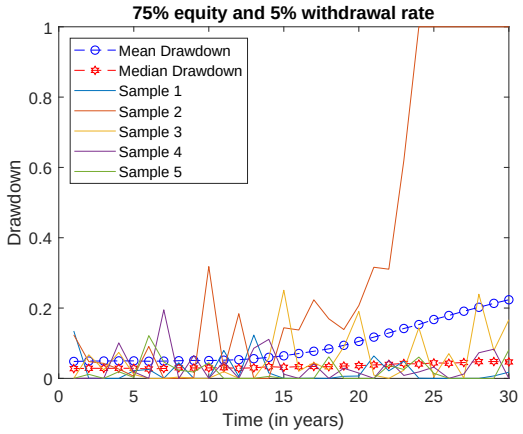
(a) Portfolio value over time



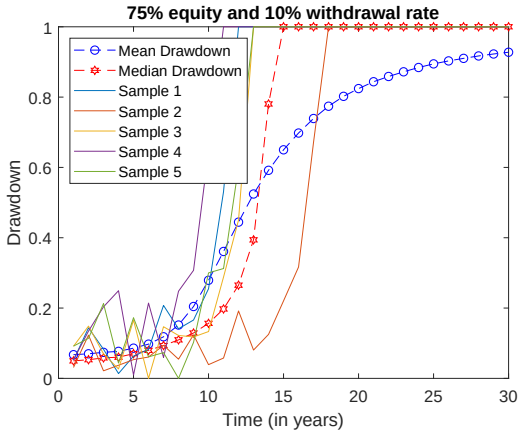
(b) Portfolio success probability over time

Figure 5.4: Portfolio descriptive measures over time.

Figure 5.5 shows the mean and median drawdown per year (measured as a proportion of the portfolio value) for the 10 000 simulated scenarios. Due to the inflation adjusted withdrawals, the drawdown increases exponentially over time. Furthermore, we also show the drawdown for the first five simulations.



(a) 75% equity with a 5% withdrawal rate



(b) 75% equity with a 10% withdrawal rate

Figure 5.5: Drawdown as a portion of the portfolio capital (the figures also include 5 sample paths of the 10 000 simulated paths).

Furthermore, the relationship between the mean yearly drawdown (as a proportion of the portfolio value), and fugit is shown in Figure 5.6. The inverse relationship between the mean yearly drawdown and fugit is clearly evident in Figure 5.6, where an average yearly drawdown in excess of 10% is unlikely to last for a 30-year period.

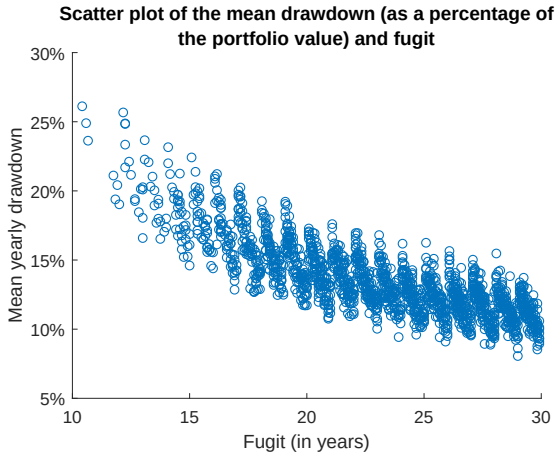


Figure 5.6: Relationship between the mean yearly drawdown (as a portion of the portfolio value) and fugit.

In Table 5.7, we incorporate longevity into our simulation. Here we are simulating the portfolio given the conditional probability of surviving[6] between x and $x+t$. Factoring longevity into the simulation improved the expected success rates of the portfolio. Furthermore, in Table 5.8, we factor fund management costs into the simulation. In our analysis, costs reduced the portfolio success by up to 4%. This is obviously not a desirable outcome on an already strained problem.

Table 5.7: Success rates with longevity.

	Withdrawal rate as percentage of initial investment value						
Asset allocation	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
100%Equity/0%Bonds	99%	96%	92%	87%	79%	72%	63%
75%Equity/25%Bonds	99%	97%	92%	85%	75%	66%	58%
50%Equity/50%Bonds	99%	97%	90%	80%	70%	60%	51%
25%Equity/75%Bonds	99%	94%	84%	72%	61%	52%	45%
0%Equity/100%Bonds	97%	87%	74%	63%	53%	46%	40%

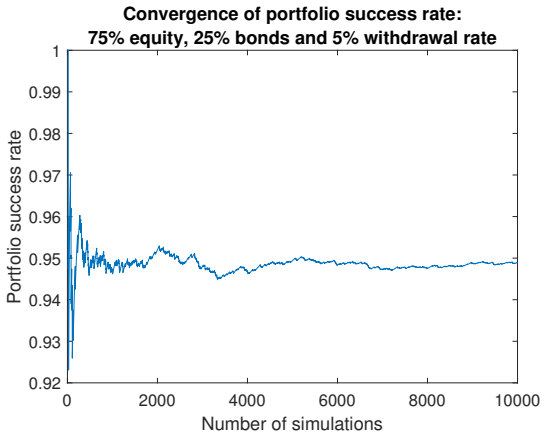
[6]The conditional survival probabilities for the simulation were obtained from CSI (2017)

Table 5.8: Success rates with fund costs and longevity.

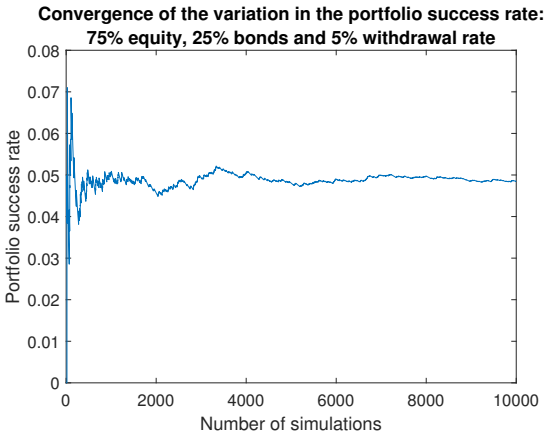
	Withdrawal rate as percentage of initial investment value						
Asset allocation	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
100%Equity/0%Bonds	98%	95%	89%	82%	75%	67%	58%
75%Equity/25%Bonds	99%	95%	89%	80%	71%	61%	54%
50%Equity/50%Bonds	99%	94%	86%	76%	64%	55%	47%
25%Equity/75%Bonds	98%	91%	79%	66%	56%	48%	42%
0%Equity/100%Bonds	94%	82%	69%	59%	50%	44%	39%

The convergence of the sample mean and variance of portfolio success rates is shown in Figure 5.7. Qualitatively, we see that the convergence graphs reach a flat region, indicating convergence of the sample moments.

Retirees are often advised to carry portfolios with lower risk at retirement. That is, they are advised to hold portfolios that have a larger asset allocation in fixed income and cash-based securities. In Table 5.9, we show that this could drastically reduce the success rates of a portfolio.



(a) Convergence of the mean portfolio success rates



(b) Convergence of the variation in the portfolio success rates

Figure 5.7: Convergence of the sample mean and variance of portfolio success rates as functions of the number of Monte Carlo simulations.

Table 5.9: Portfolio with larger cash allocation.

	Withdrawal rate as percentage of initial investment value						
Asset allocation	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
85%Equity/0%Bonds/15%Cash	55%	48%	42%	37%	33%	30%	27%
75%Equity/10%Bonds/15%Cash	53%	46%	41%	36%	32%	29%	27%
50%Equity/35%Bonds/15%Cash	50%	44%	38%	35%	31%	28%	26%
25%Equity/60%Bonds/15%Cash	47%	41%	37%	33%	30%	27%	25%
0%Equity/85%Bonds/15%Cash	44%	39%	35%	31%	28%	26%	24%

The results in Table 5.9 indicate that we need to be very conscious of our spending habits. We also need to ensure that we hold sufficient growth assets; although, these could introduce more uncertainty to our portfolios. We can, however, mitigate some of the downside risks associated with growth assets. In Table 5.10, we detail results where we protect equity holdings against downside moves. To pay for the protection we sacrifice some upside returns, in this case 50% of the return above 4%. This is done as follows:

Protection: floor = -0.03, cap = 0.04, participation = 0.5, cost = 0.001, X = one-month Equity return. Then the payoff is given as:

$$\text{payoff} = X + \max(0, \text{floor} - X) - \text{participation} \times \max(0, X - \text{cap}) - \text{cost}. \tag{5.1}$$

Table 5.10: Protecting equity returns for 5% and 10% spending rates.

		5%		10%	
Asset Allocation	**Protected equity**	**Success**	**Fugit**	**Success**	**Fugit**
75% Equities / 25% Bonds	No	83%	279.8	8%	179.7
75% Equities / 25% Bonds	Yes	95%	312.6	3%	187.0
50% Equities / 50% Bonds	No	78%	293.6	1%	166.5
50% Equities / 50% Bonds	Yes	88%	315.5	0%	165.5
25% Equities / 75% Bonds	No	59%	300.7	0%	148.7
25% Equities / 75% Bonds	Yes	61%	310.8	0%	147.8

The results in Table 5.10 are encouraging, with improvements in both the fugit and success rates in lower withdrawal rates. Moreover, these results further illustrate the main challenge encountered by retirees is that large withdrawals are not sustainable for the success of a retirement portfolio. The derivative protection strategy illustrated above can easily be implemented

by buying put options and selling call options on the index on a rolling one-month basis. Our analysis here is based on long-term average costs.[7]

5.4 Conclusion

We have presented a simulation-based approach to analyse withdrawal rates for retirement-based portfolios in a South African setting. In our approach, we examine the distribution of terminal account balances. Furthermore, we also calculate the probability that a retirees funds will not be depleted. This corresponds to the notion of a safe withdrawal rate. Our results show clearly that portfolio success rate is a rapidly decaying function of retirement spending (measured on an annual inflation adjusted percentage basis). The notion of the portfolio fugit provides insight into the expected failure time (or extent of ruin) of a portfolio and provides insight into establishing the effects of path-dependence on portfolio success.

We observe, for low withdrawal rates, that asset allocation does not have a large influence on the success of the portfolio. When we consider larger withdrawal rates, however, a higher percentage of growth assets (such as equity) is needed, and even then the portfolio is not necessarily sustainable.

We also detail that there are options to increase investment results, which entail growth asset protection by use of derivative instruments; in this chapter, we provided a proof of concept towards this. However, in Chapter 6, we will further extend the use of derivates for the protection against large losses by using observed market prices in the analysis for a more accurate assessment.

Scott et al. (2009) and Waring and Siegel (2015) criticised the framework of withdrawing a fixed rate from a naturally volatile portfolio. In particular, there is no guarantee that historical returns would be the same in the future. Therefore, the results obtained in this study indicate that moderation and caution should be applied.

Cooley et al. (2003) note that "a portfolio is only successful if it lasts as long as required by the retiree" - this motivates our use of incorporating longevity into the simulation of the

[7] Small variations in the parameters used in (5.1), resulted in small changes in the results reported in Table 5.10.

portfolio success rate. It should be noted that other measures could be used and investigated as well.

In conclusion our results showed that withdrawal rates depend critically on the investor's portfolio makeup and life expectancy. Withdrawal rates in excess of 5% is not sustainable over a 30 year period, regardless of the portfolio makeup. Furthermore, our simulation showed that the portfolio fugit is often shorter than the retiree's life expectancy for larger withdrawal rates. This indicates that retirees need to buy some life annuities in addition to living annuities. The 'optimal' incorporation of life annuities into one's retirement portfolio could safeguard against a complete loss of income – this reflects an area of future research.

Further research could be carried out in recalculating safe withdrawal rates each year based on the portfolio value, economic climate, and expected lifetime in order to decrease the likelihood of running out of money before end of life. To achieve this, forward looking return information would be vitally important. Furthermore, longevity risk remains an important and challenging factor in determining safe retirement withdrawal rates. Therefore, an alternative method to a hedging strategy is to mitigate the downside financial market risk (and achieve a given level of portfolio return) by making use of target volatility strategies (see, e.g., Olivieri et al., 2022). In addition, there is a wealth of stochastic mortality models, either in discrete-time (see, e.g., Lee and Carter, 1992) or continuous-time (see, e.g., Blackburn and Sherris, 2013), which may incorporate mortality uncertainties and shocks for the further development of risk management strategies for retirement income portfolios.

CHAPTER 6

THE RECOVERY THEOREM IN SAFE RETIREMENT WITHDRAWAL RATES

This chapter is adapted from Van Appel and Maré (2022) and addresses the third research question.

Chapter Synopsis

The focus of this chapter is on using forward-looking moments of the risk-neutral and real-world asset distributions in determining safe withdrawal rates for South African retirees. The use of forward-looking information, typically derived from traded derivative securities (rather than historical data), is essential in optimising safe withdrawal rates for retirees. In particular, we extracted the forward-looking risk-neutral and real-world distributions from option prices on the South African FTSE/JSE Top 40 (Top40) index, where the moments of the distributions were used as a signal in a simple tactical asset allocation framework. That is, when we expect the growth asset to decrease in value, we hold cash (or short the asset) and, alternatively, when we expect the growth asset to increase in value, we hold the growth asset for the period. Using this approach, we show that it is possible to sustain withdrawal rates of up to 7% compared to the commonly quoted 4% safe withdrawal rate obtained by historical simulations.

6.1 Introduction

Studies on safe retirement spending rates typically draw information from historical data (see, e.g., Cooley et al., 1998, 1999; Bengen, 1994; Maré, 2016; Van Appel et al., 2021). A prime

example of such a study is the commonly quoted "4% safe withdrawal rate" published in Cooley et al. (1998), where the authors used historical data over the period 1926 to 1995 to assess safe spending rates for retirees. The assumptions made in these studies are that the statistical properties of historical returns remain stable over time. However, it is known that historical (backward-looking) returns do not necessary predict future returns. Furthermore, in these studies, the authors assume a fixed asset allocation and a constant spending rate. Both assumptions are heavily criticised in the literature (see, e.g., Van Appel et al., 2021; Scott et al., 2009; Waring and Siegel, 2015). It is well reported that people are living longer and face a further 20 to 30 years of life with substantial probability after retirement (see, e.g., World Health Organization, 2020; Purdy, 2015; Ennart, 2012). Van Appel et al. (2021) demonstrated in an empirical study, using historical data, that for higher spending rates, a higher allocation in growth assets is needed. Even then, the portfolio is unlikely to be successful over a 30-year period (representing a typical post-retirement investment cycle). Ideally, a retiree would like to draw as much as possible, with a low probability of depleting the fund before their duration of life, or 30 years.

An important part of modelling is to generate scenarios for financial practitioners. In particular, it is important that these scenarios are as close as possible to the true representation of what could happen. In extension of the work presented by Cooley et al. (1998, 1999), Bengen (1994), and Maré (2016), the focus of this study is to improve the modelling of safe retirement spending rates by using forward-looking information rather than historical information. It is well-known that many large financial institutions regularly estimate forward-looking distributions from option prices in order to gain insights into the weights investors place on different future asset prices (see, e.g., Shimko, 1993; de Vincent-Humphreys and Noss, 2012). Therefore, modelling retirement withdrawal rates using forward-looking information should provide a more realistic assessment of safe withdrawal rates. The main advantage of using forward-looking information is that it allows for the implementation of a tactical asset allocation framework instead of a fixed asset allocation structure that is used in the literature. This allows the portfolio to potentially achieve higher asset returns, which increases the success rates of retirement portfolios. In particular, we would like to be invested in the risky (or growth) asset, when the return is expected to be favourable to the portfolio. To assess when returns will be favourable, forward-looking information should be used rather than historical data. We thereby will show that models based on historical data do not provide an optimal representation of modelling retirement portfolio success rates.

This chapter is organised as follows. Firstly, we start by introducing methods for extracting the one-month forward-looking risk-neutral and real-world return distributions from the prices of traded derivative securities. Secondly, we use the extracted moments of the forward-looking distributions in a simple tactical asset allocation framework, where we show that these forward-looking distributions are useful in forecasting the movements in the underlying asset's return. A test on whether a forecast model is of value to financial practitioners depends on its usability. Therefore, we thirdly apply the tactical asset allocation framework to determine safe retirement withdrawal rates. Lastly, we carry out a robustness test to show that the forward-looking moments of the real-world return distribution increases the success of a retirement portfolio compared to a fixed asset allocation.

6.2 Methodology

Typically, safe retirement withdrawal rates are analysed by using models based on historical data and Monte Carlo simulation (see, e.g., Cooley et al., 1998, 1999; Bengen, 1994; Maré, 2016; Van Appel et al., 2021; Scott et al., 2009; Scott, 1996; Abuizam, 2009). In particular, Scott (1996) studied the impact that a portfolio's rate of return has on safe withdrawal rates, where she found that increasing the rate of return, by increasing the equity allocation, in the retirement portfolio drastically increases safe withdrawal rates. However, this comes with higher variability in returns (or risk) (see, e.g., Van Appel et al., 2021). Therefore, the aim of this section is to use forward-looking information in a tactical asset allocation framework to maximise portfolio returns and reduce the variability in returns.

6.2.1 Forward-Looking Return Distributions

To test the usability and practicality of the real-world distribution obtained by the recovery theorem in determining safe withdrawal rates, we use the forward-looking forecasted moments in a simple tactical asset allocation framework to obtain higher returns than a model based purely on historical data with a fixed asset allocation next. In particular, we used the risk-neutral density function extracted using the model-free approach (RND) and the real-world distribution extracted using the recovery theorem with the multivariate regularisation approach (RWD-M)

as the forecast of the future asset return distributions. Furthermore, we also consider a hedging strategy by buying and selling put and call options, respectively.

6.2.2 Tactical Asset Allocation

In this section, we use the extracted forward-looking risk-neutral and real-world return distributions to forecast movements in the underlying asset returns. In particular, we extracted the forward-looking risk-neutral and real-world distributions, at the start of each month, from market-observed option prices quoted on the FTSE/JSE Top 40 index (Top40) over the period August 1996 to January 2018 (sourced from the South African FTSE/JSE), giving a total of 259 forecast months (or 259 one-month forecast distributions). The Top40 index was used as it is a key market factor in South Africa and, along with the exchange-traded derivatives on this asset, is one of the most liquid in the South African market. Furthermore, the duration of the sample (i.e., 21.5 years) would typically embrace at least three South African business cycles (see, Thomson and Van Vuuren, 2016). Thereafter, we will use the extracted risk-neutral and real-world moments in a simple tactical asset allocation framework to obtain higher returns for the portfolio.

Investors are normally in search of higher returns, skewness and kurtosis, and lower volatility. Therefore, as outlined in Audrino et al. (2014) and Flint and Maré (2017), we carried out a simple tactical asset allocation, where we hold the Top40 for the full month when the forecasted moments (mean, skewness, and kurtosis) are higher than the previous month's forecast, or when the forecasted volatility is lower than the previous month's forecast. Since trading occurs once a month, there will be a limited number of trades, resulting in negligible transactional costs (transactional costs are normally around two basis points). In particular, we found that trading costs decreased yearly returns on average by no more than 0.2%. For completeness, all results based on the tactical asset allocation framework reported in this paper will include transaction costs.

Lastly, we also consider a fixed asset allocation framework by incorporating a hedging strategy, where instead of selling the Top40 in the tactical asset allocation framework, we protect the portfolio against large losses by buying put options. This strategy involves purchasing put options with 6% out the money (OTM) strike with a 30% participation rate. Since put

options are expensive, we offset the cost by selling call options with 1% OTM strike with a 30% participation rate. Since we are using observed market-quoted prices, these parameter values were chosen such that it yielded stable and desirable results. That is, our index portfolio value will evolve as follows:

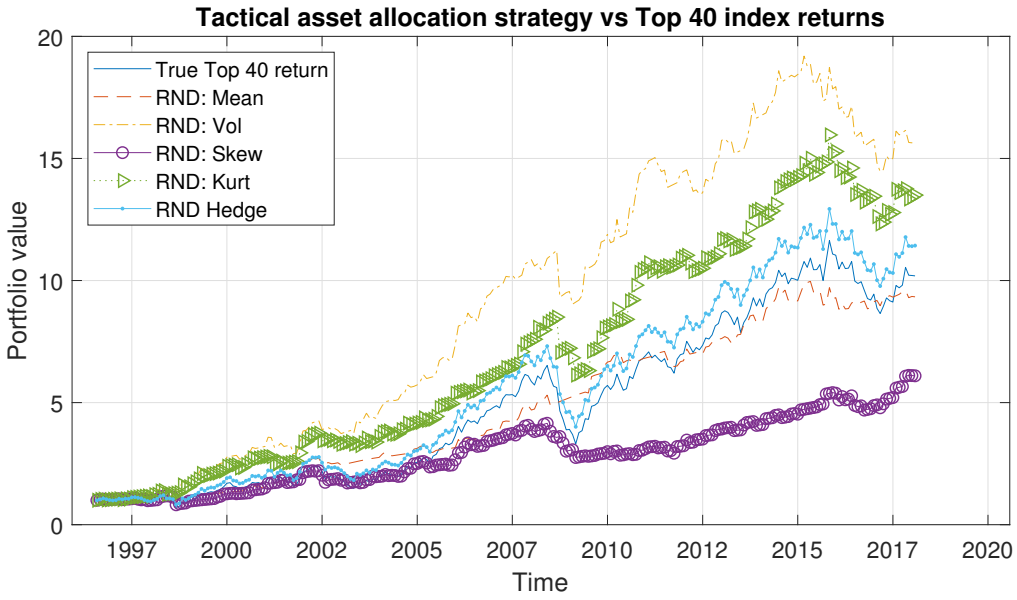
$$\begin{aligned}\pi(t) &= p^{(put)} \times \pi(t-1) \times \max\left(M^{(put)} - R_t, 0\right) \\ &\quad - p^{(call)} \times \pi(t-1) \times \max\left(R_t - M^{(call)}, 0\right) \\ &\quad + \pi(t-1)[1 - \mathrm{P}(t, M^{(put)}) \times p^{(put)} + \mathrm{C}(t, M^{(call)}) \times p^{(call)}] \times R_t, \end{aligned} \tag{6.1}$$

where $\pi(t)$ represents the portfolio value at time t, $p^{(\cdot)}$ represents the participation rates, $M^{(\cdot)}$ the moneyness rate, R_t the asset return, and $\mathrm{P}(t, M^{(put)})$ the market-quoted put option price.

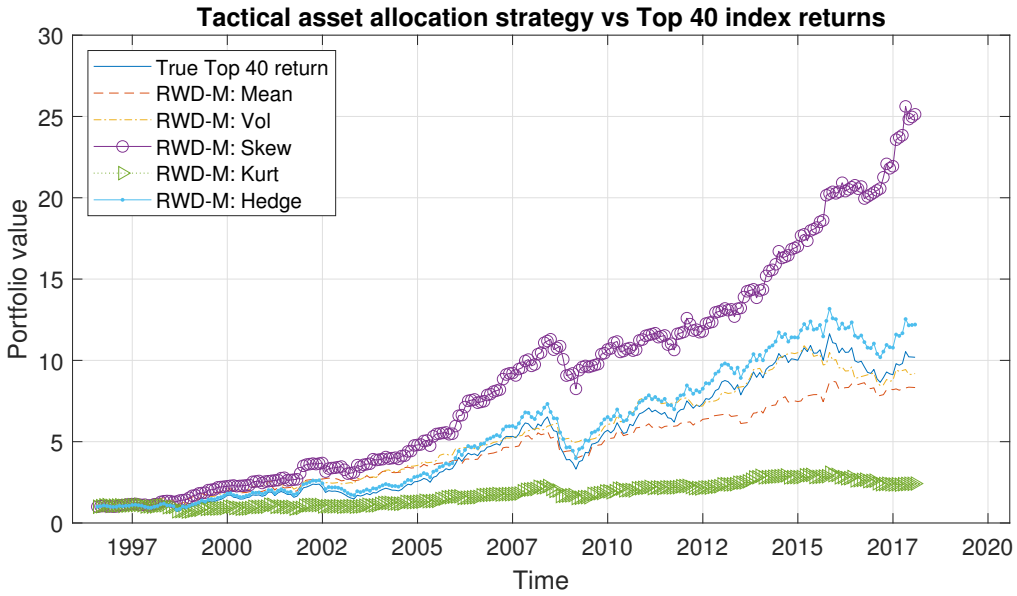
The cumulative portfolio value over the period August 1996 to January 2018 for the simple tactical asset allocation framework using the moments of the risk-neutral and real-world distributions is shown in Figure 6.1a and Figure 6.1b, respectively, if one unit of currency was invested in the Top40 on August 1996.

In particular, Figure 6.1a shows the portfolio using the forecasted risk-neutral volatility in returns as the signal in the tactical asset allocation framework yielded the best results. Moreover, the trading strategy using the risk-neutral kurtosis also outperformed the Top40, where the trading strategy using the risk-neutral mean return yielded similar results to the Top40. Furthermore, the skewness yielded similar result to around 2005, but thereafter yielded poor results. In Figure 6.1b, the real-world skewness in the simple tactical asset allocation framework yielded the best results, where the volatility and mean yielded similar returns to the Top40, while the kurtosis yielded poor results. Furthermore, Figure 6.1 shows that the tactical asset allocation based on the real-world skewness significantly outperformed the risk-neutral moments. In both the risk-neutral and real-world setting, the hedging strategy (with a fixed asset allocation as described above) involving the buying and selling of put and call options did not perform as well as the simple tactical asset allocation method. However, the hedging strategy did outperform the Top40 throughout the duration of the study.

In Table 6.1, we show some descriptive statistics of the annualised returns using the volatility in the risk-neutral tactical asset allocation framework (RND TAA), and the skewness in the real-world tactical asset allocation framework (RWD-M TAA). Furthermore, we also consider the tactical asset allocation of combining both signals from the risk-neutral volatility and



(a) A simple TAA using the risk-neurtal moments vs the Top 40 index returns.



(b) A simple TAA using the real-world moments vs the Top40 index returns.

Figure 6.1: Tactical asset allocation with withdrawals returns.

real-world skewness (RWD-M & RND TAA). That is, we hold the asset when the forecasted risk-neutral volatility is lower than the previous month's forecast and the forecasted real-world skewness is higher than the previous month's forecast. The hedging strategy, based on the risk-neutral volatility (RND Hedge) and real-world skewness (RWD-M Hedge), is also shown in Table 6.1.

Table 6.1: Descriptive statistics.

	Mean	Volatility	Sharpe Ratio	Skewness	Kurtosis
Top40	13.62%	19.90%	0.23	-0.28	5.75
RND TAA	14.85%	14.63%	0.40	0.62	8.23
RWD-M TAA	17.13%	13.01%	0.63	0.39	6.16
RWD-M & RND TAA	15.73%	8.74%	0.77	2.35	12.79
RND Hedge	13.86%	18.32%	0.27	-0.03	5.02
RWD-M Hedge	14.12%	17.89%	0.29	-0.16	4.25

The tactical asset allocation strategy involving the real-world skewness yielded the highest mean return over the sample period with a low variation in returns. The strategy involving the combination of the real-world and risk-neutral moments yielded the lowest variation in returns with a high expected return over our sample period. Low variation in returns, in conjunction with high expected returns, is obviously a desirable property for investment managers.

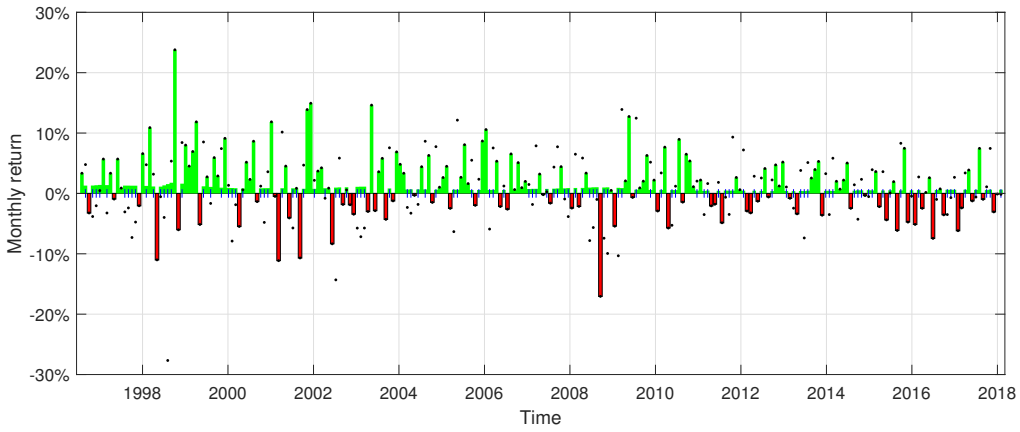
In Table 6.2, we show the number of trades carried out in each tactical asset allocation strategy shown in Table 6.1 over the total of 259 forecast months (or 21.5 years).

Table 6.2: Number of trades.

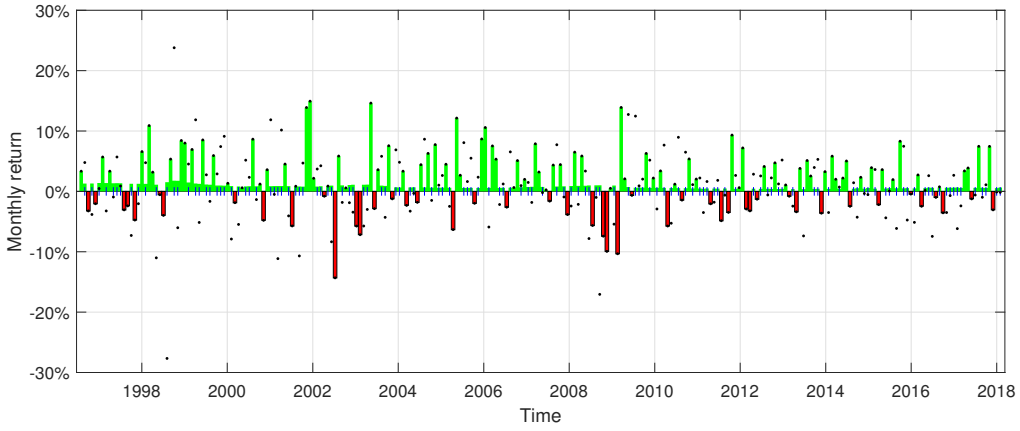
	RND TAA	RWD-M TAA	RWD-M & RND TAA
Number of trades	139	167	111

In Figure 6.2 we show the monthly returns of the TAA strategies. Furthermore, the ticker indicates a cash position. It is evident that RWD-M & RND TAA method is invested in cash more often than the other methods, which resulted in fewer negative returns over the period of study.[1]

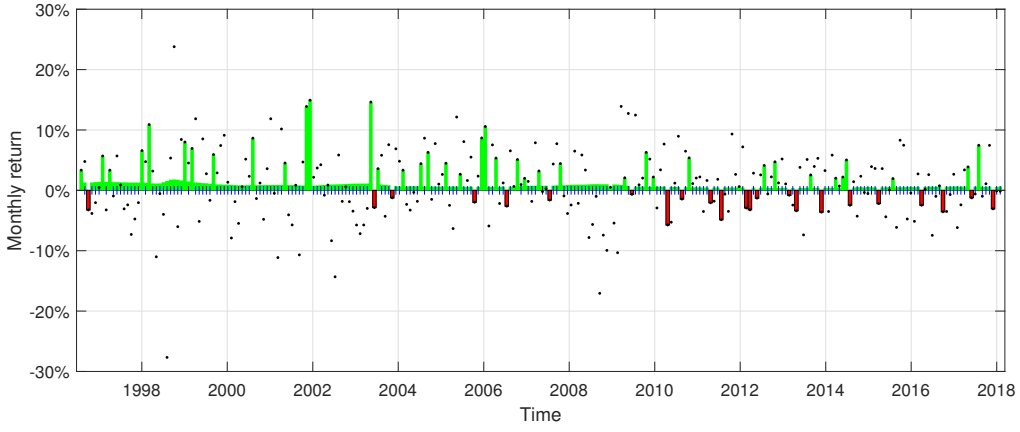
[1]Cash returns has outperformed equity returns for a number of short-term cycles during our study period in South Africa (see, e.g., Hugo, 2017), which is to the advantage of the RWD-M & RND TAA method.



(a) TAA Returns: RND



(b) TAA Returns: RWD-M



(c) TAA Returns: RND & RWD-M

Figure 6.2: The monthly returns for the tactical asset allocation methods. The blue ticker indicates a cash position and the black dots indicates the true Top40 index return.

In the next section, we examine safe withdrawal rates in a forward-looking environment using the tactical asset allocation framework.

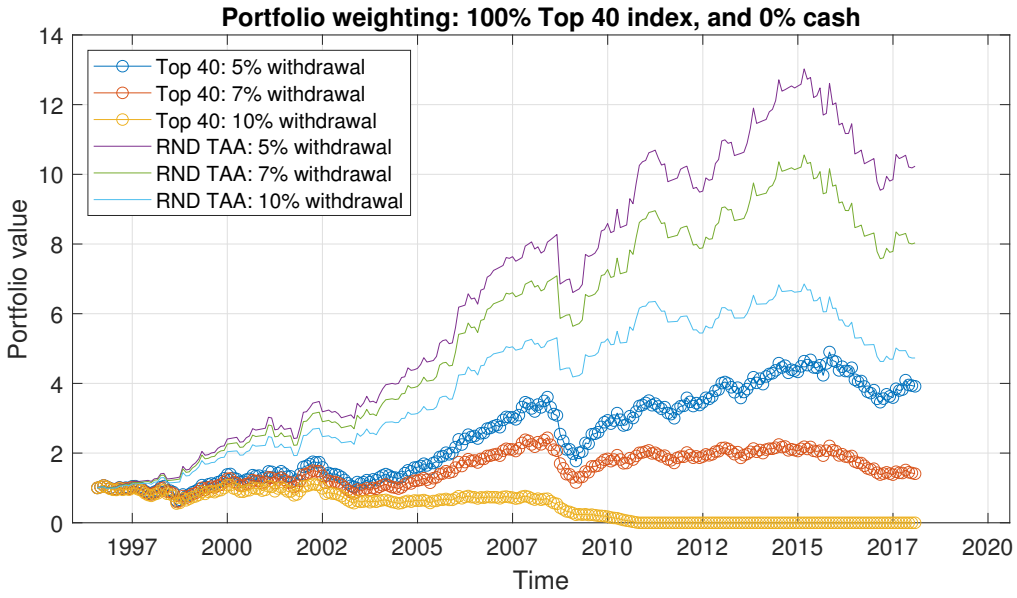
6.3 Results

In this section, we will assume that a person retired on 01 August 1996 with one unit in retirement savings. Furthermore, the retiree needs to decide how much to withdraw from the retirement fund; draw too much and carry the risk of running out of money, or draw too little and carry the risk of a compromised living standard. Therefore, in this section, we study the life expectancy of a basic retirement portfolio with three commonly used withdrawal rates used in the literature, namely 5%, 7%, and 10% per year of the initial portfolio size. Furthermore, these withdrawals will be adjusted monthly according to inflation rates and historical cash returns are used in the portfolio, which have been sourced form Firer and McLeod (1999), Firer and Staunton (2002), and I-Net.

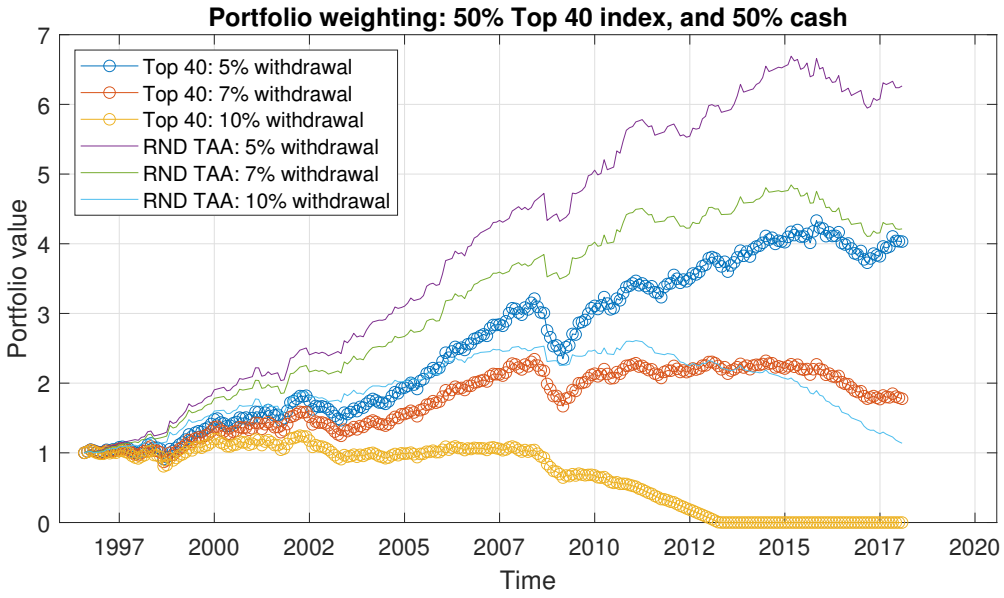
6.3.1 Safe Withdrawal Rates

In Figures 6.3, 6.4, and 6.5, we clearly see that the tactical asset allocation framework using the moments obtained from the forward-looking distributions outperformed the fixed asset allocation for the duration of the period under study. In particular, Figure 6.3 shows the accumulated portfolio value for two different asset allocations (see Figures 6.3a and 6.3b), with the risk-neutral volatility used in the tactical asset allocation framework described in the section above. Similarly, Figures 6.4 and 6.5 show the accumulated portfolio values using the real-world skewness and the combination of the real-world skewness and risk-neutral volatility in the tactical asset allocation framework, respectively. Furthermore, combining the signal from the risk-neutral volatility and the real-world skewness yielded superior fund prospects for higher withdrawal rates (see Figure 6.5). Although this strategy does not yield the highest mean return over the sample period (see Table 6.1), it has the least variation in returns. In particular, Scott et al. (2009), and Waring and Siegel (2015) criticised the notion of withdrawing a fixed real amount from an inherently volatile portfolio. This is known as sequence risk. Therefore, reducing the variation in returns is vitally important in determining safe retirement withdrawal

rates. This is particularly evident in Figure 6.5, where a high withdrawal rate of 10% yielded higher portfolio prospects than only using the real-world skewness.

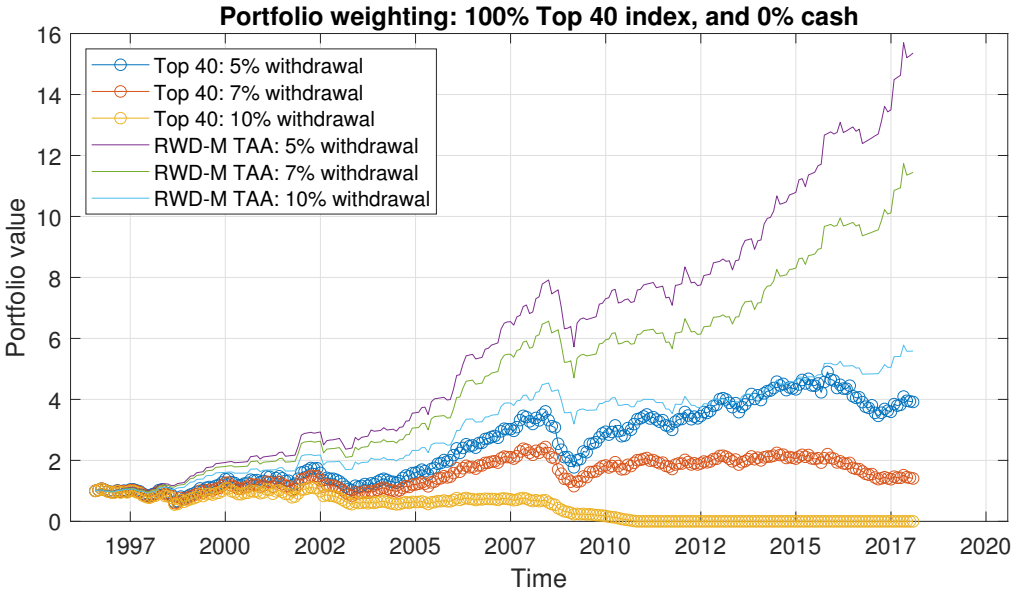


(a) 100% Equity

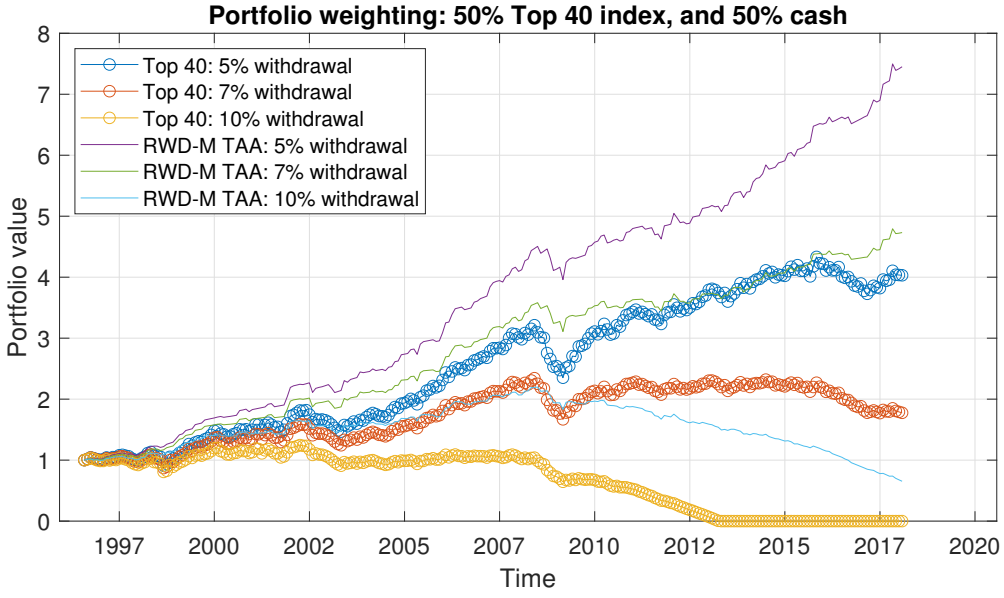


(b) 50% Equity, 50% Cash

Figure 6.3: The accumulated portfolio value for a fixed asset allocation vs. the risk-neutral tactical asset allocation (RND TAA) framework with withdrawal rates of {5%, 7%, 10%} returns.

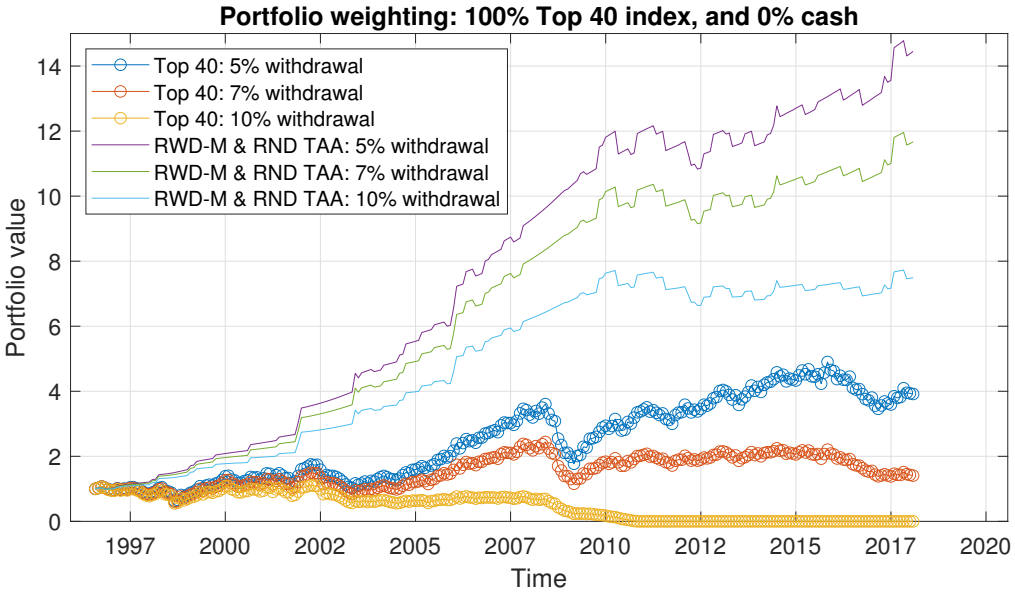


(a) 100% Equity

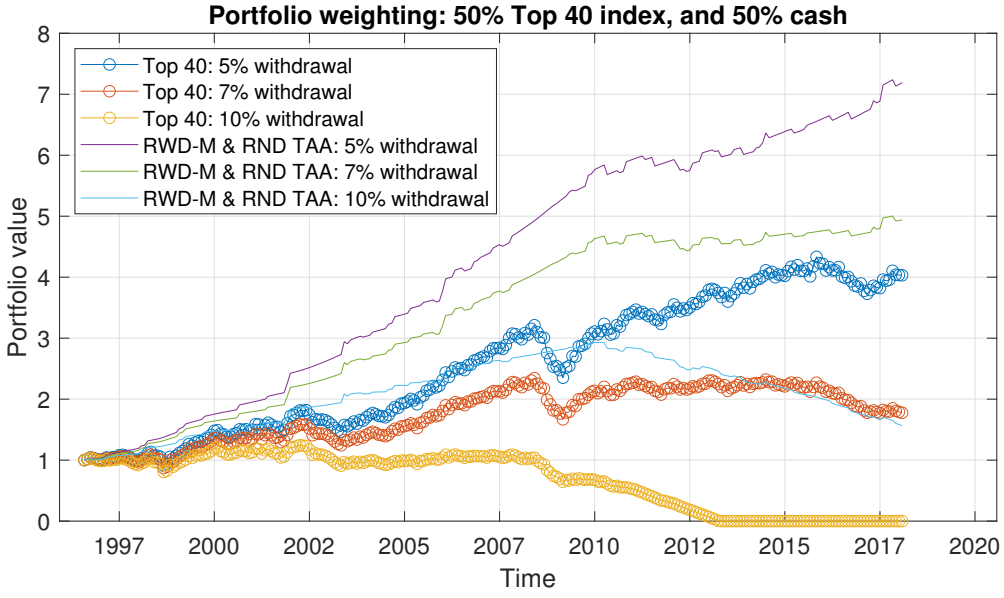


(b) 50% Equity, 50% Cash

Figure 6.4: The accumulated portfolio value for a fixed asset allocation vs. the real-world tactical asset allocation (RWD-M TAA) framework with withdrawal rates of {5%, 7%, 10%} returns.



(a) 100% Equity



(b) 50% Equity, 50% Cash

Figure 6.5: The accumulated portfolio value for a fixed asset allocation vs. the real-world and risk-neutral tactical asset allocation (RWD-M & RND TAA) framework with withdrawal rates of {5%, 7%, 10%} returns.

Next, we assess the robustness of the tactical asset allocation framework in determining safe retirement withdrawal rates. In particular, the robust analysis is carried out to determine how much, if any, of the improvements above the commonly quoted 4% safe withdrawal rate is attributed to using forward-looking information, rather than the different market or time periods used in this study.

6.3.2 Robust Analysis

In order to assess the robustness of the forward-looking distributions in modelling safe withdrawal rates, we carried out a random sampling study. We randomly selected, with replacement, a month from the sample period and used the equity, bonds, and cash returns to generate a one-month sample path. In the tactical asset allocation framework, we also used the previous month's forecasted moments of the randomly selected month to determine the portfolio asset allocation. We then continue to randomly sample from the period to simulate a 30-year period. We, therefore, simulate the evolution of the portfolio over a 30-year period and construct 10 000 such sample paths. This approach maintains the correlation structure between the assets, as we are using the true observed returns for the selected month for all assets. In Table 6.3, we show the success rates based on the fixed asset allocation versus the tactical asset allocation framework using the real-world skewness. We found similar results, in our sample, to the commonly quoted 4% safe withdrawal rate using historical (backward-looking) returns. However, by using forward-looking information in a tactical asset allocation framework, we were able to show significantly improved safe withdrawal rates. Furthermore, in Table 6.4, we calculated the fugit of the retirement portfolio. In this study, the fugit is defined as the expected duration of the portfolio given that the portfolio fails before the predefined 30-year duration.

It is evident from Table 6.3 and Table 6.4 that the simple tactical asset allocation framework, using the forward-looking real-world skewness, yielded superior success and fugit rates to the strategy that involved the fixed asset allocation. In particular, as the tactical asset allocation is based on either holding, or selling, the growth asset for a one-month period, the strategy is most prominent with a large growth asset allocation. Furthermore, these results illustrate that one can possibly achieve high portfolio success rates when making use of forward-looking information in the portfolio management.

Table 6.3: Success rates.

	Withdrawal rate						
Asset allocation	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
Fixed asset allocation:							
100% Stocks	79%	66%	52%	40%	30%	21%	14%
75% Stocks/25% Bonds	89%	77%	61%	45%	32%	20%	13%
50% Stocks/50% Bonds	96%	86%	70%	50%	31%	18%	10%
25% Stocks/75% Bonds	99%	92%	76%	54%	29%	13%	5%
Real-World TAA:							
100% Stocks	100%	98%	95%	86%	72%	59%	41%
75% Stocks/25% Bonds	100%	99%	96%	87%	71%	52%	31%
50% Stocks/50% Bonds	100%	100%	97%	85%	63%	37%	17%
25% Stocks/75% Bonds	100%	99%	93%	73%	44%	19%	6%

Table 6.4: Fugit.

	Withdrawal rate						
Asset allocation	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
Fixed asset allocation:							
100% Stocks	256	239	223	207	193	178	165
75% Stocks/25% Bonds	276	259	243	226	210	190	175
50% Stocks/50% Bonds	296	281	262	247	226	204	184
25% Stocks/75% Bonds	308	297	281	263	240	213	188
Real-World TAA:							
100% Stocks	294	283	268	257	242	226	208
75% Stocks/25% Bonds	322	308	296	276	258	239	218
50% Stocks/50% Bonds	272	313	304	288	268	245	217
25% Stocks/75% Bonds	340	314	303	285	262	235	205

6.4 Conclusion

In this study, we used forward-looking information, extracted from observed market-quoted derivative prices to determine safe retirement withdrawal rates. In particular, we extracted the forward-looking risk-neutral and real-world return distribution functions, where the distribution moments were used as a signal in a simple tactical asset allocation framework. We found that using forward-looking information in a tactical asset allocation framework yielded higher portfolio returns with a lower variation in returns compared to the portfolio with a fixed asset allocation.

It is well-known that many large financial firms frequently extract forward-looking information from derivative securities to infer market sentiment. Therefore, using a forward-looking modelling approach provided a more market consistent analysis of safe retirement withdrawal rates. We found that the portfolio based on the forward-looking real-world skewness in a tactical asset allocation framework supported safe withdrawal rates of up to 7% per annum (inflation adjusted). This strategy obtained similar success rates to the previously quoted 4% safe withdrawal rate determined from the fixed asset allocation based on historical returns. Thus, the performance of the real-world moments, used as a signal in a tactical asset allocation, allows for the possibility of higher withdrawal rates with high success rates. This confirms the usefulness of using forward-looking real-world moments in the management of retirement portfolios to improve the modelling of safe retirement withdrawal rates.

CHAPTER 7

SUMMARY AND CONCLUSION

In this thesis, we have studied several methods for extracting the forecast distribution from market quoted prices for decision making. In particular, we have extracted the historical return distribution, and the forward-looking risk-neutral and real-world return distributions using the South African Top40 index. It is well-known that distributions based solely on historical returns do not typically capture the current market sentiment, as they are backward-looking by nature. This was particularly evident during high volatility periods in our empirical study, such as, during the global financial crisis period.

Forward-looking risk-neutral distributions extracted from option prices are popular among many financial practitioners as they capture some market sentiment, and often provide a better forecast than historical based distributions (see, e.g., Shackleton et al., 2010; de Vincent-Humphreys and Noss, 2012; Christoffersen et al., 2013; Crisóstomo and Couso, 2018). However, risk-neutral distributions are typically biased estimators of the future asset return distribution, as it is used to recover traded option prices in a way that avoids arbitrage. Investors are typically risk-averse, and therefore the risk-neutral distribution does not accurately capture the true market sentiment of the possible return. However, the existence of bias in the risk-neutral distribution does not prevent option-implied information from being valuable for decision making (see, e.g., Christoffersen et al., 2013). In particular, by removing the bias in the estimator, by transforming the risk-neutral distribution into a real-world distribution, may also introduce bias in the distribution, as these methods typically involve making use of historical returns and assumptions on investor preference. This has largely hampered the use of real-world distributions by many financial practitioners. However, it is important to extract accurate real-world distributions as it contains more information than the risk-neutral distribution, such as the pricing kernel and risk-premium.

Extracting real-world forecast distributions is vitally important for forecasting and therefore formed part of our first research question. In particular, we firstly introduced the reader to methods of extracting the risk-neutral distribution from option prices and then transforming the distribution into a real-world distribution. These transformation methods include a parametric and non-parametric calibration function proposed by Liu et al. (2007) and Shackleton et al. (2010), respectively, and a behavioural transformation method proposed by Cristóstomo (2021). Secondly, we also considered the recovery theorem, proposed by Ross (2015), which recovers the real-world distribution from the richer risk-neutral transition probability matrix. In order to obtain a more accurate forecast distribution using the recovery theorem, we proposed a regularised multivariate Markov chain with prior information to estimate the risk-neutral transition probability matrix more accurately (see, Van Appel and Maré, 2018). Our empirical findings showed that this method yielded better results than the distribution extracted using the basic recovery theorem.

Research into the forecasting ability of the real-world return distribution is scarce in the literature and therefore formed part of our second research question. In our empirical analysis, we extracted weekly one-month forecast distributions for the South African JSE/FTSE Top 40 index over the period September 2005 to January 2018 for the methods studied in this thesis (see, Van Appel and Maré, 2021, 2020b). The accuracy of these return distributions were then evaluated and compared by carrying out several backtests using PIT-based goodness-of-fit tests and VaR tests. We found that certain models performed better during certain economic conditions. Therefore, we further studied the mixture of historical and option-implied information to improve the accuracy of the forecast distribution. In order to optimally mix two distributions, we proposed two mixing parameters that are both functions of the implied volatility, as it is considered a reliable proxy of the market's future state. In our empirical study, we found that the extracted real-world and mixture distributions yielded superior forecasts to the risk-neutral distribution.

Key features of the real-world distribution, such as the higher order moments, and measures such as the change in the risk-premium and pricing kernel, have become extremely powerful in forecasting future returns (see, e.g., Cuesdeanu and Jackwerth, 2018a; de Vincent-Humphreys and Noss, 2012; Christoffersen et al., 2013). In particular, the pricing kernel, risk-premium and higher moments are all related to preferences, and thereby identifying the changes in these

measures provides information about the change in the market's risk aversion (or sentiment) (see, eg., Cuesdeanu and Jackwerth, 2018a). Therefore, in this thesis, we applied the extracted distributions to the problem of modelling safe retirement withdrawal rates. This formed part of our third research question. Our empirical findings showed that a portfolio with a fixed asset allocation typically supported a safe inflation adjusted withdrawal rate of 4% per annum (see, Van Appel et al., 2021). This corresponds to the commonly advised safe withdrawal rates reported in local and international literature (see, e.g., Cooley et al., 1998, 1999, 2003; Bengen, 1994; Scott et al., 2009; Maré, 2016; Van Appel et al., 2021). These studies typically involved bootstrapping from historical returns for each asset class. However, by using forward-looking information, in a tactical asset allocation framework, we showed that it is possible to achieve safe withdrawal rates of up to 7% (see, Van Appel and Maré, 2022). This is obviously more desirable for retirees, as retirees would typically like to draw as much as possible, to not compromise their living standards, but with a low probability of running out of funds.

We leave as further research a number of extensions. Firstly, the mixing of forward-looking and historical information is a desirable method of obtaining more accurate forecasts. However, more research can be done in optimally mixing the historical and forward-looking information (see, e.g., Kapetanios et al., 2015; Timmermann, 2006). Secondly, further research could be done in testing the forecasting performance of other forward-looking measures in determining an optimal safe retirement withdrawal rate. That is, proxies such as the risk-premium and behavioural parameters are vitally important when constructing an optimal trading strategy, which could be used in the management of retirement funds. Furthermore, the use of forward-looking distributions could also be used to optimally incorporating life annuities into a retirement portfolio which will also be a valuable contribution (see, e.g., Dus et al., 2005).

APPENDIX A

THE HESTON AND BATES CLOSED-FORM EUROPEAN CALL OPTION PRICING FORMULA

A.1 The Characteristic Function for the Heston Model

The closed-form characteristic function for the Heston model is given as (see, e.g., Heston, 1993; Kienitz and Wetterau, 2012):

$$\psi(u) = \exp[A(u,t,T) + B(u,t,T)V_t + iu\{\log(S_t) + (r-d)(T-t)\}], \tag{A.1}$$

where

$$A(u,t,T) = \frac{\kappa\theta}{\nu^2}\left[(\kappa - \rho\nu ui - D)(T-t) - 2\log\left(\frac{G\exp[-D(T-t)]}{G-1} - 1\right)\right], \tag{A.2}$$

$$B(u,t,T) = \frac{\kappa - \rho\nu ui - D}{\nu^2}\left(\frac{1-\exp[-D(T-t)]}{1-G\exp[-D(T-t)]}\right), \tag{A.3}$$

$$G = \frac{\kappa - \rho\nu ui - D}{\kappa - \rho\nu ui + D}, \tag{A.4}$$

and

$$D = \sqrt{(\kappa - \rho\nu ui)^2 + u(i+u)\nu^2}. \tag{A.5}$$

A.2 The Characteristic Function for the Bates Model

The closed-form characteristic function for the Bates model is give as (see, e.g., Bates, 1996; Kienitz and Wetterau, 2012):

$$\psi(u) = \exp[A(u,t,T) + B(u,t,T)V_t + iu\{\log(S_t) + (r-d)(T-t)\} + H(u,t,T)], \quad \text{(A.6)}$$

where

$$A(u,t,T) = \frac{\kappa\theta}{\nu^2}\left[(\kappa - \rho\nu ui - D)(T-t) - 2\log\left(\frac{G\exp[-D(T-t)]}{G-1} - 1\right)\right], \quad \text{(A.7)}$$

$$B(u,t,T) = \frac{\kappa - \rho\nu ui - D}{\nu^2}\left(\frac{1-\exp[-D(T-t)]}{1-G\exp[-D(T-t)]}\right), \quad \text{(A.8)}$$

$$G = \frac{\kappa - \rho\nu ui - D}{\kappa - \rho\nu ui + D}, \quad \text{(A.9)}$$

$$D = \sqrt{(\kappa - \rho\nu ui)^2 + u(i+u)\nu^2} \quad \text{(A.10)}$$

$$H(u,t,T) = \lambda(T-t)\left(-aiu + \left(\exp\left(iu\log(1+a) + \frac{1}{2}b^2 iu(iu-1)\right) - 1\right)\right), \quad \text{(A.11)}$$

$$\mu_J = \log(1+a) - \frac{b^2}{2}, \quad \sigma_J = b, \quad a > -1, \text{ and } b \geq 0. \quad \text{(A.12)}$$

A.3 Closed-form European Call Option Pricing Formula

Using the characteristic function, ψ, we can numerically determine the value of a European call option as follows (see, e.g., Carr and Madan, 1999; Rouah, 2013):

$$\mathrm{C}(T,K) = e^{-dT}S_0 P_1 - e^{-rT}KP_2, \quad \text{(A.13)}$$

where the delta of the option, P_1, is:

$$P_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-iu\log K}\psi(u-i)}{iu\psi(-i)}\right]\mathrm{d}u, \quad \text{(A.14)}$$

and the risk-neutral probability of finishing in-the-money, P_2, is:

$$P_2 = P(S_T > K) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-iu\log K}\psi(u)}{iu}\right]\mathrm{d}u. \quad \text{(A.15)}$$

For optimal numerical methods for solving (A.13), we refer the interested reader to Carr and Madan (1999), Kienitz and Wetterau (2012, Chapter 5), and Rouah (2013, Chapter 3).

APPENDIX B

ADDITIONAL VAR BACKTESTING RESULTS

This appendix formed part of Van Appel and Maré (2020b). The recovery theorem with application to risk management, *South African Statistical Journal*, **54**(1): 65-91.

In this appendix, we give a short description of the VaR backtests that is part of the MATLAB Risk Management Toolbox (2018).

- The traffic light (TL) test classifies the number of failures into three zones, namely, green, yellow, and red using a binomial distribution, $F(x|n,p)$ (see, Basle Committee of Banking Supervision, 2011). In particular, the test computes the cumulative probability of observing up to x failures in n trails, with $p = \alpha$ and three zones:
  - Green: $F(x|n,p) \leq 0.95$
  - Yellow: $0.95 < F(x|n,p) \leq 0.9999$
  - Red: $F(x|n,p) > 0.9999$.

  This test is often used as a preliminary VaR accuracy check.

- The binomial (Bin) distribution test is an extension of Christoffersen (1998) Bernoulli test. It states that if $\mathbb{I}_t$ are i.i.d Bernoulli with parameter p, then the total number of failures, x, follows a binomial distribution with mean and variance equal to np and $np(1-p)$ respectively. Under the null hypothesis, $H_0 : p = \alpha$, the test statistic is approximated by

$$z = \frac{x - np}{\sqrt{np(1-p)}}, \tag{B.1}$$

which has a standard normal distribution.

- The proportion of failures (POF) test is a *LR* test proposed by Kupiec (1995). More specifically, the POF test determines whether the proportion of failures (i.e., number of failures divided by number of observations) denoted as $\hat{p}$ is consistent with the VaR confidence level. Under the null hypothesis, $H_0 : p = \alpha$, the *LR* test statistics is:

$$LR_{\text{POF}} = -2\log\left[(1-p)^{n-x}p^x\right] + 2\log\left[(1-\hat{p})^{n-x}(\hat{p})^x\right] \sim \chi^2(1). \tag{B.2}$$

- The time until first failure (TUFF) test, proposed by Kupiec (1995), is a *LR* test that measures the time until the first failure. Under the null hypothesis, $H_0 : p = 1/v$, where v is the time until the first failure in the sample, the *LR* test statistic is

$$LR_{\text{TUFF}} = -2\log\left[\frac{p(1-p)^{v-1}}{\hat{p}(1-\hat{p})^{v-1}}\right] \sim \chi^2(1). \tag{B.3}$$

  The TUFF test is mostly used as a preliminary test to the POF test. Furthermore, it only considers the number of failures but not the time dynamics of the failures. The test also has been shown to have a low power in identifying poor VaR models.

- The conditional coverage independence (CCI) test, also known as the Markov test, assesses whether the probability of VaR failure for any given period is dependent on the outcome of the previous period (see, Christoffersen, 1998). Using the indicator value in (4.4) and let $N_{i,j}$, $i = 0,1$, $j = 0,1$ be the number of periods in which state j occurred after state i occurred. Then let π_0 be the conditional probability of having a failure at time t, given that there was no failure at time $t-1$. Similarly, let π_1 be the conditional probability of having a failure at time t, given that there was a failure at time $t-1$. Under $H_0 : \pi_0 = \pi_1$, the *LR* test statistic is given as:

$$\begin{aligned} LR_{\text{CCI}} = &-2\log\left[(1-\pi)^{N_{00}+N_{01}}\pi^{N_{01}+N_{11}}\right] \\ &+2\log\left[(1-\pi_0)^{N_{00}}\pi_0^{N_{01}}(1-\pi_1)^{N_{10}}\pi_1^{N_{11}}\right] \sim \chi^2(1), \end{aligned} \tag{B.4}$$

  where $\pi = \pi_0 + \pi_1$.

- The conditional coverage (CC) mixed test is a combination of the CCI test and the POF test. The CC test assesses whether the failures are independent and whether the correct failure rate is obtained (see, Christoffersen, 1998). The *LR* test statistic is

$$LR_{\text{CC}} = LR_{\text{CCI}} + LR_{\text{POF}} \sim \chi^2(2). \tag{B.5}$$

A VaR model must therefore satisfy both independence and the correct failure rate in this test, making this test appealing to practitioners.

- The time between failures independence (TBTI) test proposed by Haas (2001) is an extension of Kupiec's time until first failure (TUFF) test by not only testing the time until the first failure, but also the time between all failures. Under the null hypothesis, that failures are independent from each other, the *LR* test statistic is

$$LR_{TBFI} = \sum_{i=2}^{x} \left[-2\log\left(\frac{p(1-p)^{v_i-1}}{\hat{p}(1-\hat{p})^{v_i-1}} \right) \right] - 2\log\left[\frac{p(1-p)^{v-1}}{\hat{p}(1-\hat{p})^{v-1}} \right] \sim \chi^2(x), \tag{B.6}$$

  where v_i denotes the duration between the ith and $(i-1)$th failure, v the time until the first failure and x the number of failures in the sample.

- The time between failures (TBF) likelihood ratio test, introduces by Haas (2001), is a mixed *LR* test. Under the null hypothesis, that the correct failure rate is obtained and that the failures are independent, the test statistic is

$$LR_{\text{TBF}} = LR_{\text{POF}} + LR_{\text{TBFI}}. \tag{B.7}$$

  This test statistics is $\chi^2(x+1)$ distributed, where x is the number of failures. The advantage to this test is that it is robust, since it identifies problems in dependencies and the number of failures.

Using the MATLAB Risk Management Toolbox (2018), we show the backtest results obtained for the monthly one-month $\text{VaR}_{(0.95)}$ and $\text{VaR}_{(0.90)}$ in Tables B.1 and B.2, respectively. In addition, Table B.4 shows the weekly one-week $\text{VaR}_{(0.90)}$.

Table B.1: Goodness-of-fit: one-month $\text{VaR}_{(0.95)}$ backtests.

Panel A: Monthly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Monthly (Sep 2005 - Jan 2018)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	reject	accept	reject	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept

Panel B: Weekly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	accept	reject	reject	reject
RND	green	reject	reject	accept	reject	accept	reject	accept
Heston	green	accept	accept	accept	reject	reject	reject	reject
Bates	green	accept	accept	accept	reject	reject	reject	reject
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	reject	accept	accept	accept	accept	accept
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	accept	reject	reject	reject	reject
Historical HW	yellow	accept	accept	accept	reject	reject	reject	reject
RND	green	accept	accept	accept	reject	reject	reject	reject
Heston	green	accept	accept	accept	reject	reject	reject	reject
Bates	green	accept	accept	accept	reject	reject	reject	reject
RWD	green	accept	accept	accept	reject	reject	reject	reject
RWD-M	green	accept	accept	accept	reject	reject	reject	reject
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	reject	reject	accept	reject	reject
Heston	green	reject	reject	reject	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	reject
RWD-M	green	reject	reject	accept	reject	accept	reject	accept
Weekly one-Month returns: Sep 2005 - Jan 2018								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	accept	reject	reject	reject	reject
Heston	green	reject	reject	accept	reject	reject	reject	reject
Bates	green	reject	reject	accept	reject	reject	reject	reject
RWD	green	reject	reject	accept	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject

Table B.2: Goodness-of-fit: one-month $\text{VaR}_{(0.90)}$ backtests.

Panel A: Monthly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Monthly (Sep 2005 - Jan 2018)								
Historical	green	accept	accept	accept	accept	accept	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept

Panel B: Weekly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	accept	reject	reject	reject	reject
Heston	green	accept	reject	accept	reject	reject	reject	reject
Bates	green	accept	reject	accept	reject	reject	reject	reject
RWD	green	reject	reject	accept	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	accept	accept	accept	reject	reject	reject	reject
Heston	green	accept	accept	accept	reject	reject	reject	reject
Bates	green	accept	accept	accept	reject	reject	reject	reject
RWD	green	accept	accept	accept	reject	reject	reject	reject
RWD-M	green	accept	accept	accept	reject	reject	reject	reject
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	reject	reject	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	reject	reject	reject	reject	reject
Heston	green	reject	reject	reject	reject	reject	reject	reject
Bates	green	reject	reject	reject	reject	reject	reject	reject
RWD	green	reject	reject	reject	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject
Weekly one-Month returns: Sep 2005 - Jan 2018								
Historical	green	reject	reject	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	accept	reject	reject	reject	reject
Heston	green	reject	reject	accept	reject	reject	reject	reject
Bates	green	reject	reject	accept	reject	reject	reject	reject
RWD	green	reject	reject	accept	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject

Table B.3: Goodness-of-fit: weekly one-week $\text{VaR}_{(0.95)}$ backtests.

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-week returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	accept	accept	accept	accept
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	accept	accept	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	reject	accept	accept	reject	reject
Historical HW	green	accept	accept	reject	accept	accept	reject	reject
RND	green	accept	accept	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	reject	reject	accept	reject	accept	reject	accept
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	accept
RWD-M	green	reject	reject	accept	reject	accept	accept	accept
Weekly one-week returns: Sep 2005 - Jan 2018								
Historical	green	accept	accept	accept	accept	accept	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	reject
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	accept
RWD-M	green	reject	reject	accept	reject	accept	accept	accept

Table B.4: Goodness-of-fit: weekly one-week $\text{VaR}_{(0.90)}$ backtests.

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-week returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	accept	accept	accept	accept
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	accept	reject	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	reject	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	reject	reject	reject	reject	reject
Historical HW	green	accept	accept	reject	accept	accept	accept	accept
RND	green	accept	accept	accept	accept	accept	accept	accept
Heston	green	accept	accept	reject	accept	accept	accept	accept
Bates	green	accept	accept	reject	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	accept	accept	accept	accept	accept	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	reject
RWD-M	accept	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Sep 2005 - Jan 2018								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	reject
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	accept
RWD-M	green	reject	reject	accept	accept	accept	accept	accept

BIBLIOGRAPHY

Abuizam, R. (2009). A risk-based model for retirement planning. *Journal of Business and Economics Research*, 7(6):31–43.

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, 53(2):499–547.

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1–2):9–51.

Alexander, C. (2008). *Value-at-Risk Models*. John Wiley and Sons Ltd.

Allianz (2010). Outliving Your Money Feared More Than Death. Online.

Audrino, F., Huitema, R., and Ludwig, M. (2014). An empirical analysis of the Ross recovery theorem. Available at SSRN: https://ssrn.com/abstract=2433170 or http://dx.doi.org/10.2139/ssrn.2433170.

Backwell, A. (2015). State prices and implementation of the recovery theorem. *The Journal of Risk and Financial Management*, 8:2–16.

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *The Review of Financial Studies*, 16(2):527–566.

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

Bakshi, G. and Madan, D. (2006). A theory of volatility spreads. *Management Science*, 52(12):1945–1956.

Barone-Adesi, G. (2016). VaR and CVaR implied in option prices. *Journal of Risk and Financial Management*, 9(2):1–6.

Barone-Adesi, G., Fusari, N., Mira, A., and Sala, C. (2019). Option market trading activity and the estimation of the pricing kernel: A Bayesian approach. *Journal of Econometrics*, 216(2):430–449.

Barone-Adesi, G., Mancini, L., and Shefrin, H. (2017). Estimating sentiment, risk aversion, and time preference from behavioral pricing kernel theory. *Swiss Finance Institute Research Paper No. 12-21, Available at SSRN*.

Basle Committee of Banking Supervision (2011). *Supervisory Framework For The Use of "Backtesting" in Conjunction With The Internal Models Approach to Market Risk Capital Requirements*. Available at www.bis.org.

Bates, D. S. (1991). The crash of '87: Was it expected? the evidence from options markets. *The Journal of Finance*, 46(3):1009–1044.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *The Review of Financial Studies*, 9(1):69–107.

Bengen, W. (1994). Determining withdrawal rates using historical data. *Journal of Financial Planning*, 7(1):171–180.

Berkowitz, J. (2001). Testing density forecasts, with apllications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

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

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):673–654.

Blackburn, C. and Sherris, M. (2013). Consistent dynamic affine mortality models for longevity risk applications. *Insurance: Mathematics and Economics*, 53:64–73.

Blanchett, D. M., Finke, M., and Pfau, W. D. (2013). Low bond yields and safe portfolio withdrawal rates. *The Journal of Wealth Management*, 16(2):55–62.

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

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22(11):4463–4492.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651.

Brigo, D. and Mercurio, F. (2002). Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, 5(4):427–446.

Bunn, D. (1984). *Applied Decision Analysis*. McGraw-Hill, New York.

Butler, M. and van Zyl, C. (2012a). Consumption changes on retirement for South African households. *South African Actuarial Journal*, 12:31–64.

Butler, M. and van Zyl, C. (2012b). Retirement adequacy goals for South African households. *South African Actuarial Journal*, 12:1–29.

Butler, M. B. J., Hu, B., and Kloppers, D. (2013). A comparison of probability of ruin and expected discounted utility as objective functions for choosing a post-retirement investment strategy. *South African Actuarial Journal*, 13:185–219.

Carr, P. and Madan, D. B. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 24:61–73.

Carr, P. and Madan, D. B. (2000). Factor models for option pricing. Unpublished paper.

CBOE (2009). The CBOE Volatility Index - VIX. Technical report, Chicago Board Options Exchange.

CBOE (2011). The CBOE Skew index - SKEW. Technical report, Chicago Board Options Exchange.

Chakraborty, S. (2015). Generating discrete analogues of continuous probability distributions - A survey of methods and constructions. *Journal of Statistical Distributions*, 2(6):1–30.

Chen, B. and Gan, Q. (2018). Sentiment, implied volatility slope, and risk-neutral skewness. *Available at SSRN*.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Christoffersen, P., Jacobs, K., and Chang, B. Y. (2013). Forecasting with option-implied information. *Handbook of Economic Forecasting*, 2:581–656.

Chvátal, V. (1983). *Linear Programming*. W.H. Freeman & Co Ltd.

Conrad, J., Dittmar, R. F., and Ghysels, E. (2013). Ex ante skewness and expected stock returns. *The Journal of Finance*, 68(1):85–124.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Cooley, P., Hubbard, C., and Walz, D. (1998). Retirement savings: Choosing a withdrawal rate that is sustainable. *Journal of the American Association of Individual Investors*, 10(1):40–50.

Cooley, P., Hubbard, C., and Walz, D. (1999). Sustainable withdrawal rates from your retirement portfolio. *Financial Counselling and Planning*, 20(2):16–21.

Cooley, P., Hubbard, C., and Walz, D. (2003). A comparative analysis of retirement portfolio success rates: Simulation versus overlapping periods. *Financial Services Review*, 12:115–128.

Crisóstomo, R. and Couso, L. (2018). Financial density forecasts: A comprehensive comparison of risk-neutral and historical schemes. *Journal of Forecasting*, 37(5):589–603.

Cristóstomo, R. (2021). Estimating real-world probabilities: A forward-looking behavioral framework. *Journal of Futures Markets*, pages 1–27.

CSI (2017). Report on pensioner mortality 2005-2010. Technical report, Cape Town: Continuous Statistical Investigation Committee, Actuarial Society of South Africa.

Cuesdeanu, H. and Jackwerth, J. C. (2018a). The pricing kernel puzzle in forward looking data. *Review of Derivatives Research*, 21:253–276.

Cuesdeanu, H. and Jackwerth, J. C. (2018b). The pricing kernel puzzle: survey and outlook. *Annals of Finance*, 14:289–329.

de Vincent-Humphreys, R. and Noss, J. (2012). Estimating probability distributions of future asset prices: empirical transformations from option-implied risk-neutral to real-world density functions. *Working Paper No. 455, Bank of England*.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuaracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Dillschneider, Y. and Maurer, R. (2019). Functional Ross recovery: Theoretical results and empirical tests. *Journal of Economic Dynamics & Control*, 108:103750.

Dimson, E., Marsh, P., and Staunton, M. (2016). Credit Suisse Global Investment Returns Sourcebook 2016. Zurich: Credit Suisse Research Institute.

Dobiáš, V. (2007). Empirical testing of local cross entropy as a method for recovering asset's risk-neutral pdf from option prices. In Miller, J., Edelman, D., and Appleby, J., editors, *Numerical Methods for Finance*, chapter 8, pages 150–172. Chapman and Hall/CRC, first edition.

Dus, I., Maurer, R., and Mitchell, O. S. (2005). Betting on death and capital markets in retrement: A shortfall risk analysis of life annuities versus phased withdrawal plans. *Financial Services Review*, 14:169–196.

Ennart, H. (2012). *Åldrandets gåta [The Mystery of Aging]*. Stockholm: Ordfront.

Fackler, P. L. and King, R. P. (1990). Calibration of option-based probability assessments in agricultural commodity markets. *American Journal of Agricultural Economics*, 72(1):73–83.

Figlewski, S. (2010). Estimating the implied risk neutral density for the U.S. market portfolio. In Bollerslev, T., Russell, J. R., and Watson, M., editors, *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, chapter 15, pages 323–353. Oxford University Press.

Firer, C. and McLeod, H. (1999). Equities, bonds, cash and inflation: Historical performance in South Africa 1925 to 1998. *Investment Analysts Journal*, 28(50):7–28.

Firer, C. and Staunton, M. (2002). 102 Years of South African financial market history. *Investment Analysts Journal*, 31(56):57–65.

Flint, E. and Maré, E. (2017). Estimating option-implied distributions in illiquid markets and implementing the Ross recovery theorem. *South African Actuarial Journal*, 17(1):1–28.

Garman, M. (1989). Semper tempus fugit. *Risk*, 2(5):34–35.

Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parametrization with application to the valuation of volatility derivatives. *Presentation at Global Derivatives and Risk Management, Madrid*.

Gatheral, J. and Jacquier, A. (2014). Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71.

Haas, M. (2001). New methods in backtesting. Working Paper, Financial Engineering Research Center. URL: www.ime.usp.br/rvicente/risco/haas.pdf.

Han, B. (2008). Investor sentiment and option prices. *The Review of Financial Studies*, 21(1):387–414.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Hollstein, F., Prokopczuk, M., Tharann, B., and Simen, C. W. (2019). Predicting the equity market with option-implied variables. *The European Journal of Finance*, 25(10):937–965.

Hugo, P. (2017). Don't be tempted by cash returns. Technical report, Prudential Investment Managers, https://www.investonline.co.za/wp-content/uploads/2013/06/Dont-be-tempted-by-cash-returns-Pieter-Hugo.pdf.

Hull, J. (2006). *Options, Futures and Other derivatives*. Prentice Hall, 6 edition.

Hull, J. and White, A. (1998a). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, 1(1):5–19.

Hull, J. and White, A. (1998b). Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives*, 5(3):9–19.

Institute and Faculty of Actuaries (2019). Savings goals for retirement: Policy briefing. https://www.actuaries.org.uk/system/files/field/document/Saving

Jackwerth, J. C. and Menner, M. (2020). Does the Ross recovery theorem work empirically. *Journal of Financial Economics*, 137(3):723–739.

Kapetanios, G., Mitchell, J., Price, S., and Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, 188(1):150–165.

Kienitz, J. and Wetterau, D. (2012). *Financial Modelling: Theory, Implementation and Practice*. Wiley.

Kiriu, T. and Hibiki, N. (2019). Estimating forward looking distribution with the Ross recovery theorem. *Journal of the Operations Research Society of Japan*, 62(2):83–107.

Klein, S. and Sapra, S. (2020). Optimal asset allocation, asset location and drawdown in retirement. *Quantitative Research and Analytics*, pages 1–15.

Knüppel, M. (2015). Evaluating the calibration of multi-step-ahead density forecasts using raw moments. *Journal of Business & Economic Statistics*, 33(2):270–281.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3(2):73–84.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U. S. mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Lewis, A. L. (2019). Option-based equity risk premiums. Working paper: arXiv:1910.14522.

Liu, X., Shackleton, M. B., Taylor, S. J., and Xu, X. (2007). Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking & Finance*, 31(5):1501–1520.

Malz, A. M. (2014). A simple and reliable way to compute option-based risk-neutral distributions. Technical report, Federal Reserve Bank of New York Staff Report No. 677.

Maré, E. (2016). Safe spending rates for South African retirees. *South African Journal of Science*, 112(1/2):1–4.

MATLAB Risk Management Toolbox (2018). MATLAB Risk Management Toolbox R2018b. Retrieved: https://www.mathworks.com/products/risk-management.html.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press.

Melick, W. R. and Thomas, C. P. (1997). Recovering an asset's implied pdf from option prices: An application to crude oil during the gulf crises. *Journal of Financial and Quantitative Analysis*, 32(01):91–115.

Merton, R. C. (1973). The of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183.

Meucci, A. (2005). *Risk and Asset Allocation*. Springer Finance.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM.

Milevsky, M. A. and Huang, H. (2011). Spending retirement on planet vulcan: The impact of longevity risk aversion on optimal withdrawal rates. *Financial Analysts Journal*, 67(2):45–58.

Olivieri, A., Thirurajah, S., and Ziveyi, J. (2022). Target volatility strategies for group self-annuity portfolios. *ASTIN Bulletin*, pages 1–27.

Pérignon, C. and Smith, D. R. (2010). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking & Finance*, 34(2):362–377.

Purdy, M. (2015). The first person to live to 150 has already been born. Technical report, Innovation Hub.

Rebonato, R. (2002). *Modern Pricing of Interest-Rate Derivatives: The LIBOR Market Model and Beyond*. Princeton University Press.

Richman, R. and Velcich, G. (2020). Mortality improvements in South Africa: insights from pensioner mortality. Presented at the Actuarial Society of South Africa's 2020 Virtual Convention, 6-8 October 2020.

Richman, R. D. (2017). Old age mortality in South Africa, 1985-2011. Master's thesis, University of Cape Town.

Ritchey, R. (1990). Call option valuation for discrete normal mixtures. *Journal of Financial Research*, 13(4):285–296.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–648.

Rouah, F. D. (2013). *The Heston Modeland Its Extensions in Matlab and C#*. Wiley.

Rusconi, R. (2020). Regulating South Africa's retirement funds: the case for clearer objectives. Presented at the Actuarial Society of South Africa's 2020 Virtual Convention.

Sanford, A. (2018). Recovery theorem with a multivariate Markov chain. Available at SSRN: https://ssrn.com/abstract=3247294 or http://dx.doi.org/10.2139/ssrn.3247294.

Scott, J., Sharpe, W., and Watson, J. (2009). The 4% rule - at what price? *Journal of Investment Management*, 7(3):31–48.

Scott, M. (1996). Assessing your portfolio allocation from a retiree's point of view. *Journal of the American Association of Individual Investors*, pages 16–19.

Shackleton, M. B., Taylor, S. J., and Yu, P. (2010). A Multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking & Finance*, 34(11):2678–2693.

Shimko, D. (1993). Bounds of probability. *Risk (Concord, NH)*, 6(4):33–37.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Smales, L. A. (2017). The importance of fear: investor sentiment and stock market returns. *Applied Economics*, 49(34):3395–3421.

Spears, T. J. (2013). On estimating the risk-neutral and real-world probability measures. *Masters Thesis*, University of Oxford.

Statman, M., Thorley, S., and Vorkink, K. (2006). Investor overconfidence and trading volume. *The Review of Financial Studies*, 19(4):1531–1565.

Thomas, P. J. (2016). Measuring risk-aversion: The challenge. *Mesurement*, 79:285–301.

Thomson, D. and Van Vuuren, G. (2016). Forecasting the South African business cycle using fourier analysis. *International Business & Economics Research Journal*, 15(4):175–192.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4(4):1035–1038.

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, chapter 4, pages 135–196. Elsevier.

Van Appel, V. and Maré, E. (2018). The Ross recovery theorem with a regularised multivariate Markov chain. *ORiON*, 34(2):133–155.

Van Appel, V. and Maré, E. (2020a). Die herwinningstelling met toepassing op risikobestuur. *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie*, 39(1):133.

Van Appel, V. and Maré, E. (2020b). The recovery theorem with application to risk management. *South African Statistical Journal*, 54(1):65–91.

Van Appel, V. and Maré, E. (2021). Estimation of forecast distributions: A comparison of risk-neutral and real-world distributions in the South-African market. *Working paper*.

Van Appel, V. and Maré, E. (2022). Determining safe retirement withdrawal rates using forward-looking distributions. *South African Journal of Science*, 118(3/4):38–44.

Van Appel, V., Maré, E., and Van Niekerk, A. J. (2021). Quantitative guidelines for retiring (more safely) in South Africa. *South African Actuarial Journal*, 21:75–91.

Van Zyl, N. and Van Zyl, D. J. (2016). The impact of behavioural economics and finance on retirement provision. *South African Actuarial Journal*, 16:91–125.

Venter, G. G. (2010). Advances in modeling of financial series. *Working Paper. Casualty Actuarial Society*.

Waring, M. B. and Siegel, L. B. (2015). The only spending rule article you will ever need. *Financial Analysts Journal*, 71(1):91–107.

World Health Organization (2020). Global health observatory data repository. https://apps.who.int/gho/data/view.main.SDG2016LEXv?lang=en.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320.