THE GARCH-EVT-COPULA MODEL AND SIMULATION IN SCENARIO-BASED ASSET ALLOCATION

by

Peter Gareth Fredric McEwan

Submitted in fulfilment of the requirements for the degree of

Master of Science

in the

Faculty of Science

at the

Nelson Mandela Metropolitan University

December 2015

Supervisor: Prof. G.D. Sharp

Declaration

I, Peter Gareth Fredric McEwan, hereby declare that this dissertation entitled “The GARCH-EVT-Copula Model and Simulation in Scenario-based Asset Allocation” for Master of Science is my own work and that it has not previously been submitted for assessment or completion of any postgraduate qualification to another University or for another qualification.

_____

Peter Gareth Fredric McEwan

Dedication

This dissertation is dedicated to my parents, Peter McEwan and Polly van Niekerk, my wife, Tashreeqa McEwan, my son, Mikail McEwan, and my brother, JC van Niekerk, for their unbounded support, continuous inspiration, invaluable understanding and, most importantly, unconditional love. Last, but not least, I dedicate this to my late brother, Tyrone McEwan, gone forever away from our loving eyes and who left a void never to be filled in our lives. Among the most special people to walk the earth, we love you and miss you beyond words.

Acknowledgements

I would like to express my sincerest gratitude to the many people who helped make this research and journey possible. In particular, I would like to thank:

- My amazing mom, Polly van Niekerk, for your unwavering, unshakeable and loving support all my life.
- My brilliant dad, Peter McEwan, for being my best friend and continuous inspiration.
- My beautiful wife, Tashreeqa McEwan, for your deeply-treasured love, limitless support and unfailing encouragement.
- My awesome son, Mikail McEwan, thank you simply for being in my life.
- My brother, JC van Niekerk, for your moral support, good character and uproarious humour over the years.
- My supervisor, Prof. Gary Sharp, for your insight and guidance throughout this study.
- My friend, Vsevolod Gorlach, for inspiring the vision of this research and for your guidance when things got tough.
- Jeanne Steven at INET BFA for your valuable assistance with the data.

This research was not entirely an isolated, domestic endeavour, so I would like to thank those who contributed from afar, in particular, Dr Alexios Ghalanos, Guillaume Pealat and Professor Eric Zivot.

For their financial support, I would like to thank the following institutions:

- The National Research Foundation of South Africa, grant number SFH13091035571.
- The Dormehl-Cunningham Trust.
- The Nelson Mandela Metropolitan University RCD Department.

Abstract

Financial market integration, in particular, portfolio allocations from advanced economies to South African markets, continues to strengthen volatility linkages and quicken volatility transmissions between participating markets. Largely as a result, South African portfolios are net recipients of returns and volatility shocks emanating from major world markets. In light of these, and other, sources of risk, this dissertation proposes a methodology to improve risk management systems in funds by building a contemporary asset allocation framework that offers practitioners an opportunity to explicitly model combinations of hypothesised global risks and the effects on their investments.

The framework models portfolio return variables and their key risk driver variables separately and then joins them to model their combined dependence structure. The separate modelling of univariate and multivariate (MV) components admits the benefit of capturing the data generating processes with improved accuracy.

Univariate variables were modelled using ARMA-GARCH-family structures paired with a variety of skewed and leptokurtic conditional distributions. Model residuals were fit using the Peaks-over-Threshold method from Extreme Value Theory for the tails and a non-parametric, kernel density for the interior, forming a completed semi-parametric distribution (SPD) for each variable. Asset and risk factor returns were then combined and their dependence structure jointly modelled with a MV Student *t* copula. Finally, the SPD margins and Student *t* copula were used to construct a MV meta *t* distribution.

Monte Carlo simulations were generated from the fitted MV meta *t* distribution on which an out-of-sample test was conducted. The 2014-to-2015 horizon served to proxy as an out-of-sample, forward-looking scenario for a set of key risk factors against which a hypothetical, diversified portfolio was optimised. Traditional mean-variance and contemporary mean-CVaR optimisation techniques were used and their results compared. As an addendum, performance over the in-sample 2008 financial crisis was reported.

Keywords: GARCH, Extreme Value Theory, Copula, Simulation, Conditional Value-at-Risk, Portfolio Optimisation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The secret of all victory lies in the organisation of the non-obvious.
Marcus Aurelius

1.1 Brief South African Macro Risk Landscape

An overarching characteristic of global financial markets is the trend towards unification of financial markets in advanced economies with those in emerging market economies. Factors such as increasing globalisation, financial and exchange rate liberalisation and financial innovation are key contributors strengthening the linkages between international markets (Lane & Milesi-Ferretti, 2008).

Emerging market (EM) economies have been, partly as a result, absorbing a much larger share of outward portfolio investment from advanced economies than was the case prior to the 2008 financial crisis. For example, the International Monetary Fund [IMF] (2014) observed that, between 2002 and 2012, equity portfolio allocations to EM economies from advanced economies increased from 7 percent of the total stock of advanced economy portfolio investment to almost 20 percent. Similarly for bond portfolio allocations, from 4 percent of the total stock of outward portfolio investment from advanced economy markets in 2002 to almost 10 percent in 2012. These portfolio allocations are principally directed towards only a few destination countries. Of the portfolio allocations to EM equities in 2012, 80 percent was invested in 12 of the 190 emerging market economies. Similarly for EM fixed income, with 75 percent directed to the same 12 economies. South Africa is among the 12 destination markets. Concomitantly, a significant degree of portfolio concentration to emerging market economies hails from only a handful of advanced economies (viz., the United States [US], United Kingdom [UK], Hong Kong SAR and Singapore). One consequence is an increasing synchronisation in asset price movements and volatilities between these dualistic market-types.

Existing globalised portfolio investment channels facilitate market liquidity that may rapidly be mobilised, enabling shocks from advanced economies to quickly propagate to emerging market economies. Amplifying the transmission of such shocks is an appreciable uptrend in “herding” behaviour observed among not only investor groups and across investment styles, but predominantly in the exceptionally large $25 trillion United States mutual fund industry (IMF, 2015). In the strict, investor-behaviour sense, herding refers to actions taken only because investors see other investors taking them. Relaxing this definition, the degree of “correlated trading” within an investor group serves as an adequate proxy measure for herding (Lakonishok, Shleifer & Vishny, 1992). The trend in rising herding levels, observed in the largest U.S. investment vehicles, is highest too for emerging market asset and high-yield asset groups. Combined with increasingly mobile cross-jurisdictional

market liquidity, these factors should add to the list of sources of contagion risk to emerging markets.

Investor decisions, particularly when they aggregate on the downside, can activate shock externalities that transmit through asset classes and across financial markets. Linking asset co-movements, correlation spillovers and market liquidity is an important step towards appreciating contemporary risk undercurrents. Such risk undercurrents are continuously evolving as a function of, notably, increasingly homogenous global markets, technological change, regulation and the shifting composition of market participants. It is well documented (e.g., Fenn, et al., 2011; IMF, 2015; J.P. Morgan, 2011) that cross-asset correlations and correlation levels among major asset classes have risen markedly in recent years, particularly since the 2008 financial crisis. There has also been a substantial rise in correlations between asset markets in advanced and emerging market economies. The major asset classes considered generally fall into the categories of international equities, government bonds, corporate bonds, exchange rates, hard commodities and soft commodities, with further differentiation along emerging market and developed market lines. At a more granular level, asymmetries emerge in exceedance correlation structures of asset returns and return volatilities (Baruník, Kočenda & Vácha, 2014; Christoffersen, Errunza, Jacobs & Langlois, 2012; Longin & Solnik, 2001; Mashal & Zeevi, 2002). Asymmetric co-movements in asset returns implies higher return and volatility correlations in market downturns than in upturns, where large negative returns are more correlated than large positive returns. When negative returns in an asset class or market are driven to extreme levels, the inverse relationship between contemporaneous returns and their conditional volatility cause volatility levels to spike. If such idiosyncratic price shocks are large enough, volatility spikes in one (typically important) market cause not only volatility, but also correlation, to “spill over” to other markets. In a heightened correlation world with advanced financial networks, “contagion” can quickly transmit across markets. This has been observed over the regularly occurring financial crises experienced in the last 30 years (Forbes, 2012).

South African (SA) financial markets are subsumed in the momentum of unification, becoming more synchronised both laterally with emerging markets and vertically with developed markets. For an emerging market, they are also relatively deep and informational-efficient (McKinsey & Company, 2013). In an investigation of the degree of South Africa’s global interdependence, Chinzara and Aziakpono (2009) analysed returns and volatility linkages between the SA equity market and six major world equity markets, those of Australia, China, Germany, Japan, UK and US. The authors found the SA equity market to be a statistically significant net recipient of returns and volatility shocks from the majority of the markets in the study. Importantly, exogenous volatility is quickly transmitted to the SA equity market which, in turn, could spill over to other domestic asset classes. Indeed, Duncan and Kabundi (2011) found significant interdependence in volatility across three of South Africa’s major asset classes: rand/dollar currency, yields on SA 10-year government bonds and returns to the Johannesburg Stock Exchange (JSE) all-share equities index. The SA equity market

is, in general, the primary source of domestic volatility transmission to the currency and bond markets.

1.2 On the Traditional Model and Model Risk

The global financial crisis of 2007-2008 was followed by significant losses in portfolios across the globe and, indeed, in portfolios of South African investors. It ushered in a protracted period of market volatility, with traditional risk management models failing because of increased correlation among all asset classes (Stefanova, 2015). The pendulum of investors' attention has since swung away from portfolio returns as the singular measure of success to a more holistic domain focusing on risk management practices.

Traditional portfolio risk management models are those developed under the auspices of Modern Portfolio Theory (MPT). Introduced by Markowitz (1952), MPT revolutionised asset allocation decision-making by using statistical methods to determine how to efficiently distribute wealth across a portfolio of risky assets. Investment allocation decisions are based on a mean-variance optimisation (MVO) problem with the objective of maximising portfolio returns for given levels of portfolio risk (equivalently, minimising portfolio risks for given levels of portfolio return), subject to certain constraints. The theory is consistent with the assumptions in a normal distribution probabilistic setting, where sample mean and variance estimators proxy the first two population moments of asset returns data (i.e., returns and risk) and a sample variance-covariance matrix captures return dynamics. The latter structure is equivalently a scaled correlation matrix deemed, in the MPT setting, an appropriate statistic for defining diversification (Fabozzi, Kolm, Pachamanova & Focardi, 2007). Diversification is the core concept of MPT. In formalising the mathematics behind combining risky assets in a portfolio setting, Markowitz quantified a path for portfolio managers to efficiently reduce exposure by diversifying away non-systematic risks. This "traditional" asset allocation framework is used extensively in present day institutional investment management. In what may suffice as an indicator for the industry, a 2013 risk study by Six and Wiedemann (2013) found that more than two-thirds of the 104 institutional investors surveyed in Germany use MVO when making investment decisions. Fabozzi, Focardi and Jonas (2007) surveyed 38 large and medium-sized U.S. and European equity investment managers managing a combined $4.3 trillion in assets. The authors found that 83% of the managers employed MVO. Other research affirms mean-variance optimisation as the dominant asset allocation framework among institutional investors (see, for example, Amenc, Goltz & Lioui, 2011; Dempster, Pflug & Mitra, 2009; Idzorek, 2006; IMF, 2011).

Diversification, as a means to reduce portfolio risks and extract risk premia, is an effective concept over long-term investment horizons. It provides investors with the best reward per unit risk through an efficient combination of individual assets. The demand for greater diversification is, in fact, one of

the drivers of globalisation, as investors increasingly trade their "home bias" for exposure outside their home markets (Boston Consulting Group, 2014; Institute of International Finance, 2014; Phillips, 2014). However, short-term, systemic crashes destroy diversification benefits as asset class correlations converge and shocks rapidly propagate through global financial networks. The traditional MPT framework does not protect portfolios against severe losses, or wealth destruction, over such periods. In "normal" markets, the distributional setting in MPT may serve investors well, but financial crises punctuate this state far more often than the normal distribution suggests (Xiong, 2010 and references therein). In "non-normal" markets, current risk management approaches underestimate downside risks, supporting the continued investigation of new approaches and applications of statistical methods.

Extreme market events represent incidences of "non-normality": environments in which asset classes exhibit so-called "stylised facts". Stylised characteristics of asset returns are well documented for both the univariate (Chakraborti, Toke, Patriarca & Abergel, 2011; Cont, 2001) and multivariate schema (e.g. Delatte & Lopez, 2013; McNeil, Frey & Embrechts, 2005; Nyström & Skoglund, 2002a). The following stylised facts are typical for univariate returns:

1. Returns are not independent and identically distributed (i.i.d.). They may also show low absolute values for a first-order autocorrelation coefficient.
2. Serial correlation in returns is not significant, whereas corresponding absolute or squared returns are mostly autocorrelated (i.e., they exhibit volatility persistence or "long memory").
3. Return volatilities exhibit conditional heteroskedasticity (i.e., time-varying volatility).
4. Return distributions are leptokurtic, reflecting fatter tails and "peakedness" around (mostly[1]) positive arithmetic means (i.e., higher probabilities for extreme events and greater central tendency, respectively, than what the normal distribution would generate).
5. Returns exhibit skewness, reflecting asymmetry in the tails. Distributions are generally negatively skewed, implying higher probabilities of negative returns.
6. Extreme returns are observed closely in time (i.e., volatility clustering); the clustering itself generates excess volatility, or fat tails.

Similarly for multivariate return series:

1. There is little evidence of cross-correlations between return series (i.e., insignificant similarity between one series and shifted/lagged copies of another, as a function of the lag), except for when returns are not shifted (i.e., strong evidence of contemporaneous cross-correlations between returns). In other words, there is time-dependent co-dependence between return

[1] It is likely an artefact, in this study, of the sampling frequency of the data that two variables exhibit negative arithmetic means.

series.

2. There is strong evidence of cross-correlation between absolute or squared returns.
3. Contemporaneous correlations vary over time.
4. Extreme returns in one series are often accompanied by extreme returns in several other series (i.e., evidence of non-linear correlation or spillover effects).

These empirical properties represent real limitations in traditional asset management, as well as imply real "model risks" in using MPT. On the other hand, the MPT models are theoretically and computationally tractable. Therefore, one of the aims of this paper is to incorporate these properties in a way that extends and supports the basic Markowitz portfolio theory.

1.3 Qualitative Aim of the Paper

Assets under management in South African pension funds, collective investment schemes (CIS) and hedge funds are in the region of $302 billion, $147 billion and $4.6 billion, respectively[2] (PricewaterhouseCoopers, 2015). As net recipients of global returns and volatility shocks originating exogenously and from a structurally changing and expanding pool of risk, it may materially benefit investors to be able to explicitly model combinations of these global risks and the effects on their portfolios. As well, the bulk of the variation in SA equity markets is driven by global, rather than local, risk factors (Polakow & Flint, 2014), with co-movement in world equity markets significantly higher during crisis than non-crisis periods (Duncan & Kabundi, 2014). These notions form a macroeconomic backdrop highly conducive to a flexible, bottom-up, macro-driven model to augment top-down South African portfolio management systems; a quantitative framework able to capture anticipated shifts in global risk-return expectations across an array of asset classes and market environments.

This paper proposes a technically advanced, yet tractable, quantitative decision support framework designed to work in lockstep with economists' and experts' opinions on core (statistically or feasibly meaningful) risk drivers to portfolio asset classes. Non-normal stylised facts of univariate and multivariate financial data are explicitly modelled in the framework. The resulting model is an empirically- and theoretically-consistent simulation engine customised to a portfolio of asset classes and corresponding set of risk drivers. The model is the centre of a Monte Carlo simulation framework capable of extrapolating the observed underlying multivariate and univariate data generating processes beyond past observations. Simulations are ultimately used for conditioning on any number of combinations of forward-looking views in the form of range expectations of key asset class drivers. Sets of scenarios are formed under which portfolio optimisations take place with the goal of enriching the insight and risk management practices of practitioners. The model admits fat-tailed events and

[2] As at December 2014 and exchange rate USD/ZAR 11.55. Also, 8.4% of pension assets were invested in CIS.

asymmetric expected returns to be explicitly factored into the asset allocation process. Optimisation is implemented as a function of maximised expected portfolio return per given level of portfolio conditional value-at-risk[3] (CVaR), as opposed to portfolio variance[4]. Each resulting efficient frontier of asset class weights is, hence, optimally diversified against downside risk as well as against any backdrop of forecast scenarios of a practitioner's choosing. This approach significantly improves on the traditional Markowitz MVO method, as is shown in an out-of-sample evaluation.

The framework lends itself well to evaluating stochastic spillover effects on portfolios. Not only can magnitudes of univariate return and/or volatility shocks be adjusted, but they may be done so under a sliding scale of multivariate co-dependence. Forward-looking scenarios may be set and portfolios evaluated under varying degrees of multivariate co-dependence. For each stress-period scenario considered, portfolio diversification is maximised along an efficient frontier. However, while merely using sophisticated distributional models and downside risk measures may lead to more effective diversification, it may not lead to substantially smaller losses in extreme stress-period scenarios. By taking a step further and analysing scenario-conditioned portfolios along the frontier and corresponding expected profits and losses in portfolio assets, areas and degrees of vulnerability in portfolios can be identified in advance of extreme stress periods. Clear-sighted hedging instruments or strategies may be determined *ex ante*, as well as proactive response plans mandated. In addition, common portfolio risk measures, such as value-at-risk (VaR) and CVaR, may be generated for analysis. By becoming a key technology in risk management practices, the conditional Monte Carlo framework may serve to add material value through actionable insight.

The approach in this paper builds on research by Inanoglu and Ulman (2009), Mashal and Zeevi (2002), Nyström and Skoglund (2002a), Wang, Sullivan and Ge (2012) and Xiong and Idzorek (2011). All data in this paper were prepared using Microsoft Excel and all results produced using the statistical package R (R Development Core Team, 2014), with various R packages used in subsets of analysis. The computer used to implement codes is a Lenovo laptop with Intel Core i5-4200M processor with a clock speed of 2.50 GHz and 8 GB memory, running on a 64-bit Windows 7 Operating System.

The remainder of this paper is organised as follows: Chapter 2 provides a survey of the literature and forms the basis for the methodology chapter; Chapter 3 describes the theories and methodologies used in building the framework and presents results; Chapter 4 concludes and gives some suggestions for further research.

[3] Equivalently, expected shortfall (ES), expected tail loss (ETL), tail-VaR, average VaR (AVaR), mean excess loss and mean shortfall in the literature.

[4] This paper assumes an asymmetric risk preference for risk-averse investors and uses CVaR as the preferred measure of downside risk. Optimisation may be modified to optimise using other downside risk measures (e.g., Omega, Kappa measure or Sortino ratio).

Chapter 2

Literature Review

Much of the real world is controlled as much by the "tails" of distributions as by means or averages: by the exceptional, not the mean; by the catastrophe, not the steady drip; by the very rich, not the "middle class". We need to free ourselves from "average" thinking.
Philip W. Anderson, Nobel-prize-winning physicist.

This chapter provides a literature review of the topics relevant to the dissertation. It is divided into four sub-sections governing the main theoretical concepts used: Time Series Analysis (TSA), Extreme Value Theory (EVT), Dependence and Portfolio Optimisation.

2.1 Time Series Analysis

The subject of correct distributional form for univariate financial data is a dynamic and deeply researched area in finance. Louis Bachelier, through his 1900 doctoral dissertation, is credited as the first to assign a probability distribution to financial process data, that of "*la loi de Gauss*", or "the law of Gauss". (Bachelier, 1900; Lévy, 1940, p. 487). That the normal distribution describes stochastic (equity) return processes was further supported in Osborne (1959).

However, empirical research deepened and, as Kim and Kon (1994) showed, "sampling independent observations from an identical Gaussian distribution" has proved less and less accurate in describing financial returns. Mandelbrot (1963) found that the many outliers in the tails of empirical commodity return distributions tended to pull the central peak of the Gaussian distribution "much lower and flatter" than what the data suggested (i.e., the data are leptokurtic: too many observations near the mean and too many in the extreme tails than what the normal distribution generates). The author further identified differing slopes between the two tails in graphs of returns data (i.e., skewness in returns) and that "large [price] changes tend to be followed by large [price] changes - of either sign - and small [price] changes tend to be followed by small [price] changes" (i.e., volatility clustering in returns). Such alternating periods of high and low volatility contradict the idea that sampled observations are i.i.d.

These (and other) univariate stylised facts have since been found to be pervasive across all major asset classes: leptokurtosis in equities (Fama, 1965), fixed income (Amin & Kat, 2003), foreign exchange (Huisman, Koedijk, Kool & Palm, 2002) and listed real estate (Lizieri, Satchell & Qi, 2007); skewness in equities (Fama, 1965), fixed income (Alles & Kling, 1994), foreign exchange (Brunnermeier, Nagel & Pedersen, 2009) and listed real estate (Hutson & Stevenson, 2008); volatility

clustering in equities (Fama, 1965), fixed income (Cappiello, 2000), foreign exchange (Baillie & Bollerslev, 1989) and listed real estate (Cotter & Stevenson, 2007). Cont (2001) provides a detailed survey of univariate financial return stylised facts (see also Pagan, 1996).

Since the rejection of the normal distribution as adequate in describing return characteristics, many distributional forms have been investigated. These forms can be described by either unconditional models (i.e., time-independent models, where the assumption is that return distributions and distribution moments are constant through time) or conditional (i.e., time-dependent) models.

Before conditioning on time became widely adopted, research effort was aligned with extending the unconditional modelling framework. The unconditional approach considers financial return data as random draws from a static, time-invariant distribution. Returns are assumed to be i.i.d., implying distribution moments evolve unchanged through time, with each observation representing a draw from an identical distribution. The focus of this research typically centred on adequately representing empirically fatter-than-normal tails.

Mandelbrot (1963) modelled the excess kurtosis of commodity returns under this approach using the stable-Lévy, or stable Paretian, family of probability laws, which nests the Gaussian distribution as a limiting case. Fama (1963, 1965) found support for the stable laws applied to equity returns. A number of drawbacks of this class of distributions exist, however, such as non-existent second and higher moments for characteristic exponent less than two and varying fitted characteristic exponents (or tail indices) under temporal aggregation (Akgiray & Booth, 1988). As a consequence, many of the assumptions in classical financial theory (e.g., in portfolio theory and the capital asset pricing model) could not hold as they did for the Gaussian model, leading other researchers to propose simpler, finite-variance alternative distributions.

A number of competing distributions in the unconditional framework have been considered. Box and Tiao (1962, 1973) introduced the exponential power distribution, more commonly known as the generalised error distribution (GED) as per Nelson (1991). Blattberg and Gonedes (1974) proposed the Student *t* distribution; Mittnik and Rachev (1993), the Laplace and Weibull distributions. Unconditional mixture distributions have been proposed (Akgiray & Booth, 1988; Ball & Torous, 1983; Kon, 1984; Peiró, 1994; Tucker, 1992). Temporal dependencies in return series, however, violate too many constraints implicit in unconditional distributions. This directed the trend in financial research towards investigating capacities of time-dependent, conditional approaches (Bollerslev, Chou & Kroner, 1992; Kim & Kon, 1994; Tucker, 1992; Yu, 2001).

Time series analysis (TSA) techniques are built around identifying underlying structures that manifest in the data through the changing characteristics of financial variables. The theory develops through the point of view of returns as ordered sequences of values of a variable at equally spaced time

intervals (i.e., a time series). The concept of time-ordering is the premise underpinning conditional distribution models. Parameter dynamics in such models are modelled on information of past market movements or, at times, a more general or exogenous information set.

In univariate financial time series, a dynamic conditional mean model may be specified to filter out serial correlation in the series. Logarithmic returns (log returns) are predominantly used as the time series as they have attractive properties, such as stationarity and ergodicity (Cambell, Lo & MacKinlay, 1997), independence of the unit of time and stability under time-aggregation (Longin, 1999). Financial time series returns, however, at lower frequency levels (such as weekly or monthly), rarely exhibit non-spurious serial dependence. The conditional means in the ARMA model may, in such instances and where significant, be specified as a constant (i.e., set to the unconditional mean or simply to zero). The benchmark conditional mean specification in TSA is that of a low order autoregressive moving-average (ARMA) process. The specification is favoured due to its ability to produce acceptably accurate short term forecasts of time series (Nyström & Skoglund, 2002b). In its standard form, the ARMA framework implies a homoskedastic distribution reflective of constant conditional volatility. A natural extension is to assume a weakly stationary white noise innovation process for the ARMA model (Cryer & Chan, 2008; Tsay, 2012) and to model the innovation process with a varying, or dynamic, conditional volatility structure. This extension shifts the conditional mean equation out of the (largely unsupported) domain of conditional homoskedasticity and into the domain of ubiquitously observed conditional heteroskedasticity (Dias & Embrechts, 2003; Nyström & Skoglund, 2002a). The innovation process may, thus represented, be modelled by one of the many generalised autoregressive conditionally heteroskedastic (GARCH) family of models to capture, most notably, the stylised fact of volatility clustering in the underlying data. The combination of ARMA model for the conditional mean process and GARCH model for the conditional variance of the ARMA innovation process is referred to as a composite ARMA-GARCH model.

Engle (1982) introduced the autoregressive conditional heteroskedastic (ARCH) model and Bollerslev (1986) the GARCH generalisation. The authors showed the effectiveness of using the respective specifications in modelling the conditional volatility of financial time series. Since these seminal publications, a great deal of related research and application has taken place. Engle and Bollerslev (1986) defined a “persistent variance” integrated GARCH (IGARCH) model, where conditional volatility is considered integrated of order one and shocks to the system permanent. Caution has been advised to first investigate the possibility of omitted structural breaks prior to using the IGARCH model (Diebold, 1986). Hypothesising that conditional volatility may be related to risk premia on assets, Engle, Lilien and Robins (1987) presented a more sophisticated specification for the conditional mean: the ARCH-in-mean model (GARCH-M), where conditional variance is estimated as a regressor in the conditional mean equation. To the extent that observed asymmetric responses of volatility to positive and negative return shocks are significant, Nelson (1991) and

Glosten, Jagannathan and Runkle (1993) defined, respectively, the exponential GARCH (EGARCH) and GJR-GARCH models. These models admit leverage terms for the modelling of asymmetric volatility clustering. Ding, Granger and Engle (1993) proposed the asymmetric power ARCH (A-PARCH) specification as an alternative ARCH-based model to capture asymmetric volatility clustering. Instead of assuming conditional volatility in square form, the authors suggested a varying power term in the heteroskedastic equation, to be estimated directly from the data. The A-PARCH specification nests a number of ARCH-based sub-models for different power and leverage parameter settings. Zakoïan (1994) introduced the threshold GARCH (TGARCH) model, the first-order specification of which is a variant of the GJR-GARCH model. The model introduces a threshold effect into the innovation volatility process to capture GARCH features within a regime-switching framework.

On the whole, the abovementioned models are among the more popular GARCH-family specifications for conditional volatility. “Essentially”, as Focardi and Fabozzi (2004, p. 379) point out, “one wants to understand how the decisions of a large number of economic agents do not average out, but produce cascading and amplification phenomena”. Identifying and capturing this effect as quickly as possible is an overarching theme in conditional volatility modelling. A comprehensive review of different GARCH specifications may be found in Bollerslev, Chou and Kroner (1992) and, for more recent work, Carmona (2014), Francq and Zakoïan (2010) and Tsay (2010). Of interest to note is the Hansen and Lunde (2005) study conducted as an out-of-sample test of the superior predictive ability of over 330 different conditional volatility models. The authors showed that the best models do not provide a significantly better forecast than the parsimonious GARCH(1,1) model for daily exchange rate data (DM-$). However, for daily equity data (IBM shares), the symmetric GARCH(1,1) model is inferior to models that account for asymmetric volatility clustering. Conditional densities used were limited to the Gaussian and Student *t* specification.

In the GARCH model, the shape of the conditional distribution of the innovations is the same shape as the conditional distribution of future returns. This has implications for under- or overestimating future risks. The standard form of the model assumes a conditional Gaussian distribution. While the ARMA-GARCH-family structure may capture dependency in the conditional time variation of the distributional parameters of the mean and variance, employing a conditional Gaussian distribution does not account for leptokurtosis or skewness (Bai, Russell & Tiao, 2003); conditional heavy tails typically remain after filtering returns for volatility clustering. The Gaussian law places an exponential speed at which the downside tail of the distribution decays. This is empirically too fast a decay, prompting the consideration of non-normal conditional densities to improve capturing of stylised facts. In terms of optimal asset allocations, Xiong and Idzorek (2011) demonstrated the adverse effects of ignoring skewness and kurtosis.

Candidate (standardised) conditional distributions considered in this dissertation include the following: Gaussian, Student *t* (Bollerslev, 1987; So & Yu, 2006), GED (Nelson, 1991) and their associated skewed variants (Fernández & Steel, 1998; Ferreira & Steel, 2006), as well as the Generalised Hyperbolic [GHYP] (Prause, 1999), Normal Inverse Gaussian (NIG), nested in the GHYP distribution, Generalised Hyperbolic Skew Student *t* [GHST] (Aas & Haff, 2006) and Johnson's reparameterised SU [JSU] (Rigby & Stasinopoulos, 2005, 2010). Ghalanos (2014a) provides a concise overview of the candidate distributions. In terms of the Fernández and Steel (1998) skewing mechanism used to distort distributions, the reader is referred to Li, Wang and Tian (2013) for the standardised skewed normal, Bao, Lee and Saltoğlu (2007) for the standardised skewed Student *t* and Bao, Lee and Saltoğlu (2004) for the standardised skewed GED distributions, as well as to Palmitesta and Provasi (2006) for the latter two distributions.

Despite the useful features of the ARMA-GARCH-family model, the structure may still be challenged in capturing extremal behaviour, in the asymptotic sense, of asset returns (Davis & Mikosch, 2009; Mikosch & Stărică, 2000). The added flexibility of controlling the thickness and skewness in conditional returns via the innovation distribution "allows for some off-load of the impact of outliers from the volatility estimates" (O'Brian & Szerszen, 2014). However, regardless of choice of distribution, GARCH-family innovation processes still tend to exhibit uncaptured density in the tail extrema. The ARMA model produces model residuals, which are modelled with a GARCH structure, producing filtered residuals. The conditional density choices for the filtered residuals, even with fat-tail and skewness optionalities, are generally unable to allocate sufficient density deep in the tails, or outlier areas, of the typical financial variable. There is a forced compromise in the fit of a single set of parameters to competing conforming and non-conforming distributional behaviour. The bulk of the innovations may be fit well, but at the compromise of a sub-standard fit to the scarcely populated tails.

To this end, TSA serves two purposes: it provides techniques capable of capturing a large degree of stylised facts in return data and, in so doing, pre-whitens the data for EVT analysis. The GARCH-filtered innovation series may be standardised by its conditional volatility series. The resulting standardised series is assumed to follow an uncorrelated strict white noise process, approximately i.i.d., with zero mean and unit variance. Byström (2004, 2005), Diebold, Schuermann and Stroughair (2000), Ghorbel and Trabelsi (2009), McNeil and Frey (2000), Nyström and Skoglund (2002b) and Rocco (2010), among others, advocate this method of pre-whitening in order to produce (approximately) i.i.d. data, as EVT analysis requires data to be i.i.d. Jalal and Rockinger (2008) showed that even for cases when the ARMA-GARCH-family model is misspecified, applying EVT to the resulting filtered standardised residuals remains a robust technique that "delivers good results".

2.2 Extreme Value Theory

The weight of individual tails prevails
and drives the aggregate process.
Focardi and Fabozzi (2004, p. 380)

A critical element in modelling samples of univariate financial returns is proper management of the set of extreme observations populating, essentially, all such data. The EVT branch of statistics provides a formal framework for studying the statistical behaviour expected in the tail(s) of a distribution. In the sense that the central limit theorem (CLT) is a limit law for the mean, EVT is a limit law for extremes (Ergen, 2010). The CLT states that the Gaussian distribution, regardless of the finite-variance underlying distribution (or stable Paretian distribution, regardless of the infinite-variance underlying distribution), is the limiting distribution for sums and averages of i.i.d. random variables, provided the sample size is sufficiently large. Similarly, but with regards to extremes, EVT deals with the convergence of sample maxima[5]. Specifically, EVT deals with the distribution of the smallest and largest order statistics (Kuester, Mittnik & Paolella, 2006). To effectively do so, the theory differs from classical statistics in that it is developed without imposing a distribution (or mixtures thereof) over the entire sample; rather, focus is on the density in each tail region, allowing the density in each region to be modelled as distinct from the whole sample. This difference implicitly admits asymmetry in tail modelling, creating an opportunity to improve accuracy in capturing the probabilistic structure behind high-risk market moves. A further distinguishing feature of EVT is the calibration of a parametric form to discrete, empirical tail aspects of the random variable. The parametric nature of EVT enables users to extrapolate to "parts of the distribution that have yet to be observed in the empirical data" (LeBaron & Samanta, 2005). This has clear advantages for simulating tail events beyond the range of historical data.

Fisher and Tippett (1928) derived the three foundational limit laws describing the type of distributional form a set of suitably normalised maxima may belong[6]. The form of the asymptotic distribution is independent of the process generating the maxima; only the distribution's parameter values depend on the process (Longin, 1996). The three standard limiting distributions can therefore be mapped onto different domains of attraction, belonging to the type of either Fréchet, Gumbel or Weibull. Depending on the speed of tail decay of the density function (and in order of left-tail favour[7]),

1. slower, fat-tail, power-law decay, such as that described by (unbounded) stable Paretian, Cauchy, Burr, loggamma and Student *t* distributions, belong to the Fréchet domain[8];

[5] Noting that $\max(X_1, X_2, \dots, X_n) = \min(-X_1, -X_2, \dots, -X_n)$, results for the distribution of maxima hold similarly for the distribution of minima and vice versa. For convenience, results of maxima are discussed.
[6] Distributions of the same "type" are obtained from one another through normalisation by using appropriate location and scale transformations (McNeil, Frey & Embrechts, 2005).
[7] With respect to financial modelling of downside risks (McNeil, et al., 2005).
[8] The ARCH class of distributions also belongs to the domain of attraction of the Fréchet law (Jansen & de Vries, 1991).

2. quicker, exponential-decaying distributions (bounded or unbounded, but with all finite moments), such as the Gaussian, exponential, gamma and lognormal, belong to the Gumbel domain; and
3. bounded tail decay distributions, such as the uniform and beta distributions, having finite endpoints, belong to the Weibull domain of attraction.

All three limit laws are, however, subsumed in the unifying, parametric generalised extreme value (GEV) distribution (Jenkinson, 1955; McFadden, 1978).

There are two common parametric approaches to EVT that differ on how maxima and their corresponding limit models of behaviour are defined: the initial Block Maxima (BM) approach and the more contemporary Peaks-over-Threshold (POT) approach. The BM approach models fluctuations of normalised maxima in the framework of the GEV distribution. A set of maxima is constructed by dividing the sample into non-overlapping blocks and selecting the maximum observation from each block. Sets constructed in this way may be expected to approximate an i.i.d. series more closely than sets created using the POT approach, but at a cost of losing (often limited) data points and increasing parameter estimate uncertainty. Since only one data point within each block is used, large observations that are smaller than the local maxima (but which may be significant for risk management purposes) are discarded from analysis. At the same time, data points selected as block maxima over low volatility periods are regarded as extrema and included in the analysis. Consequently, unacceptable levels of estimation bias enter into the BM framework. The method is therefore considered data inefficient. The POT method, on the other hand, is less wasteful of data in that the focus is on modelling excess losses above high thresholds (i.e., in the tails of the distribution). Sets of extrema are populated with observations exceeding some high predetermined threshold, enabling extreme data other than just the block maxima to be modelled. No clear method of threshold selection has, however, been advised in the literature. The user is forced to assume, arbitrarily, that the tail of the parent distribution starts at the threshold value. The reader is referred to Daníelsson and de Vries (1997), Dupuis (1999) and McNeil and Saladin (1997) for treatments on threshold selection, as well as to Chapter 3.4.3 of this dissertation. Nyström and Skoglund (2002b) conducted Monte Carlo experiments for different parent distributions and sample sizes and showed that the maximum likelihood (ML) estimator was virtually invariant to the choice of threshold for threshold values in the range of 5-13% of the data. Arbitrarily, this paper sets the threshold at 10% of each side of the (ordered) data as per guidance in Chavez-Demoulin and Embrechts (2004), Chavez-Demoulin, Embrechts and Sardy (2014) and Nyström and Skoglund (2002b). With reference to the BM method, the reader is referred to the texts of Embrechts, Klüppelberg and Mikosch (2001) and Jondeau, Poon and Rockinger (2007).

While both approaches have advantages and disadvantages, it is primarily the POT approach's

efficient use of scarce tail data that makes it the method of choice in modelling financial time series extrema (Chavez-Demoulin, et al., 2014; Davison & Smith, 1990; Embrechts, et al., 2001; McNeil, 1999; Nyström & Skoglund, 2002a, 2002b). This dissertation employs the POT method.

The mathematical proofs extending the classic Fisher and Tippett (1928) limit laws and laying the groundwork supporting the POT approach may be found in Balkema and de Haan (1974) and Pickands (1975). The latter author proved the generalised Pareto distribution (GPD) to be a natural, albeit approximate, limit law for modelling excesses over asymptotically high thresholds, conditional on the excesses being above the threshold. The threshold exceedance and associated limit law approach introduced an alternative and statistically sound method for examining extrema in the EVT literature. The concept is similar to how convergence to the non-degenerate Gaussian distribution is proved in the CLT. The "Pickands-Balkema-de Haan" theorem proves that essentially all commonly encountered continuous distributions are in the maximum domain of attraction of the GPD, but only in terms of modelling appropriately normalised excesses from these distributions above a high threshold and as this threshold tends to the right endpoint, or limit, of the respective distribution. That is to say, appropriately normalised sample maxima from the parent distribution converge in distribution and in the limit to the non-degenerate GPD (McNeil & Saladin, 1997). Figure 2.1 depicts a graphical description behind the POT theory, where a hypothetical generalised Pareto distribution is fit to excesses above a high threshold.

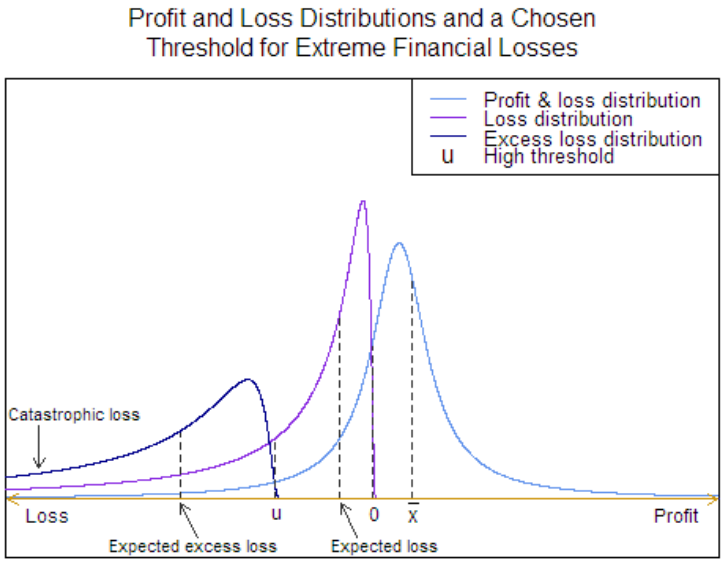


Figure 2.1: Graphical Concept behind Extreme Value Distributions

Hosking and Wallace (1987) investigated methods of estimating the GPD on simulated data samples of size 15 to 500. The authors found preference for using the sample-size based method of probability-weighted moments (PWM), compared to the method of maximum likelihood and method of moments. However, the preference towards the better small sample properties of PWM decreased as sample size increased and, as a result, mean square errors among the competing methods

converged. On the other hand, Smith (1987) chose maximum likelihood estimation (MLE), exclusively, to examine and compare asymptotic properties of various GPD estimators. MLE is supported in Longin (1996) in that the method yielded efficient estimates that were "unbiased, asymptotically normal and of minimum variance". Grimshaw (1993) showed MLE estimates of GPD parameters to be, in most cases, asymptotically normal and asymptotically efficient. Rootzén and Tajvidi (1997) showed that, in the domain of attraction of the heavy-tailed Fréchet distribution, PWM estimates became significantly biased and severely and systematically underestimated the quantiles of the distributions of maxima. The MLE estimates remained significantly less biased, although they exhibited a higher degree of variability. For reasons mentioned above and for the fact that it is the more commonly used method (McNeil, et al., 2005), this dissertation uses the MLE method for two-tailed GPD estimation.

Longin (1996) applied the BM method using the GEV distribution framework to model raw, un-declustered, unfiltered daily, weekly and monthly log return percentages of the most traded stocks on the New York Stock Exchange (NYSE). The data set spanned a period of over a century. The author reported a number of notable findings. Among them are that (1) the asymptotic distribution of extreme returns obeyed a Fréchet distribution, (2) as the sample period increased so too did the distribution of extremes shift to the right for maxima and to the left for minima, while the shape of the distribution remained the same, (3) the shape of the tails remained stable over the entire period, including the Great Depression era, and (4) the limiting distribution was stable under temporal aggregation (i.e., extrema sets selected from daily, weekly and monthly returns consistently obeyed the Fréchet law).

It has been shown that the restrictive i.i.d. assumption of EVT may be relaxed without a consequent change in the resulting limit theory. Conditions do apply, but the assumption holds for ARMA-GARCH filtered innovation series (Embrechts, et al., 1997, Chapter 4.4). McNeil and Frey (2000) applied the POT method to the returns series of the S&P 500 index, the DAX index, the BMW share price, the U.S. dollar British pound exchange rate and gold. The authors approximated the GPD to tails of standardised residuals, pre-whitened through an AR(1)-GARCH(1,1) filter. The two-stage, conditional EVT approach (or GARCH-EVT approach), taking into account the current volatility background, improved VaR and CVaR estimates compared to corresponding estimates from unconditional EVT and the classical GARCH model with normal and Student *t* innovations. The two-stage approach added risk measurement gains by enabling periods of high and low volatility to be more accurately reflected through time-varying tail quantiles. Fernandez (2005) applied the same two-stage methodology to thirteen equity indices spanning the United States, Europe, Asia and Latin America (i.e., developed and emerging market equities). The author's conclusions were consistent with those of McNeil and Frey (2000). In a related study, Byström (2004) applied the POT methods used in McNeil and Frey (2000) to Sweden's AFF and the United States' DOW (both equity) indices.

The author extended the conditional EVT approach to include the BM method and found similar performance across both methods, as well as improved risk measure estimates consistent with the findings of McNeil and Frey (2000). Krehbiel and Adkins (2005) showed similar results for the extended methods of Byström (2004) applied to daily spot and futures contract returns of the commodities traded on the NYMEX (i.e., West Texas Intermediate (WTI) crude oil, Brent crude oil, natural gas, heating oil and unleaded gasoline). Marimoutou, Raggad and Trabelsi (2009) evaluated daily WTI and Brent crude oil returns and found support for the modelling gains resulting from the conditional EVT approach. Furió and Climent (2013) modelled daily log returns from three equity indices (the S&P 500, UK's FTSE 100 and Japan's Nikkei 225) using the same conditional and unconditional EVT methods in McNeil and Frey (2000). To simultaneously address the question of whether conditional EVT quantile estimates differ under different GARCH specifications[9], the authors expanded the scope of GARCH models to include the asymmetric EGARCH and TGARCH specifications. The authors found that the three GARCH specifications produced robust results across all indices and that the conditional EVT framework was superior to the unconditional framework[10].

In order to complete the univariate density model, the two individual parametric GPD tails need to be spliced into the remaining high-density central portion of the distribution. A resulting "semi-parametric", approximate piecewise, constant density function may be created using linear interpolation to splice together the GPD tails into a smoothed gradient interior. Nyström and Skoglund (2002a) recommended standard non-parametric methods to model the interior of the distribution, thereby avoiding assumptions on the nature of the empirical distribution. The authors chose the empirical cumulative distribution function (CDF). This approach is supported in Carmona (2014) and Zivot and Wang (2006). MacDonald, Scarrott, Lee, Darlow, Reale and Russell (2011) and Wang, et al. (2012) chose a non-parametric smoothing kernel method based on a mean zero Gaussian probability density function (PDF) for the interior distribution. Other kernel functions are available (e.g., Epanechnikov, triangular, cosine or box). MacDonald, et al. (2011) noted, however, that the choice of kernel function is not critical in the semi-parametric setup, provided the selected kernel is not unreasonable. This is motivated by how the tail behaviour associated with the kernel will be diminished by the mechanism through which the kernel distributes, or averages, probability mass in the less-populated neighbourhoods of the tails (see also Alexander, 2008, p. 165). This dissertation implements a GPD-Gaussian kernel-GPD conditional EVT approach (i.e., a semi-parametric "tail-interior portion-tail" distribution).

The completed semi-parametric probabilistic model is a key component in the portfolio risk

[9] Chavez-Demoulin, Davison and McNeil (2005) claimed that one drawback of the conditional EVT methodology was that it, being a two-stage procedure, would result in EVT analysis sensitive to the fitting of the GARCH model to the dataset in the first stage.
[10] See also Jalal and Rockinger (2008).

management framework proposed in this study. It allows the granular modelling of variations of each marginal[11] asset, composing the set of positions held in a portfolio, and of the risk factors to the portfolio. The univariate step segues into the multivariate step: fitting the category of copula models well-suited to capturing non-linear multivariate co-variation in asset class returns, notably during periods of high market volatility. The completed semi-parametric distribution (SPD) represents a CDF, which is used to map from the x-domain (i.e., standardised innovations) to the y-range (i.e., the uniform $[0,1]$ range). The latter, transformed series, termed "pseudo-observations", forms inputs to the copula-fitting step.

According to Malevergne and Sornette (2006, p. 273), "concerning the cross-dependence between assets … copulas are the most fundamental concept and tool. They should … constitute a cornerstone of modern risk management practices". For the multivariate component, the second step captures and represents the collective dependence in the portfolio with a copula model.

2.3 Dependence

Capturing the [data generating] mechanism well enough to be able to explain events that did not occur in the given sample, but do occur later on or in other related populations is the essence of statistical modelling.
Peter C.B. Phillips

This chapter explains the approach to modelling multivariate dependence between asset returns in the portfolio and risk factors driving the returns.

There are several methods used in quantitative analysis to measure dependence. Arguably the most widely used measure of dependence in risk management is Pearson's linear correlation coefficient[12]. It is a quintessential part of MPT, used in determining covariance structure inputs to the Markowitz mean-variance framework. The estimator measures the statistical, linear dependence between two jointly normally distributed variables (Embrechts, McNeil & Straumann, 1999). It is assumed to fully describe the relationships between financial returns and so directly influences diversification benefits. However, the stylised facts of univariate returns practically ensure that respective distributions are non-normal, giving way to non-linear co-variation between them. Malevergne and Sornette (2006) and Mashal and Zeevi (2002) discuss some of the inadequacies of the estimator as a descriptor of multivariate dependencies, particularly its inability to characterise extreme co-

[11] Instead of "univariate", the term "marginal" is used here to indicate the transition to multivariate terminology. The individual univariate distributions in a multivariate distribution are known as marginal distributions.
[12] This is also, arguably, a function of the ubiquitous use of Markowitz-based portfolio theory. See, for example, Kat (2003) and Rodgers and Nicewander (1988).

movements in the underlying variables. Chicheportiche and Bouchaud (2012) noted that for any case other than the multivariate Gaussian, the linear correlation matrix is unable to describe non-linear dependence. Embrechts, et al. (1999) listed a number of shortcomings to using the Pearson's correlation coefficient. Among them are:

1. The scalar nature of the measure implies it can only partially capture the dependence relationship between risks. Simply as a measure of the degree of dependence, it is unable to characterise the structure of dependence. It has been widely observed that market crashes or financial crises often occur in different countries or asset classes at about the same time period, even when the correlation among those markets is fairly low; hence, only a partial capture of dependence relationships is reflected.
2. The measure assumes all correlation values in the range $[-1,1]$ are attainable. This range, however, depends on the marginal distributions of the risks and may not, in fact, be attainable.
3. Perfectly positively (negatively) dependent risks do not necessarily have a correlation of 1 (-1).
4. A correlation measure of zero (i.e., uncorrelatedness) does not imply independence between risks.
5. The measure is not invariant under non-linear strictly increasing transformations of the risks[13] (e.g., the correlation between $\log X$ and $\log Y$ generally does not equal the correlation between X and Y).
6. Correlation is only defined for finite-variance risks. It is undefined for infinite-variance risks and performs poorly for very heavy-tailed risks [14]. In other words, the measure cannot sensibly model dependence in jointly non-elliptically[15] distributed risks, nor even in jointly elliptically distributed risks with heavy-tailed marginal distributions (Embrechts, Lindskog & McNeil, 2003).
7. The measure is sensitive to extreme values (Kowalski, 1972).

The above shortcomings imply that traditional MVO portfolios are not efficient with respect to their effective risk profiles (Boubaker & Sghaier, 2013) and diversification benefits tend to be higher than expected.

To alleviate some of the shortcomings (those of 2, 3, 5, 6 and 7), Embrechts, et al. (1999), McNeil,

[13] One implication is that while the returns may be uncorrelated, the prices are correlated, or vice versa (Sun, Rachev, Fabozzi & Kalev, 2009).
[14] Very heavy-tailed risks refer here to power-law distributed random variables with tail index in the interval $(2,4]$ (Malevergne & Sornette, 2006, p.148).
[15] With reference to elliptical distributions (e.g. Gaussian and Student *t*) and non-elliptical distributions (shapes exhibiting multimodality, extreme skewness and/or heavy tails) see Jondeau, et al. (2007, Section 6.2), McNeil, et al. (2005, pp. 89-102) and Miller (2014, pp. 93-103).

et al. (2005) and Nelsen (2006), among others, suggest alternative dependence estimators, derived from copulas, known as rank correlation measures. These measures, the popular ones being Kendall's tau τ (Kendall, 1938) and Spearman's rho ρ_S (Spearman, 1904), measure the intensity of monotonic (as opposed to linear) dependence between two random variables on a quantile scale (Schweizer & Wolff, 1981; Wolff, 1980 and references therein). The assumption of a monotonic relationship is less restrictive than a linear relationship. The middle image in Figure 2.2 illustrates the point well: a non-linear dependence relationship exists, but the relationship is monotonic and is suitable for analysis by a rank-order estimator, but not by a linear correlation estimator. This dependence concept provides a valid alternative to linear correlation as a measure for non-elliptical distributions, for which linear correlation is an inappropriate measure of dependence and is often misleading. In contrast to the shortcomings listed above, rank-based estimators do not depend on marginal distributions[16], can attain values in the range $[-1,1]$ for joint distributions, are invariant under monotonic marginal transformations[17], are defined for margins with infinite variance (i.e., robust to heavy-tailed data) and, finally, represent outliers well, in that they ignore exact numerical values of variable attributes and consider only the ordering (ranking) of the values. In terms of selecting a rank estimator, Giplin (1993) and Newson (2002) argued the superiority of Kendall's tau over Spearman's rho. In terms of expedient calculation for elliptically contoured distributions, Fang, Fang and Kotz (2002) selected Kendall's tau over Spearman's rho and further noted its invariance property for extension to the class of meta-elliptical distributions[18]. Genest, Nešlehová and Ghorbal (2011) noted the preference for rank-based procedures in dependence calculations in that they to guard against misspecification of marginal distributions[19]. Differences between the two estimators are discussed in Fredericks and Nelsen (2007) and Nelsen (2006). Figure 2.3 depicts the generic relationships between Kendall's tau, Spearman's rho and Pearson's linear correlation coefficient. The relative magnitudes of dependence are shown for Spearman's ρ_S and Kendall's τ in relation to Pearson's linear correlation coefficient ρ of the Gaussian copula C_ρ^{Ga}. The relationships hold for other normal variance mixture distributions with correlation parameter ρ, such as the Student *t* copula $C_{\nu,\rho}^t$ (McNeil, et al., 2005). This dissertation uses the tau measure.

The rank-based estimators, however, still yield a pairwise scalar measurement and, also, measures of zero do not imply independence between risks (i.e., shortcomings 1 and 4 remain). Shortcoming 1 is mitigated by τ and ρ_S serving as inputs to the larger copula dependence structure. A further point worth noting is that the measures, being scale invariant, do not directly tie into the mean-variance

[16] The choice of marginal is irrelevant, as the rank-order estimators depend only on the copula (Ruppert, 2011, p. 184).
[17] A monotonic transformation is a means of transforming one set of numbers into another set of numbers in a way that preserves the order of the numbers. For example, inputting an arbitrary random variable X through its own CDF will map to a transformed variable, called the *grade* of X, uniform on the unit interval $[0,1]$. See Figure 2.5.
[18] A multivariate distribution is said to be "meta-elliptical" if its dependence structure is governed by an elliptical copula and where the marginal distributions are arbitrary. For example, a multivariate meta Student *t* distribution (i.e., meta t_ν), with elliptical Student *t* copula $C_{\nu,\rho}^t$ with arbitrary marginal distributions, say, Gaussian, exponential or skew Student *t* margins.
[19] Used in the calibration of the multivariate copula model.

framework.

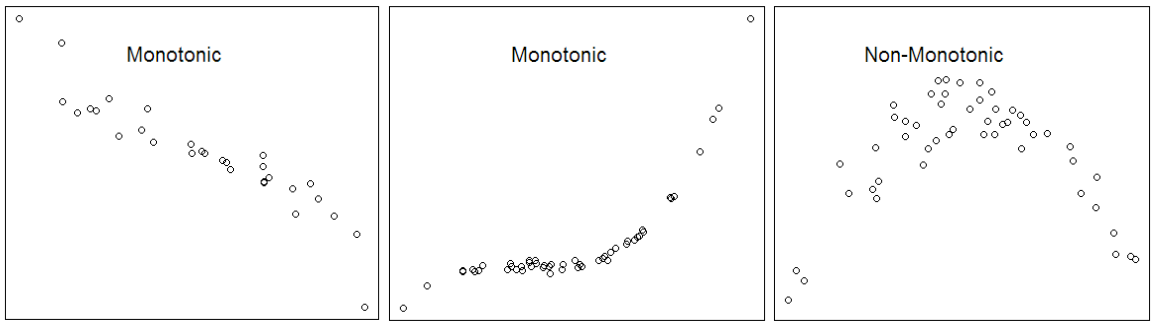


Figure 2.2: Monotonic and Non-Monotonic Relationships

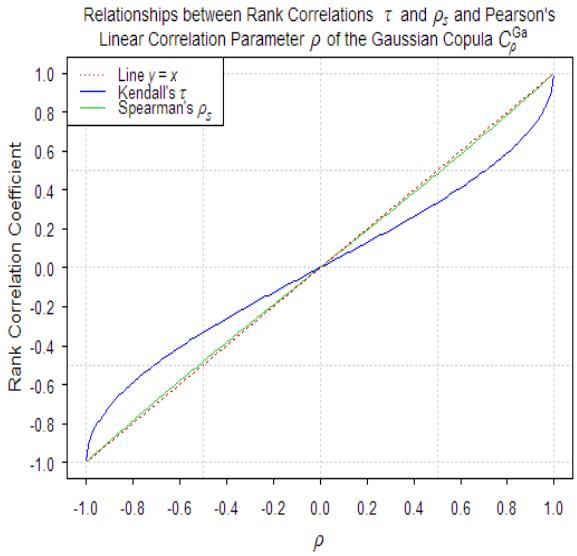


Figure 2.3: Relationships between Rank Correlation Measures and Pearson's Linear Correlation Measure

Under MPT, the dependence structure is contained within the multivariate Gaussian distribution model. The model is based on marginal normally distributed variables whose pairwise dependence relationships are captured by Pearson's linear correlation coefficient. The aforementioned building blocks[20] of the MPT framework allocates zero[21] density to extreme joint tail events (such as multi-asset market crashes) leading to severely overestimated diversification benefits. Including copula theory in a portfolio modelling framework offers a way of improving not only multivariate return distribution models, but also co-variation in the model framework.

Optimisation, risk management and stress testing in a portfolio can be organised around "the key idea that the risk of a set of positions can be decomposed into two major components:

1. The marginal risk associated with the variations of wealth of each risky position,
2. The cross-dependence between the change in the wealth of each position." (Malevergne

[20] That is, the multivariate Gaussian distribution, marginal Gaussian distributions and Pearson's linear correlation coefficient matrix. The implied copula in this framework is the Gaussian copula.
[21] The Gaussian copula has zero asymptotic tail dependence (Malevergne & Sornette, 2006, p. 212).

& Sornette, 2006, p. 272).

The decomposition is justified by the introduction of copula theory.

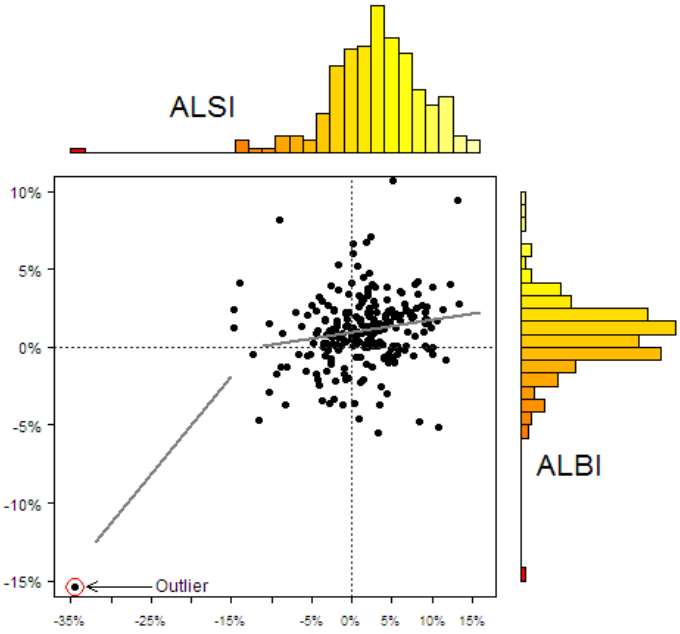


Figure 2.4: Dual Correlation Structure between the FTSE/JSE All Share Index (ALSI) and FTSE/JSE All Bond Index (ALBI)

In Figure 2.4, empirical outliers exert influence, which disrupts the efficacy of the linear correlation coefficient. The straight line segment leading up from the outlier into the bulk of the data represents the initial section of a local polynomial (non-linear) regression. The line segment in the bulk of the data represents the numeric Pearson correlation coefficient response of the ALBI to linear predictor ALSI, with the respective contemporaneous outlier removed. The complete dependence relationship appears monotone, indicating that a measure of monotone association may better summarise the relationship than that of a linear measure.

Figure 2.5 illustrates the concept of mapping a random variable through its CDF. The CDF F_X of a random variable X represents a non-linear, strictly increasing, monotone function acting as a two-way transformation mapping (a) from $F_X(x) \sim U_{[0,1]}$ and, the reverse, (b) from $F_X^{-1}(U_{[0,1]}) \sim f_X$, where f_X is the probability distribution function of X. The CDF of X in copula vernacular is called the "grade" of X and its inverse $F_X^{-1}(\cdot)$ the "quantile" of X.

A multivariate distribution function may be factorised into it's purely univariate components (i.e., the marginal distributions) and a purely joint, or collective dependence, component (i.e., the copula function). Essentially, a copula is a multivariate CDF dependence structure.

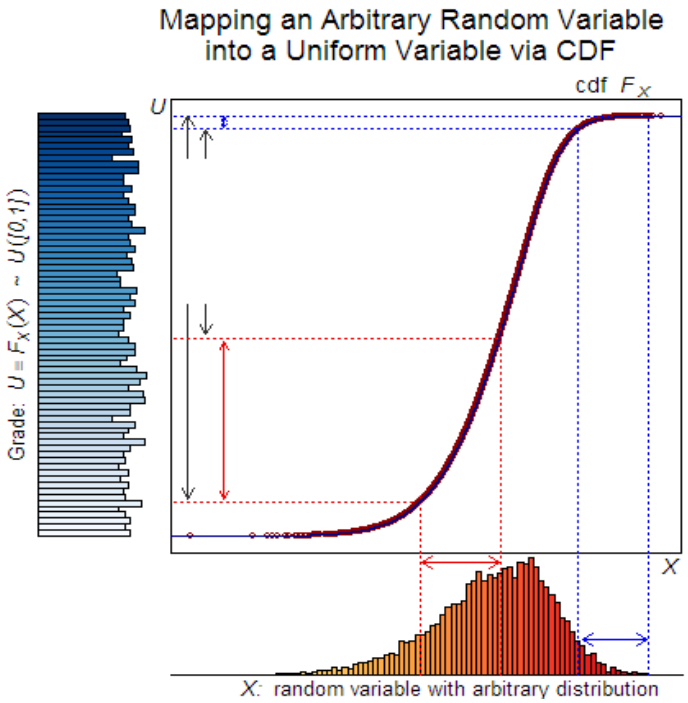


Figure 2.5: Mapping an Arbitrary Random Variable to a Uniform Variable

The multivariate distribution decomposition provides a flexible approach to managing aggregate randomness in the context of portfolio management. The factorisation, or copula technique, enables an elegant separation of marginal distributions from the joint dependence structure. There is an advantage to separate modelling at the two levels, the univariate and multivariate. At each level, the opportunity to fit more precise distributions is afforded. At the multivariate level, the copula framework subsumes both elliptical and non-elliptical distributions. Since most real-world multivariate distributions are non-elliptical, copula theory not only facilitates a universally valid alternative to the modelling of empirical distributions with rudimentary correlation-based distributions, but it also increases the range of applicability beyond the domain of traditional correlation-based approaches.

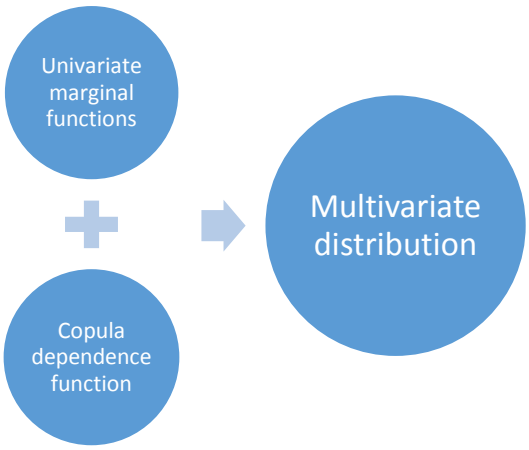


Figure 2.6: Factorisation of a Multivariate Distribution into Component Parts

Sklar (1959) formally introduced the copula notion into mathematical and statistical theory. The copula function is the multivariate CDF, on the unit hypercube $[0,1]^d$, of the d marginal grades (each on the unit interval $[0,1]$) that captures the pure dependence between (transformed) random vectors.

Embrechts and Hofert (2014) described a graphic worth noting and is adapted here. It may also be used to highlight the importance of shortcoming 5 on the Embrechts, et al. (1999) list, above. Figure 2.7 shows two empirical relationships between two hypothetical data sets (X_1, X_2). In what follows, H represents a multivariate distribution function, $F_1 = F_2 = F$ represents a univariate distribution function and C represents a copula function. Data is simulated for $n = 500$ realisations of $(X_1, X_2) \sim H$. In the left panel, CDFs F_1 and F_2 are standard normal $N(0,1)$ with $F(x) = \Phi(x)$; in the right panel F_1 and F_2 are standard exponential $Exp(1)$ with $F(x) = 1 - exp(x)$. The generated data sets may proxy, for example, historical risk-factor returns. The questions posed by the authors regarding dependence between $X_1$and X_2 concern, firstly, how the dependence may be modelled and, secondly, which of the left-hand or right-hand panels show a "stronger" dependence relationship. The margins in Figure 2.7 are supported on distinctly different scales and, thereby, disguise the underlying dependence structure.

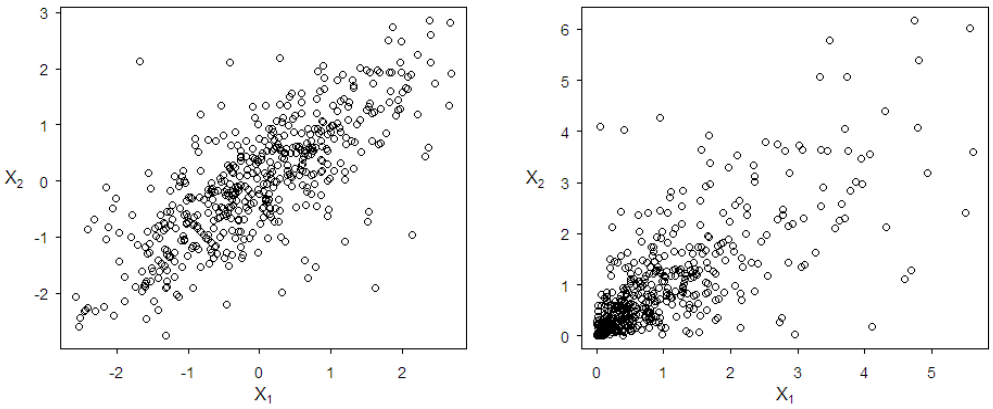


Figure 2.7: Simulations from Standard Normal and Standard Exponential Distributions

The Pearson correlation coefficients differ for the two panels, with the right panel showing stronger correlation. The copula approach, however, makes a like-for-like comparison possible. Margin influences are removed by transforming each variable via their respective CDFs to output on the uniform $[0,1]$ scale.

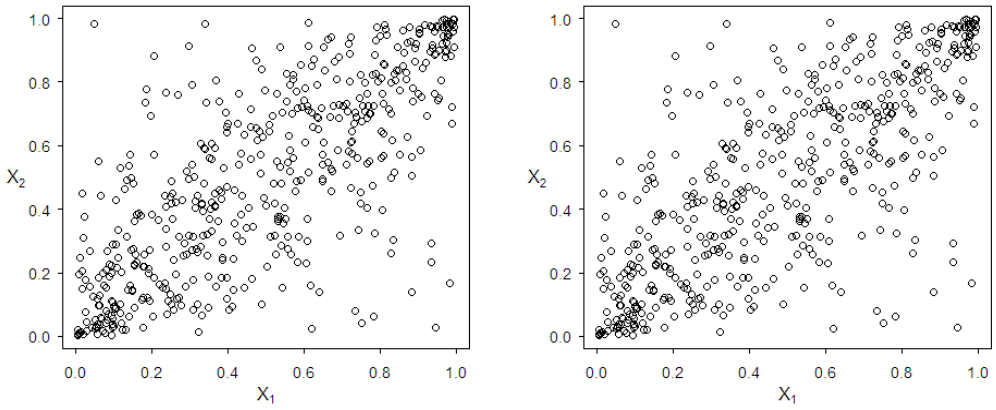


Figure 2.8: Copula Dependence Structure Underlying the Corresponding Simulated Data

The results in Figure 2.8 are obtained by fitting a copula C (in this case, a Student *t* copula) to $(U_1, U_2) = \big(F_1(X_1), F_2(X_2)\big) \sim C$. They show that when the data sets (X_1, X_2) are monotone-transformed, in this case by their respective non-linear marginal CDFs, their underlying dependence structure (i.e., the copula C) is exactly, not only in distribution, equal. Given that the margins are not known in practical applications, the typical approach is either to (1) assume a distributional form, (2) use the empirical CDF form or (3) use the semi-parametric form described in the previous section. Copulas have the property of being invariant to monotonic transformations of the margins, validating any of the three CDF inverse-transform approaches to creating pseudo-copula data.

The Student *t* copula may be considered a first-step extension of the Gaussian correlation-based dependence structure and is one of the most popular copulas in modelling multivariate financial data. The *t* copula is the elliptical copula derived from the elliptical multivariate *t* distribution. It extends the Gaussian copula (used implicitly in MPT) by way of an additional degrees of freedom (df) parameter[22]. Whereas the normal copula is not parameterised to assign probability mass to joint tail events, the addition of the df parameter ν does so in the *t* copula formulation. The parameter controls the degree of extreme co-movements, thereby also measuring the extent of departure from the Gaussian dependence structure (Mashal, Naldi & Zeevi, 2003). Schmidt (2006) highlighted the *t* copula advantage over the Gaussian in a bivariate portfolio setting, citing similar densities in the centre of the distribution, but with the *t* copula allocating density to joint tail observations. Klüppelberg, Kuhn and Peng (2007) noted growing interest in the *t* copula, primarily due to its properties of avoiding the curse of dimensionality[23], providing a robust way of dealing with tail dependence and its simplicity in simulating multivariate extremes.

[22] The Gaussian copula is nested within the *t* copula family. As the df estimate gets bigger, so too does the *t* copula resemble more and more the Gaussian copula. Functionally, $\{\nu: \nu > 2, \nu \in \mathbb{R}\}$ to ensure a finite second moment. Where $\nu > 30$ the multivariate *t* copula begins to "close-in" on the Gaussian representation; at the limit (where $\nu \to \infty$), the *t* copula degenerates to the Gaussian copula. In fact, for $\nu > 10^3$, Mashal and Zeevi (2002) show that the two copulas are essentially indistinguishable.

[23] That is, the issue of data sparsity encountered when fitting models to high dimensional data. In high dimensions, local neighbourhoods tend to be empty, affecting the error properties of density estimates.

Mashal and Zeevi (2002) investigated dependence patterns and extremal behaviour using the Gaussian and *t* copula framework applied to three representative financial asset classes (G5[24] equities, commodities and foreign exchange). The authors found empirical support for extreme co-movements in all data sets. In addition, the Gaussian copula was rejected in all data sets (with negligible error probability) and dimensionality effects were found to characterise the data, where extreme co-movements became increasingly significant and more pronounced as the number of underlying assets increased (dimension $d = 2,3,\ldots,9$ assets). Kole, Koedijk and Verbeek (2007) stress tested the adequacy of three copula models (the Gaussian, *t* and Gumbel[25]) and their ability to capture the risk of joint downward movements in a portfolio of stocks, bonds and real estate (proxied, respectively, by the S&P 500 Composite Index, JP Morgan Government Bond Index and National Association of Real Estate Investment Trusts [NAREIT] All REITs Index). The authors found that the trivariate Gaussian (resp. Gumbel) copula significantly underestimated (resp. overestimated) the risk of joint downward movements which, in turn, translated to overestimated (resp. underestimated) diversification benefits in the portfolio. The Student *t* copula, on the other hand, allocated an empirically commensurate and statistically significant amount of density in both the centre and the tails of the portfolio distribution. Martellini and Meyfredi (2007) chose the *t* copula to model the non-linear dependence structures underlying key risk factor proxies to a set of fixed income portfolios. The copula model was part of a framework proposed as an alternative to calculating VaR estimates and was tested on a bond portfolio of nine French benchmark-quality bond issues of varying durations. The authors showed a marked improvement in risk calculations compared to those obtained with standard Gaussian VaR and historical VaR methods. Fischer, Köck, Schlüter and Weigert (2009) investigated the performance of higher dimensional copula construction schemes. The authors set the (static) multivariate *t* copula as the benchmark dependence structure against which they compared over two dozen competing copula schemes in the four-dimensional case. The data sets studied came from three different markets (German equity, foreign exchange and commodities) from each of which four representative variables were selected. The authors showed the *t* copula to be superior to the Gaussian and Archimedean[26] copulas and also, due to the complexities associated with higher dimensionality and the large number of parameters, to the D-vine pair-copula decompositions considered. Wang, et al. (2012) built a dynamic, risk-on/risk-off, regime-based portfolio construction framework using the *t* copula to represent co-dependence in a portfolio of five asset classes (global equity, commodities, real estate, high-yield bonds and investment grade bonds). The authors chose the *t* copula for its noted ability "to better capture the effects of fat tails and allocate non-zero probabilities to observations that may

[24] G5, the Group of Five of the world's leading industrialised nations: France, Germany, Japan, the UK and US.
[25] The Gumbel copula, parameterised to reflect mostly tail dependence, extends univariate EVT techniques to the multivariate setting.
[26] Archimedean copulas with dimension 3 or higher allow only positive association between variables (Yan, 2007). This constraint supports the decision in this dissertation to use the *t* copula over Archimdedean copulas in the modelling framework.

occur outside of the range of historical returns".

Boubaker and Sghaier (2013) considered the bivariate case of five Archimedean copulas (the Gumbel, survival Gumbel, Frank, Clayton and survival Clayton) as a first step in a portfolio optimisation investigation. The authors modelled the copulas on an equity pair (the Dow Jones Industrial Average [DJIA] Index and the French CAC40 Index) and an exchange rate pair (USD/EUR and JPY/EUR) using, first, the raw log return series and, second, filtered residuals from an ARFIMA-FIGARCH model fit to the raw log returns. The Gumbel copula was the appropriate model in both cases, justified by its asymmetric and EVT-based characteristics. An important observation worth noting, however, was the higher co-dependence estimates reported in the filtered series than in the raw series, the implication being that the paired raw returns masked the true dependence structure. Termed a "long memory in volatility" characteristic, the ARFIMA-FIGARCH filtering process uncovered this masking characteristic by virtue of that dependence intensity outputs were higher in filtered returns than in unfiltered returns. In short, the authors reported that "the true dependence between financial returns series is higher than the one observed". In the second step, the authors showed that the traditional mean-variance efficient frontier constructed using unfiltered returns underestimated (resp. overestimated) portfolio risk as risk appetite grew (resp. shrunk). The efficient frontier constructed using a copula for dependency and accounting for long memory volatility in the univariate series uncovered higher risk-per-given-return features. The traditional investor assumed more risk than expected. Riccetti (2013) compared traditional MVO macro asset allocation decisions with a number of copula-constructed portfolios comprising two, three and four asset classes (a representative commodities index, the DJIA equity index, the Merrill Lynch US Treasury 1-10 years fixed income index and the USD/EUR exchange rate). Copulas considered were the Clayton, Gumbel, Frank, Gaussian, Student *t* and Canonical Vine copulas. For the specified two- and three-asset portfolios, the *t* copula performed best. For the four-asset portfolio, the author reported mixed results.

In terms of an ideal copula, Allen and Satchell (2014) argued that the following four characteristics should be accommodated:

1. Tail dependence;
2. Asymmetric tail dependence;
3. Heterogeneous tail dependence;
4. Scalable in high dimensions.

The authors proposed and implemented a skew Student *t* copula as an asymmetric dependence model for a portfolio of five asset classes, viz. US equities (S&P 500), MSCI EAFE[27] equities,

[27] MSCI EAFE Index: Morgan Stanley Capital International's Europe, Australasia and Far East equities index.

corporate bonds (Barclay's Corporate Bond Index), commodities (GSCI[28]) and US real estate (FTSE/EPRA REITs[29]). Although the authors reported net gains over competing copula models (such as the Gaussian and Student *t*), they also noted the increased complexity embedded in the skewed *t* framework. Indeed, Wei, Li, Cao, Sun, Liu and Li (2013) noted the computationally challenging and structurally complex task associated with modelling asymmetries and skewness of joint distributions of returns in high dimensional space. Comparatively, the *t* copula is not only computationally tractable, but it also admits ideal characteristics 1, 3 (partially) and 4. The *t* copula admits tail dependence via its functional structure, allows for partial heterogeneous levels of dependence via the pair-wise construct of the elements of its (rank-measured) covariance matrix parameter (Owen & Rabinovitch, 1983) and is scalable in high dimensions. Tsuchida, Giacometti, Fabozzi, Kim and Frey (2014) included the skew Student *t* copula in a study of copula dependency in sovereign bond returns across seven Eurozone countries (France, Germany, Greece, Ireland, Italy, Portugal and Spain). The authors used ARMA-GARCH models based on five different innovation distributions and, for each model, applied four copula dependence structures: Gaussian, Student *t*, skewed Student *t* and multivariate normal tempered stable. The study concluded that the multivariate Student *t* distribution gave the best fit to the empirical distribution.

A fifth characteristic may be noted in that, generally, the copulas used in finance and risk management are easily simulated, lending themselves well to Monte Carlo studies of risk (McNeil, et al., 2005).

This paper uses the radially symmetric[30] Student *t* copula to represent multivariate dependence between asset and risk factor returns which, respectively, are modelled with (arbitrary) semi-parametric distributions. The completed distribution is referred to as a "meta *t* distribution". Inanoglu and Ulman (2009) defined a "meta-x distribution" as a "joint distribution created from an 'x'-copula model using arbitrary margins" (see also Demarta & McNeil, 2005; Fang, Fang & Kotz, 2002; Quessy & Bellerive, 2013). Copula theory lends itself well to a top-down, macro-driven portfolio management system in allowing, via the univariate channel, for a "bottom-up approach to multivariate model building" (McNeil, et al., 2005, p. 185).

There are several methods to fit copulas to data, such as method-of-moments (using rank correlation), exact maximum likelihood (which attempts to simultaneously estimate all parameters of the marginal and copula models in a single optimisation) and inference-functions for margins (where ML estimation is split into two steps: margins estimated parametrically and the likelihood function maximised with respect to the vector of marginal parameter estimates from step one). This

[28] GSCI: Goldman Sachs Commodity Index.
[29] FTSE EPRA/NAREIT: The Financial Times Stock Exchange European Public Real Estate Association/NAREIT US Super Liquid Index.
[30] Copulas of multivariate elliptical distributions are radially symmetric (Joe, 2015, p. 65; McNeil, et al., 2005, p. 196).

dissertation uses canonical maximum likelihood (CML), where copula parameters are also estimated in two steps, but without specifying the marginal distributions. The distribution-free margins may be estimated using the empirical CDF or semi-parametric CDF (discussed in the previous section), a pseudo-sample of uniform $(0,1)$ variates[31] formed via inverse transformation and the copula parameters estimated by ML as a function of the pseudo-sample. The articles by Bouyé, Durrleman, Nikeghbali, Riboulet and Roncalli (2000, pp. 23-28) and Scaillet and Fermanian (2002) and the text of Cherubini, Luciano and Vecchiato (2004, pp. 153-174) may be referenced for details on copula fitting methods. Specifically, this dissertation follows the procedure detailed in Ruppert (2011, pp. 188-191). Details on CML are given in Chapter 3.5.2.

Copula theory is reviewed in the monographs of Joe (2015), Nelsen (2006) and Ruppert (2011); for empirical applications, the monographs of Cherubini, et al. (2004), Jondeau, et al. (2007) and McNeil, et al. (2005). For articles on the subject, the reader is referred to Bouyé, et al. (2000), Embrechts, et al. (1999), Embrechts, et al. (2003) and Meucci (2011).

In summary, the standard correlation measure has been shown to have a number of shortcomings when applied to financial data, similarly for univariate and multivariate Gaussian distributions. Copula theory admits a flexible methodology to capture and overcome many of the shortcomings. Given that bivariate data is easier to visualise than multivariate data, Figure 2.9 provides a graphical summary of the methodology in the bivariate case. The copula dependence structure in a portfolio comprising a representative equity asset class and bond asset class is output. The portfolio's marginal return distributions are first transformed via their respective CDFs into uniform variates, before "gluing" them together to form the joint dependence distribution. This process admits simulation methods for the fitted elliptical *t* copula and transfers a major advantage to practitioners modelling portfolios. The advantage is discussed next in the Portfolio Optimisation section.

[31] Uniform $[0,1]$ is scaled to uniform $(0,1)$, where required. This is because the maximum likelihood routine evaluates the copula density at each uniform observation and, for pseudo-copula data on the boundary of the unit cube, may evaluate to infinite density at the corresponding boundary observation (McNeil, et al., 2005, p. 233).

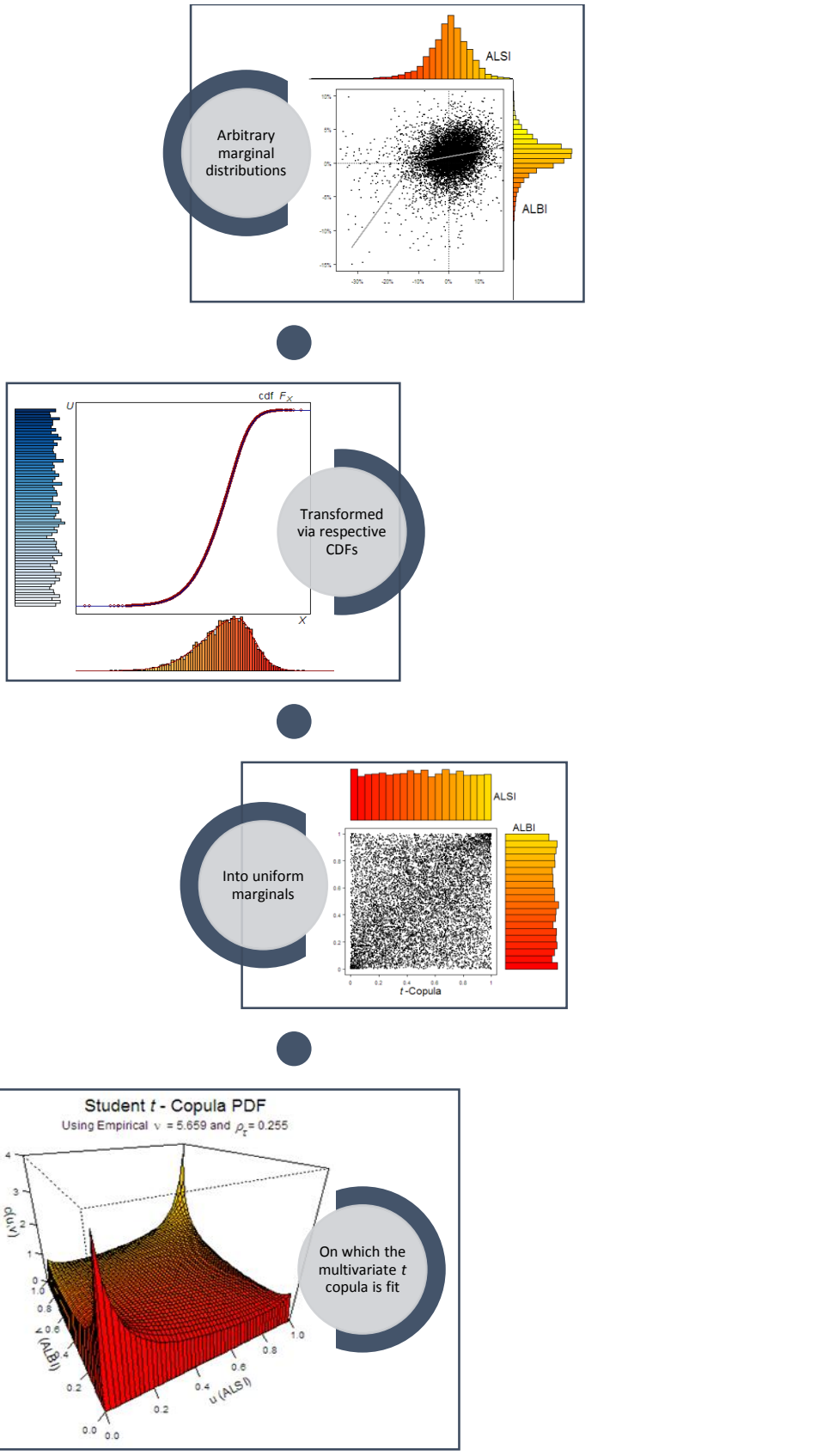


Figure 2.9: Flow Diagram of a Bivariate Student *t* Copula Density Construction

2.4 Portfolio Optimisation

The framework of this dissertation is based on Monte Carlo simulations (draws from a parametric distribution) of log returns for portfolio asset classes that have been conditioned on forward-looking scenarios. The concept behind the Monte Carlo approach used is to repeatedly simulate a correlated random process for the multivariate financial variable (i.e., the portfolio, its risks and the collective dependence structure) and then use the output as input to simulate random processes for the univariate financial variables (i.e., asset classes and risk factors). The resulting distribution of portfolio values covers a wide range of possible scenarios and allows sampling from anticipated states of the world. Importantly, the simulated returns maintain the dependence structure induced by the copula (Brandimarte, 2014, p. 283). Scherer (2002, p. 152) noted that, while it required considerably more effort from investors than traditional Markowitz optimisation[32], scenario optimisation is worth the effort if returns are non-normally distributed.

Nyström and Skoglund (2002a) emphasised sound economic reasoning and informed, considered judgment, with respect to constructing forward-looking scenarios[33], as being a value-add approach over strictly quantitative (e.g., cointegration) approaches. By simulating forward-looking scenarios for all assets, the practitioner is able to "turn a stochastic problem into a deterministic problem [and] in this form the problem can now be solved using mathematical (mostly linear) programming techniques" (Scherer, 2002, p. 141). Meucci, Gan and Lazanas (2007) built a comprehensive scenario-based market representation using Monte Carlo simulations. Alexander (2008, p. 364) provides a guideline to creating a mathematically coherent hypothetical distribution scenario for a vector of risk factors to a portfolio.

Grégoire, Genest and Gendron (2008) used simulations from a fitted *t* copula to construct an empirical predictive price distribution of crude oil and natural gas contracts. Wang, et al. (2012) used a fitted *t* copula to simulate forward-looking scenario-based outcomes to extrapolate portfolio CVaR beyond the range of historical observations. Brandimarte (2014, p. 283), Cherubini, et al. (2004, p. 181) and McNeil, et al. (2005, p. 193) provide versions of the algorithmic steps required in simulating random variates from the multivariate *t* copula.

In 1996, the Basel Committee adopted the VaR quantile measure into its capital-adequacy framework (Basel Committee on Banking Supervision, 1996). VaR has since become widely used as a risk management tool by corporate treasurers, dealers, fund managers, financial institutions and regulators (Alexander & Baptista, 2004). Notwithstanding it being an industry standard, the risk measure has a number of criticisms. Bouyé, et al. (2000) and McNeil, et al. (2005) noted systematic

[32] The ease of use attributed to widely available mean-variance optimisation software.
[33] The term "scenarios" is consistent with the term "information hypercubes" referred to in Nyström and Skoglund (2002a).

underestimation of risk in Gaussian distribution-based VaR models. On the other hand, the copula approach to portfolio optimisation enables not only VaR, but also other risk measures (e.g., variance and CVaR) to be expressed without the Gaussian hypothesis. Artzner, Delbaen, Eber and Heath (1999) defined a set of axioms a risk measure should satisfy in order to qualify as a "coherent measure of risk". Importantly, VaR has been shown to disobey the axiom of subadditivity[34], a consequence of which may result in otherwise higher risk concentration in the portfolio. This is at odds with the key principle of diversification. Furthermore, VaR, as a financial metric, represents the minimum amount that a portfolio can expect to lose within a specified time period, at a given confidence level. The metric is silent on the possibility and magnitude of extreme events (i.e., high-impact, low probability tail events beyond the given confidence, or quantile, level remain undetected by construction). Daníelsson (2002) and Szegö (2002) detail some of the criticisms of VaR. See Figure 2.10 for a graphical description of VaR and CVaR for probability level, in this case, 5%[35]. A 5% VaR reflects the expected minimum loss that the portfolio can expect 5% of the time. A 5% CVaR reflects the size of the expected loss, given that a 5% VaR event occurs. Sarykalin, Serraino and Uryasev (2008) provide a thorough mathematical and case-study comparison of VaR and CVaR in the risk management and optimisation context.

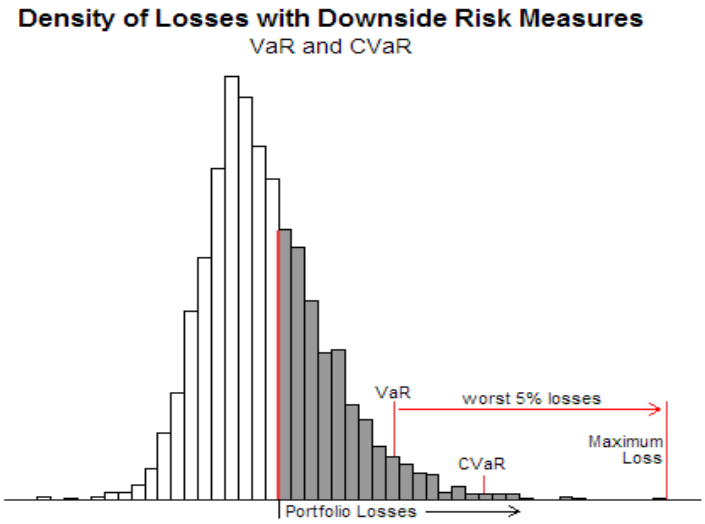


Figure 2.10: Graphical Description of VaR and CVaR Risk Measures

A coherent measure of risk that has gained popularity is the CVaR asymmetric, downside risk measure[36]. For general distributions, both continuous and discrete, CVaR measures the weighted average of VaR and the expected losses strictly exceeding VaR (i.e., the expected loss in the tail,

[34] The subadditivity property implies that the measure of risk in a portfolio should be less than or equal to the sum of the measures of risk of the instruments comprising the portfolio. It is the mathematical description of the diversification effect (Fabozzi, et al., 2007). Formally, for a measure ρ of the risk X in a set of all risks $\mathcal{G}$, for all $X_1, X_2 \in \mathcal{G}$, $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$. VaR is not always subadditive outside the elliptical distribution framework (Embrechts, et al., 1999). Outside of this framework (e.g., in a fat-tailed framework), it has been shown to be subadditive only when calculated very deep in the tail (Daníelsson, Jorgensen, Samorodnitsky, Sarma & de Vries, 2013).

[35] It is convention to invert the VaR and CVaR x-axis, meaning that risk and losses increase as one moves into the right tail of the distribution. Positive, larger values represent increasing downside risk (i.e., bigger tail losses).

[36] Coherency, with respect to Artzner, et al. (1999), proved by Pflug (2000).

given that the loss is greater than or equal to the VaR value). Rockafellar and Uryasev (2000) developed a linear programming (i.e., non-smooth programming) optimisation technique efficient in simultaneously minimising the CVaR risk measure, subject to a minimum expected return constraint, and calculating VaR. The linear programming technique is effective under both normal and non-normal distributions. The authors showed that, for Gaussian distributed loss functions, the solutions to minimising variance in the standard Markowitz MV framework and to minimising CVaR generated the same optimal portfolio. An optimal, or efficient, portfolio is the investor-preferred set of asset weights that maximises the investor's risk-return preferences (i.e., the portfolio that best approximates the risk-return characteristics of the investor's utility function). Optimal sets plot an "efficient frontier", depicting an equivalence representation of concave reward and convex risk functions (see Figure 2.11). Krokhmal, Palmquist and Uryasev (2002) adapted the Rockafellar and Uryasev (2000) technique to optimisation problems with CVaR constraints, as opposed to expected return constraints per Rockafellar and Uryasev (2000). Specifically, the authors swapped the CVaR function and expected return in the Rockafellar and Uryasev (2000) problem formulation: the reward function is optimised subject to, among others, a CVaR constraint. The authors constructed efficient frontiers, for both mean-variance and mean-CVaR optimisations, of portfolios comprising stocks from the S&P 100 index and a cash asset as the investible set for the portfolio. It was shown that as the CVaR constraint decreased (i.e., as the quantile threshold decreased), the expected return increased. In other words, as tail risk (CVaR) constraints decreased, the portfolio expected return increased. Important to the case supporting mean-CVaR optimisation, the authors showed that, for any given return level, the MV portfolio reflected higher CVaR tail risk levels than in the efficient mean-CVaR portfolio. The difference in CVaR risk levels between the two portfolios further increased as the constraint increased. The authors also confirmed (per Rockafellar & Uryasev, 2000) the closeness of solutions of CVaR and MV optimisation problems as being a function of "close-to-normal" distributions of returns. For non-normal, especially asymmetric, distributions, however, the authors noted that MV and mean-CVaR optimisations may lead to significantly different optimal portfolios.

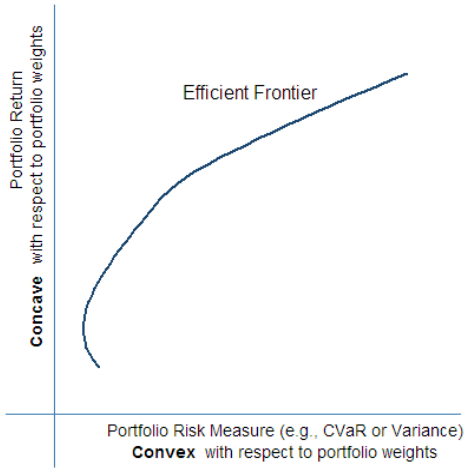


Figure 2.11: Efficient Frontier as an Equivalence Representation

Investors' risk preferences tend to be asymmetric, where agents place greater weight on downside risks relative to upside gains (Ang, Chen & Xing, 2006). Ang, et al. (2006) isolated a robust premium associated with downside risk exposure in a cross-section of equity returns[37]. Fabozzi, et al. (2007) remarked on the mean-variance and mean-CVaR methodologies. On the one hand, the MV approach defines risk symmetrically by the variance statistic of the returns distribution; hence, incorporating information from both lower and upper tails of the distribution, with losses and gains contributing equally to the risk magnitude. On the other hand, the mean-CVaR approach considers (asymmetrically) only the segment of the tail of the distribution that contributes to high losses. Sheikh and Qiao (2010) incorporated and advocated the use of CVaR in their asset allocation model, noting the measure as an improved risk quantifier over the variance measure in a non-normal environment. The authors showed the measure's ability to capture investors' asymmetric risk preferences as well as left tail incidences of events induced by skewed and leptokurtic return distributions. Wang and Zheng (2010) used CVaR as the downside risk measure in a study that included a comparison of the measure's use to the traditional MVO method. Monthly returns were considered on indices representing seven asset classes[38]. The authors showed sub-optimal performance of the MV portfolios in terms of underestimating downside risk. Optimal mean-CVaR portfolios reflected slightly higher variances than portfolios on the optimal mean-variance efficient frontier. In compensation, though, the optimal mean-CVaR portfolios exhibited a net comparable gain in terms of containing much less downside risk. Xiong and Idzorek (2010) compared the variance risk measure in MVO to the CVaR measure in mean-CVaR optimisation. The authors simulated hypothetical asset classes and compared optimisations under various moment characteristics by varying skewness and kurtosis levels. A representative asset class case-study was also investigated. Variance considers strictly the first two moments of a variable; CVaR, the first four moments (i.e., additional kurtosis and skewness moments outside of the Gaussian domain). The authors noted similar allocations between the MVO and mean-CVaR optimisations under Gaussian conditions and, more generally, under symmetric conditions combined with uniform kurtosis characteristics. When skewness and kurtosis levels varied among asset classes, the two frameworks output differing asset allocations, with the mean-CVaR optimisation preferring assets with "higher skewness, lower kurtosis and lower variance".

Meucci, et al. (2007) succinctly summarised the above three portfolio risk measures as follows:

1. Volatility as the regular fluctuations in return.
2. VaR as the best of worst case scenarios.

[37] Listed equities restricted to the NYSE, AMEX and NASDAQ exchanges.
[38] MSCI All World equity index, GSCI for representative global commodities, a blend of benchmark indices for credit, Barclays Capital US Treasury Intermediate index for interest rate, a blend of benchmarks for Treasury Inflation Protected Securities for inflation, NAREIT for real estate and Citi 3-month US Treasury bills for cash.

3. CVaR as the mean of worst case scenarios.

Stoyanov, Rachev, Racheva-Iotova and Fabozzi (2010) summarised the attractive properties of CVaR as a measure that:

1. Gives an informed view of losses beyond VaR (by accounting for risks in excess of VaR).
2. Is a convex function with respect to portfolio weights. Convexity eliminates the possibility of a local minimum being different from a global minimum (this is a key property, computationally, in optimising portfolios in terms of CVaR being continuously differentiable with respect to a quantile threshold[39]). In a scenario simulation model[40], VaR is a non-convex, non-smooth function of weights, with multiple local extrema (Uryasev, 2000). See also Figure 2.11.
3. Is subadditive, making it practical beyond the one-instrument setting, as well as satisfies the Artzner, et al. (1999) set of intuitively appealing coherent risk measure properties.
4. As a form of expected loss (i.e., conditional expected loss), is convenient for use in scenario-based portfolio optimisation. It is also a natural form of risk-adjustment to expected return.

Not accounted for in the above list, but which may be considered an attractive property of CVaR is:

5. Linear programming and non-smooth optimisation algorithms allow the "handling [of] portfolios with very large numbers of instruments and scenarios.… For instance, a problem with 1,000 instruments and 20,000 scenarios … can be optimised … in less than one minute." (Uryasev, 2000).

It is worth noting that CVaR-optimisation is not a panacea for risk management, as is shown in a minimum-variance versus minimum-CVaR investment style study[41]. Tokpavi and Vaucher (2012) showed the outperformance of the variance measure compared to the CVaR measure in portfolios that seek to obtain (and maintain) minimum global risk. However, this dissertation does not seek to obtain the global minimum CVaR portfolio; the objective is to obtain sets of portfolios with maximised expected returns across varying levels of CVaR (equivalently, with minimised CVaR for given levels of returns).

[39] Rockafellar and Uryasev (2000, pp. 23-26).
[40] Here, the term "scenario" is used in the context of VaR and CVaR calculations where the analytical representation of the portfolio density function is not available. Calculations are sample-built, on historical portfolio observations, or from Monte Carlo simulations (Uryasev, 2000).
[41] The "minimum-variance", or low-volatility (or low-beta), equity investment style has gained popularity since the 2008 global financial crisis (Brochart, Taillardat & Jourovski, 2013). This is true also in South Africa. For example (though not strictly "minimum-variance"), low-beta, mid-capitalisation companies are used in the construction of the long side of a portfolio in a hedged investment strategy, with the polar high-beta, large-capitalisation domain of companies used in the construction of the short hedge side (Tower Capital, 2013).

In summary, non-normality can impact asset allocation by way of different asset classes exhibiting different downside risks. For example, equity type investments entail greater degrees of downside risk than fixed income type investments (Sheikh & Qiao, 2010). Incorporating non-normality, however, necessitates an alternative to Markowitz's mean-variance framework. CVaR is an improved risk measure over variance in terms of optimising on tail risk. It also makes no assumptions on the underlying distributions of assets. The potential for marginal gains in switching to mean-CVaR optimisation is acknowledged and is the method that will be used in this study. The illustration in Figure 2.13 provides an integrated summary of the selection-to-simulation-to-optimisation process (Absa Capital, 2013). The following steps, demarcated in the graphic, correspond to the steps taken in the forward-looking scenario-based simulation and asset allocation process:

1. Select risk factor input variables and construct scenarios.
2. Simulate returns from multivariate distribution using fitted *t* copula.
3. Match scenarios and extract corresponding portfolio returns.
4. Measure the conditional asset class return distributions.
5. Build efficient frontiers using mean-CVaR as the optimisation criteria.

Figure 2.12 is used to introduce the subsequent methodology chapter. The graphic captures the essence of the methodological steps relevant to the dissertation.

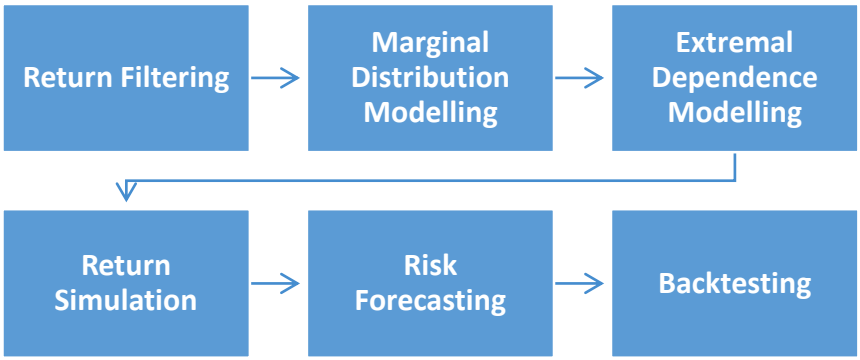


Figure 2.12: Methodological Steps Used to Build the Framework

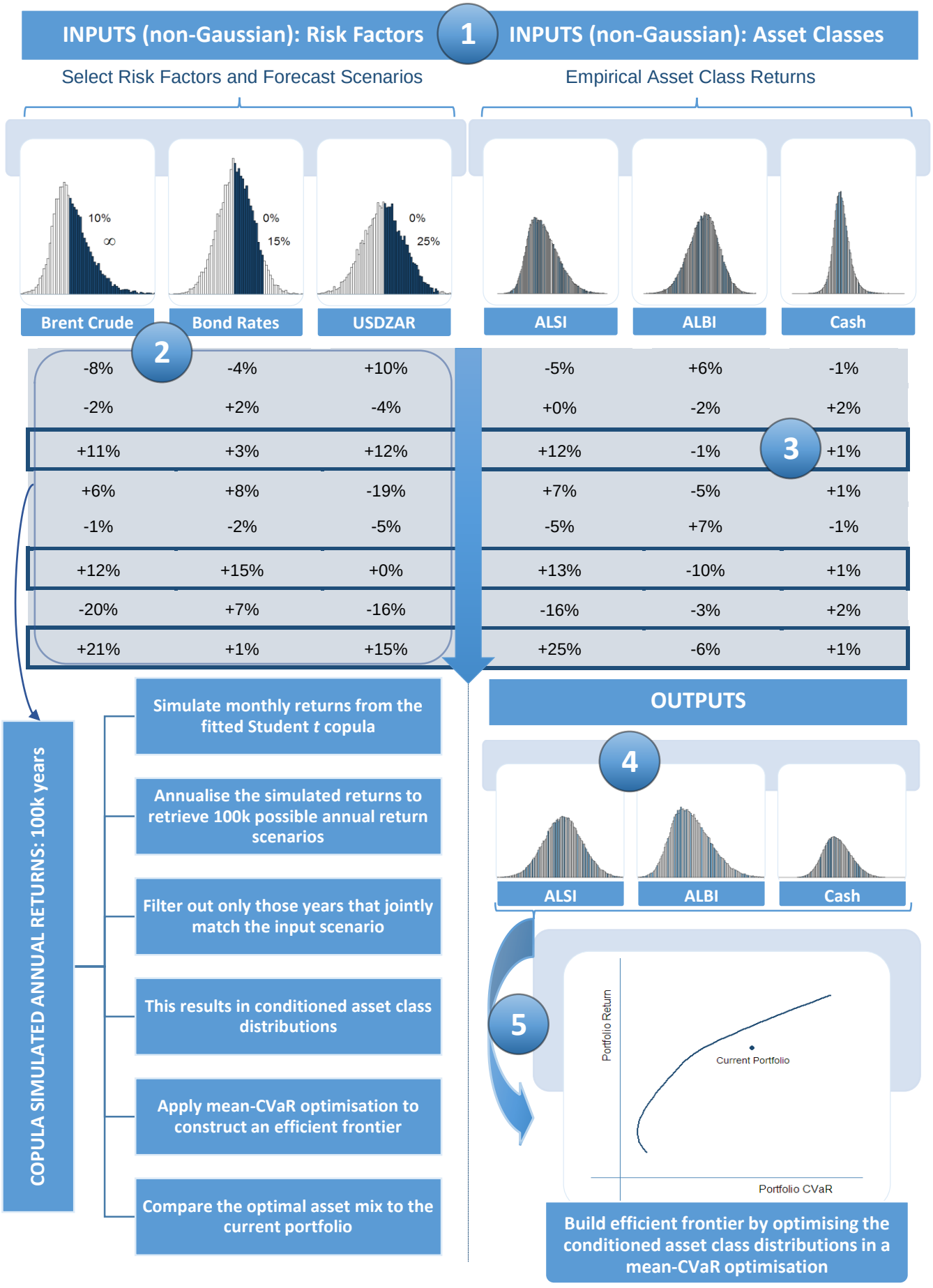


Figure 2.13: Illustration of the Forward-Looking Scenario-Based Simulation and Asset Allocation Process

Chapter 3

Methodology and Results

This chapter combines two purposes: first, it describes the methodology and theory used in building the framework and, second, it presents the main results from each methodology of the framework, concluding with an out-of-sample evaluation. It is divided into six sub-sections: Data Description, Exploratory Data Analysis, TSA, EVT, Dependence and Portfolio Optimisation.

3.1 Data Description

This study uses 18 variables in the model: 6 asset class variables and 12 key sector and asset class return driver variables. Asset classes are selected based on a representative, diversified portfolio. Asset classes (e.g., equities) may be decomposed into sector and sub-sector levels (e.g., industrial sector → sub-sectors: aerospace and defence, electronic and electrical equipment, general industrials, industrial engineering, industrial transportation, etc.) with, on the one hand, the caveat that the significance of the copula model diminishes as correlation among portfolio variables becomes less varied (Riccetti, 2013) but, on the other hand, with the confidence that the copula model is robust to the opportunity set typical of a sophisticated investor[42] (Wang, Sullivan & Ge, 2012).

The data were obtained from INET BFA. End-of-month observations, spanning 20 years, from the period 31 March 1994 to 30 April 2014 were used to fit the model. End-of-month observations for the period 30 April 2014 to 30 April 2015 were used to evaluate the model out-of-sample.

Indices were used to represent portfolio asset classes and portfolio return drivers. For portfolio asset classes: FTSE/JSE All Share Total Return Index (ALSI) for South African equities, FTSE/JSE All Bond Index (ALBI) for South African fixed income, MSCI World Index (MSCI.WRLD.ZAR) for international equities[43], Alexander Forbes money-market index (GMC1) for cash, JP Morgan Global Government Bond Index (GLOUS) for international bonds and the FTSE/JSE Listed Property Index (J253T) for South African real estate. For portfolio return drivers: domestic currency (USDZAR), dollar strength/weakness proxy[44] (EURUSD), Brent crude oil[45] (BRSPOT), Goldman Sachs

[42] Typical opportunity set of a sophisticated (U.S.) investor includes the following asset classes: large value equity, large growth equity, small value equity, small growth equity, non-U.S. developed market equities, emerging market equities, commodities, non-U.S. REITs, U.S. REITs, U.S. TIPS, U.S. bonds, non-U.S. bonds, global high yield bonds and cash (Xiong & Idzorek, 2011).
[43] Converted from USD to ZAR to be consistent with the perspective of a holding in a South African portfolio.
[44] The USDZAR and EURUSD exchange rate components admit a means of representing USD strength (via EURUSD returns) as distinct from ZAR weakness (via USDZAR returns).
[45] Changes in the oil price typically impact on the domestic expected rate of inflation.

Commodity Index[46] (GSCI), gold price[47] (GLFX), platinum price[48] (PLAT), MSCI Emerging Market equities (MSCI.EM.USD), S&P 500 developed market equities (FSPI), JP Morgan Emerging Market Bond Index (JPEMBI), JP Morgan US Government Bond Index (USALCI), the 90-day Banker's Acceptance Rate (RBAS) for domestic short-term interest rates and the 10-year point on the yield curve (JAYC10) for the domestic medium-to-long-term South African Treasury bill yield[49]. Figure 3.1 illustrates heterogeneity across portfolio asset classes based on the empirical monthly log returns.

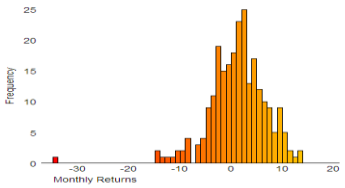
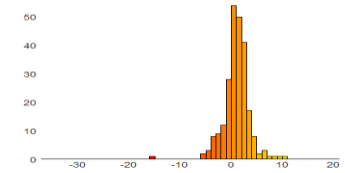
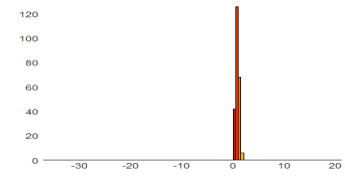
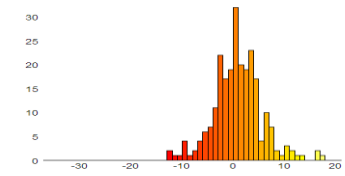
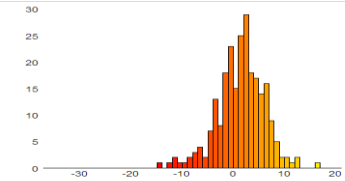
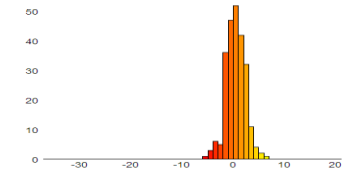
Series	Mean %	Standard Deviation %	Distribution %
FTSE/JSE All Share Index	1.2	5.6	
FTSE/JSE All Bond Index	0.9	2.5	
Money-Market Cash	0.8	0.3	
MSCI World Int'l Equities	0.9	4.7	
FTSE/JSE Listed Property Index	1.5	4.6	
JP Morgan Global Govt Bond Index	0.5	1.9	

Figure 3.1: Summary Statistics and Empirical Histograms

[46] Changes in hard commodity (i.e., metal) prices typically impact on domestic equity prices (e.g., mining and industrial sectors).
[47] Changes in gold prices may proxy for global inflation expectations as well as proxy for market instability through its status as being a safe-haven asset.
[48] Changes in platinum prices typically impact on domestic equity prices (e.g., in mining and industrial sectors).
[49] Changes in interest rates typically impact on ALSI, ALBI and sector returns.

Table 3.1 presents summary statistics for the 20 year sample of monthly log returns of all variables. For each displayed is the mean, median, maximum, minimum, standard deviation, skewness, excess kurtosis and test statistics and *p*-values of the Jarque-Bera test for normality, the Augmented Dickey-Fuller (ADF) test for a unit root (i.e., for non-stationarity) and the Ljung-Box (LB) test for independence (at the first autocorrelation lag). In the column of means, RBAS and JAYC10 evaluate to negative numbers, suspected to be an anomaly of the low sampling frequency of the data. The JP Morgan Global Government Bond and US Government Bond Indices were the only two variables that possibly could be approximated by a Gaussian distribution (Jarque-Bera test *p*-values of 0.135 and 0.017, respectively). The money-market cash variable (GMC1) exhibited a degree of statistically significant non-stationarity (ADF test *p*-value of 0.082). The Global Government Bond (GLOUS) and commodities (GSCI, GLFX and PLAT) indices showed minor evidence of dependence in the returns (LB test *p*-values of 0.089, 0.084, 0.013 and 0.011, respectively), while, naturally, the money-market cash and domestic short-term interest rate (RBAS) variables showed strong evidence of dependence in the returns (LB test *p*-values are both approximately zero).

For simplicity, portfolios are constructed with the first 4 of the 6 asset classes (ALSI, ALBI, MSCI.WRLD.ZAR and Cash). Cash is the flagship asset class in a risk-averse investor's portfolio, typically exhibiting the lowest average real return. The advantage of a cash holding, however, is the certainty of its nominal return not becoming negative. The MSCI World Index, converted to Rand denomination, is used to proxy offshore equities and the allocation limit is constrained to the domestic limit of 25% of the portfolio. The EURUSD time series admits a switch in the sample at the time the euro was introduced to world financial markets. In lieu of a synthetic historical price series constructed as weighted averages of the previous currencies, the INET BFA sample transitions from the Deutsche mark (DEM), as the prior assumed benchmark European currency, to the euro on 1 January 1999, at an exchange rate of USD 1.172 per euro (European Central Bank, 1999) [50]. Prior to the switch date, the EURUSD series is proxied by the DEMUSD series.

3.2 Exploratory Data Analysis

Since prices in financial series are mostly non-stationary, they are transformed to returns for modelling. In particular, logarithmic returns are constructed as:

$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right) = \ln(p_t) - \ln(p_{t-1}), \forall t \in \mathbb{Z}, \quad (1)$$

where r_t denotes the continuously compounded return (equivalently, the log return) at time t, p_t denotes the asset price (or value of the series) at time t, $\ln$ denotes the natural logarithm and $t = 1, \dots, n$, for sample size n.

[50] Personal correspondence with INET BFA.

Table 3.1: Descriptive Statistics of the Data Set

Index Name	Mean %	Median %	Max. %	Min. %	Std. Dev. %	Skewness	Kurtosis	Jarque-Bera Test[51]		ADF Test		Ljung-Box Test	
								Statistic	*p*-Value	Statistic	*p*-Value	Statistic	*p*-Value
ALSI	1.22	1.60	13.43	-34.58	5.62	-1.27	6.24	457.29	0.000	-6.39	0.01	0.33	0.568
ALBI	0.98	1.12	10.69	-15.37	2.52	-0.85	7.93	663.58	0.000	-7.39	0.01	0.27	0.606
MSCI.WRLD.ZAR	0.89	0.67	17.14	-12.79	4.72	0.22	1.28	18.48	0.000	-4.90	0.01	0.44	0.510
GMC1 Cash	0.85	0.85	1.72	0.42	0.31	0.48	-0.50	11.69	0.003	-3.24	0.082	240.81	0.000
GLOUS	0.47	0.30	6.82	-5.12	1.85	0.07	0.62	4.00	0.135	-7.69	0.01	2.88	0.089
J253T	1.47	1.81	16.71	-15.00	4.60	-0.36	1.03	15.84	0.000	-6.08	0.01	1.58	0.209
USDZAR	0.46*	0.43	17.02	-12.64	4.45	0.56	1.46	34.03	0.000	-5.97	0.01	0.13	0.713
EURUSD	0.06*	0.17	9.77	-10.32	2.95	-0.21	0.92	10.42	0.005	-6.17	0.01	0.09	0.754
BRSPOT	0.86	1.66	33.22	-49.20	9.80	-0.70	3.20	122.74	0.000	-6.85	0.01	0.20	0.652
GSCI	0.55*	1.34	20.93	-37.97	6.61	-1.07	4.52	251.61	0.000	-6.43	0.01	0.56	0.453
GLFX	0.50	0.14	16.20	-12.44	3.71	0.36	1.68	33.66	0.000	-5.08	0.01	2.98	0.084
PLAT	0.52	0.81	22.01	-38.20	6.19	-1.29	6.76	527.09	0.000	-6.13	0.01	6.13	0.013
MSCI.EM.USD	0.31*	0.69	15.41	-34.65	7.04	-1.14	3.54	178.49	0.000	-5.72	0.01	6.54	0.011
FSPI	0.60	1.21	10.23	-18.56	4.42	-0.90	1.66	60.27	0.000	-5.22	0.01	2.11	0.146
JPEMBI	0.91	1.41	10.17	-33.88	4.04	-2.99	22.55	5489.29	0.000	-6.15	0.01	0.04	0.833
USALCI	0.47	0.52	5.48	-3.38	1.29	-0.12	0.87	8.17	0.017	-7.23	0.01	2.02	0.155
RBAS	-0.24*	0.00	28.09	-13.44	4.21	0.62	8.96	824.74	0.000	-5.07	0.01	39.17	0.000
JAYC10	-0.19*	-0.51	22.86	-15.34	4.46	0.57	3.33	124.78	0.000	-7.28	0.01	0.25	0.619

Selection of representative portfolio asset classes

Actual portfolio asset classes used in this dissertation

Selection of asset class return drivers (used in this dissertation)

[51] Results are consistent with those found with the Anderson-Darling and Shapiro-Wilk tests for normality.

* Variables did not require demeaning (refer Chapter 3.3.1).

Financial return time series represented by equation (1) exhibit many of the so-called stylised facts, mentioned in Chapter 1.2. To illustrate the univariate data modelling process, we consider the ALSI variable as a representative variable for the data.

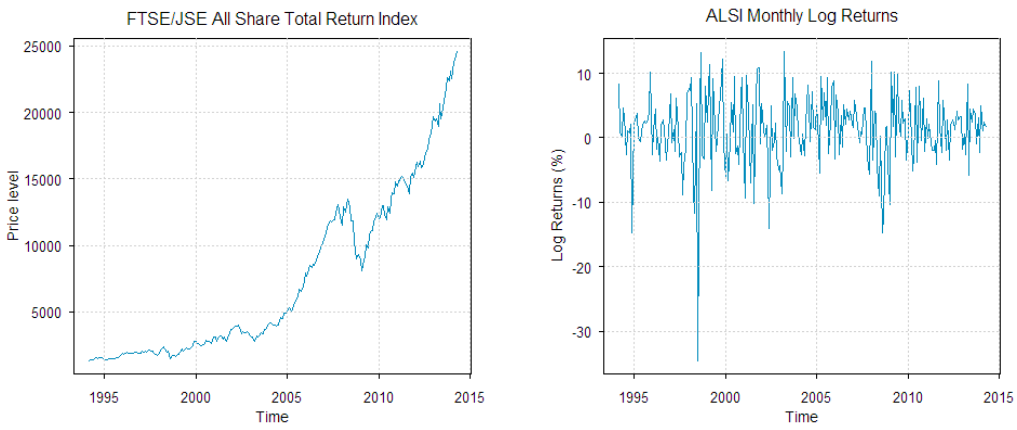


Figure 3.2: ALSI Data in Price Level and Log Return Level

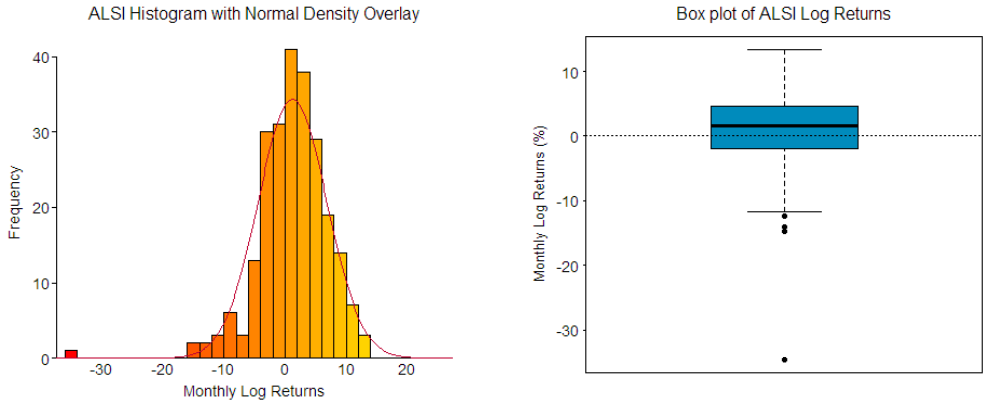


Figure 3.3: ALSI Histogram with Normal Density Overlay and Boxplot of ALSI Returns

The left panel in Figure 3.2 depicts the reconstructed FTSE/JSE All Share Total Return Index, using share price and dividend data, in raw price format. The series is clearly non-stationary, hence the transformation via log return function, displayed in the right panel. There is visual evidence of volatility clustering in the log returns. From Table 3.1, the mean of the log return series is 1.22% per month and is less than the median of 1.60%. The fluctuations are around the mean and the mean does not change with time (an indication of stationarity in the log return series). The largest loss occurred over August 1998, at -34.58%. The highest return occurred over May 2003, at 13.43%. The mean less than the median signals an asymmetric return distribution with a negative skew. Asymmetry is statistically confirmed via the skewness estimate of -1.27. The excess kurtosis estimate is 6.24, describing a heavy-tailed, or leptokurtic, distribution. The left panel of Figure 3.3 plots the normal density curve over the empirical histogram and illustrates the negative skew and leptokurtic characteristics. The negative outlier pulls the left tail beyond the reach of the fitted Gaussian density and the peakedness around the mean pushes the distribution up higher than that of the Gaussian fit. The boxplot (see Appendix) in the right panel displays an alternative view of the

negatively skewed and heavy tailed returns.

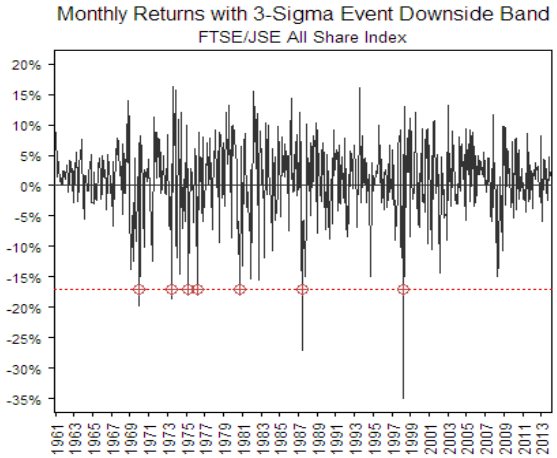


Figure 3.4: ALSI Returns with 3-Sigma Event Downside Band

The Gaussian distribution assumes that a 3-sigma lower-tail event (e.g., an asset return observation more than three standard deviations below the asset's mean return value) has approximately a 0.135% probability of occurring, or roughly one occurrence per 1000 return periods. From 30 June 1960 to 30 May 2014, the FTSE/JSE All Share Index had a monthly mean return of 1.03% and monthly standard deviation of 6.05%. These statistics imply that a negative 3-sigma event would result in a monthly loss of at least $3\sigma - \mu \approx 17.12\%$. Figure 3.4 plots the long ALSI time series, spanning 54 years, with a 3-sigma event downside band. There are 7 occurrences in the 3-sigma neighbourhood (out of 648 observations). Scaling to 1000 return periods, there may be 10.8 such downside occurrences, or roughly ten times as many empirical losses beyond the three standard deviation level as the normal distribution generates. As Xiong (2010) pointed out, underestimating such extreme risks through modelling with thin-tailed distributions (e.g., normal and lognormal) can severely impair wealth accumulation, particularly for investors approaching retirement.

The Jarque-Bera test is a formal, numerical, asymptotic, or large-sample, goodness-of-fit test which assesses whether both skewness and kurtosis in the data are consistent with a Gaussian model. The test statistic follows a chi-square distribution with two degrees of freedom under the null hypothesis of normality and is given by:

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \sim X_2^2 \,,$$

where n is the sample size, $S = \frac{E(X-\mu)^3}{\sigma^3}$, the theoretical skewness measure, $K = \frac{E(X-\mu)^4}{[E(X-\mu)^2]^2}$, the theoretical kurtosis measure and μ and σ, the theoretical mean and standard deviation, respectively, of a random variable X (e.g., returns data). The value of the JB statistic is expected to be zero under normality (i.e., where S and K jointly equal zero and three, respectively). Sample kurtosis and skewness values differing widely from three and zero, respectively, may lead to rejection of

normality. The large JB test statistic (equivalently, the small *p*-value) for the ALSI sample leads to rejection of the assumption of normality in the ALSI distribution, at the 99.9% confidence level. The quantile-quantile plot in the left panel of Figure 3.7 graphically supports the rejection of normality in the data.

A stationary process is characterised by a mean and variance that are constant over time, with the covariance between any two time points dependent only on the lag between the two points and not on the actual time at which the covariance is calculated (Gujarati & Porter, 2009). The variance, or second moment, of a stochastic return process $\{r_t\}_{t\in\mathbb{Z}}$ is given by:

$$Var(r_t) = E(r_t - \mu)(r_t - \mu) = \sigma^2, \forall t \in \mathbb{Z}, \mu = E(r_t)\ \forall t.$$

The autocovariance is the covariance of r_t with its own previous values (i.e., it determines how r_t is related to its previous values) and is given by the autocovariance function:

$$\gamma_r(s,t) = Cov(r_s, r_t) = E[r_s - E(r_s)][r_t - E(r_t)], \text{ for } s,t \in \mathbb{Z}, [= Var(r_t) \text{ when } s = t].$$

Autocorrelation is a measure of the strength of linear dependence of a variable with itself at two points in time. For the stochastic process $\{r_t\}_{t\in\mathbb{Z}}$ the autocorrelation measure is given by:

$$\rho_k(s,t) = Corr(r_s, r_t) = \frac{Cov(r_s,r_t)}{\sqrt{Var(r_s)Var(r_t)}} = \frac{\gamma_r(s,t)}{\sqrt{\gamma_r(s,s)\gamma_r(t,t)}}, \text{ for } s,t \in \mathbb{Z},\ variances < \infty \text{ and } -1 < \rho_k < 1$$

and a plot of ρ_k against k is called the autocorrelation function (ACF). The ACF is a useful qualitative tool to assess the presence of autocorrelation at different individual lags in a time series. Figure 3.5 illustrates the usefulness of ACF plots. By construction of the ACF, mutual linear dependence on other variables between the two time points (s,t) considered may distort the measure. Therefore, the sample partial ACF (PACF), which removes inter-variable linear dependence, is typically also computed and displayed. The graph in the left panel of Figure 3.5 shows the sample ACF of the ALSI price series decaying slowly over its lagged values. The slow decay indicates non-stationarity or possibly (though highly unlikely in most financial data) stationarity with long-memory dependence. The middle panel graphs the sample ACF for the ALSI return series and the graph is typical of a stationary series. Bounds, in red, are set at 0.05, the level of the test of the null hypothesis of an autocorrelation coefficient being equal to zero (i.e., insignificant autocorrelation). The right panel graphs the sample PACF. The figure is consistent with non-stationarity in a time series.

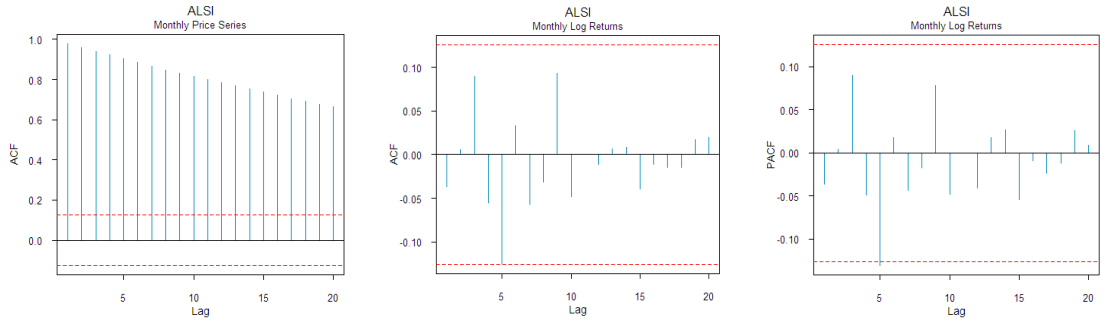


Figure 3.5: Sample ACF for the ALSI Price Series and Sample ACF and PACF for the ALSI Return Series

One of the stylised facts of financial returns is autocorrelation in absolute or squared returns. Figure 3.6 displays the sample ACF (left) and sample PACF (middle) for the absolute returns in the ALSI series, as well as the ACF for squared returns (right). There are a number of spikes breaching the test bounds, indicating autocorrelation in the absolute and squared returns. The presence of significant autocorrelations suggests that a GARCH model with lagged variances and lagged squared innovations may serve as an appropriate model for the log return series.

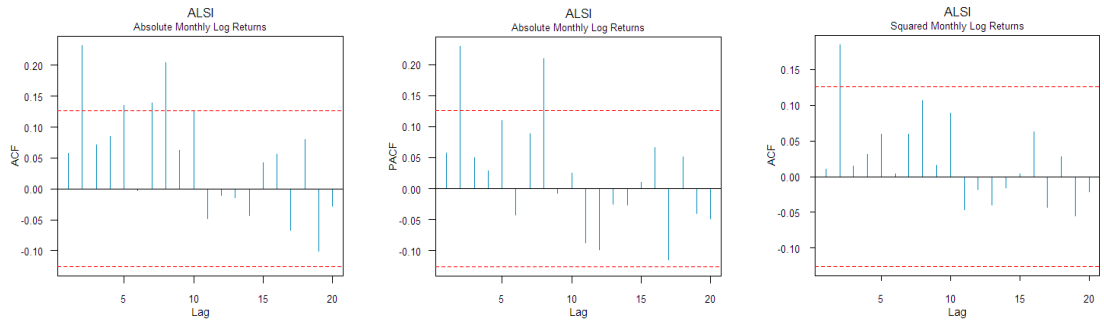


Figure 3.6: Sample ACF and PACF for the ALSI Absolute Returns and Sample ACF for the ALSI Squared Absolute Returns

The ADF test is a unit root test used to evaluate a time series for non-stationarity. The Dickey-Fuller test tests for the presence of a unit root in an AR(1) time series model representation. The ADF test augments the Dickey-Fuller test by expanding the time series model to admit testing for additional "drift" and "trend" characteristics. The ADF test is performed using the following (augmented) AR(p) model:

$$\Delta r_t = \mu + \gamma t + \delta r_{t-1} + \sum_{i=1}^{p} \beta_i \Delta r_{t-i} + \epsilon_t \,,$$

where Δ is the difference operator, r_t a time series, μ a constant (i.e., drift), γt a linear trend (i.e., trend), p is the lag order, ϵ_t a white-noise innovation, and choosing p (the number of lagged Δr_{t-i} terms to include[52]) such that the test statistic is insignificant for autocorrelation in the innovations. The ALSI log return series rejected the null of non-stationarity at the 99.9% confidence level. This is

[52] For $p = 0$ the standard Dickey-Fuller test is performed.

supported graphically in the middle panel of Figure 3.5. Similarly, drift and trend characteristics were rejected with high confidence[53]. Of the testing for each of unit root, drift and trend characteristics, all variables, except the cash variable, were stationary. The cash variable exhibited a degree of unit-root presence and a degree of drift in the returns level, but exhibited no significant degree of trend presence. This was a challenging variable to model. To manage the explosive nature of the variable in the AR representation, the AR coefficient was restricted to the bounds of $-0.9 < \delta < 0.9$ before the GARCH component of the model was fit. Further details are discussed in the GARCH-fitting section of the dissertation.

Financial data at returns level typically do not exhibit serial correlation. The Ljung-Box (LB) test[54], developed by Ljung and Box (1978), is used as a portmanteau joint test of linear dependence and is used to test a time series for autocorrelation. The test statistic is:

$$Q^* = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \sim X_m^2\,,$$

where n is the sample size, m the maximum lag length, $\hat{\rho}_k$ the autocorrelation coefficient at lag k, estimated from the sample, and where the Q^*-statistic is asymptotically distributed as a X_m^2 (i.e., chi-square with m df). When $Q^* > X_m^2$, for a given significance level, the joint null hypothesis that the first m autocorrelation coefficients are zero is rejected. The ALSI log returns show, in Table 3.1, statistically significant independence at lag one, with a *p*-value of 0.568. Table C.1 and Table C.2 in the Appendix show the *p*-values of the LB test on log returns and absolute log returns, respectively, evaluated at lags one through twelve (suggested by Daníelsson, 2011, pp. 12-14). Since data are sampled monthly, the presence of autocorrelation in returns is unlikely and, if present, may simply be spurious. However, the highly-predictable and slow-moving GMC1 (cash) and RBAS (90-day Banker's Acceptance Rate) variables, expectedly, showed significant dependence in the returns. In both cases, an AR(1) conditional mean model was used to capture lag-one autocorrelation.

As seen in Table C.2, there is evidence of statistical dependence (i.e., autocorrelation) in the absolute log returns of the variables. This is evidence of the stylised fact of volatility clustering in financial return data, also known as ARCH effects in the data[55]. A second statistical test for conditional heteroskedasticity (i.e., volatility clustering) is the Lagrange multiplier test of Engle (1982). The test is applied to demeaned data, so we construct demeaned data sets as: $\alpha_t = r_t - \mu_t$ (Tsay, 2012, pp. 182-189). The ARCH test is performed on the linear regression:

$$\alpha_t^2 = \beta_0 + \beta_1\alpha_{t-1}^2 + \cdots + \beta_m\alpha_{t-m}^2 + \epsilon_t \text{ for } t = m+1, \ldots, n$$

[53] Drift is represented by the constant μ; trend, by the constant μ and time component γt. In both "drift" and "trend" cases the series will typically be integrated of order one (i.e., contain a unit root).
[54] The LB test is preferred to the precursor Box-Pierce test due to improved small-sample properties (Gujarati & Porter, 2009, p. 754).
[55] The evidence is similar for squared log returns.

where ϵ_t denotes the error term, m is a prespecified lag term and n is the sample size. This is also a joint test of linear dependence, where the null hypothesis is $H_0: \beta_1 = \cdots = \beta_m = 0$ (i.e., no ARCH effects) and the alternative hypothesis is $H_a: \beta_i \neq 0$ for some i between 1 and m (i.e., ARCH effects). Table C.3 in the Appendix gives the *p*-values of the ARCH test on the demeaned ALSI log return series. The ALSI log returns exhibit ARCH effects at lags 2, 3 and 4 (*p*-values equal to 0.023, 0.052 and 0.104, respectively). Table C.3 shows evidence of ARCH effects in the majority of the variables and at multiple lags. The right panel of Figure 3.7 illustrates evidence of ARCH effects. The group of absolute log returns from the ALSI series that are greater than the 40th largest absolute return (of 242 total observations) is plotted against time. The 40th largest absolute value was arbitrarily chosen as the threshold and used to highlight the presence of volatility clustering in the data. As is seen in the graph, volatility clustering appears to be present in the returns series. Overall, the ARCH testing provides support for the use of a GARCH structure to model the conditional variance.

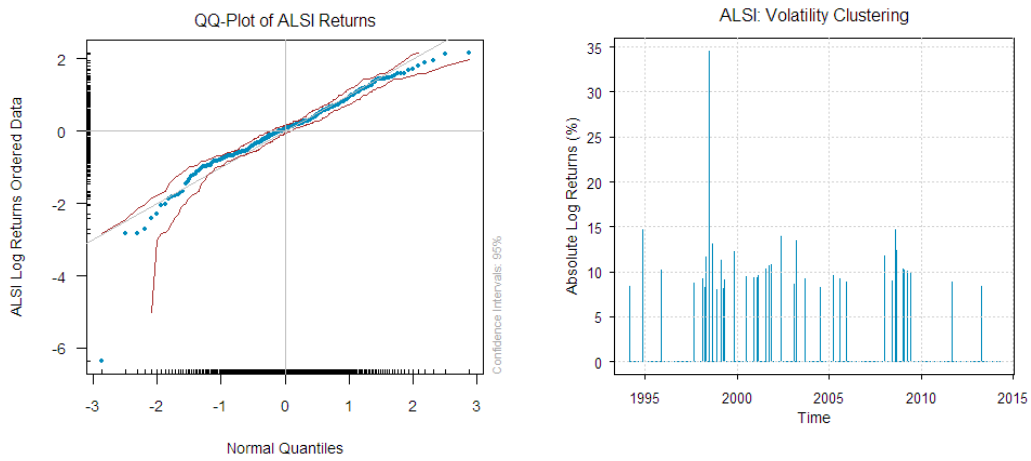


Figure 3.7: QQ Plot of the ALSI Returns and Plot of Volatility Clustering in the ALSI Absolute Returns

The left panel in Figure 3.7 shows the quantile-quantile (QQ) plot, comparing quantiles from the ALSI log return data to quantiles of a reference distribution, in this case the normal distribution[56]. If the two sample sets come from a population with the same distribution the points lie along the 45-degree line plotted in the graph. This is not the case, with the QQ plot showing negative skewness and heavy-tailedness in comparison to the Gaussian reference.

3.3 Return Filtering

Having determined support in the returns data for GARCH modelling, marginal model fitting may commence. This dissertation allows for a conditional mean equation[57] and an ARMA structure to model stationarity and any significant serial dependence (AR term) and residual influence (MA term)

[56] Here, quantiles are the standardised data values taken at regular intervals along the domain of the inverse function of the CDF of the random variable. The quantiles of the ALSI log return data are plotted on the y-axis and the theoretical standard normal quantiles are plotted on the x-axis.
[57] With a view to parsimony and practical use, the conditional mean equation is restricted to the series' unconditional mean plus an optional term for modelling ARCH-in-mean effects.

in the return series. The GARCH-family extension is included to represent the conditional variance component. The AR(*m*) component captures autocorrelation between returns and the MA(*n*) component captures shocks to the long-term average over time. Compared to linear AR models, MA structures can be used to model and forecast variables exhibiting less persistence and shorter memory (Bao, 2015).

3.3.1 Conditional Mean Modelling

The conditional mean equation is defined as follows:

$$\mu_t = \mu + \xi \sigma_t^k \,, \tag{2}$$

where μ is the unconditional mean and the second term is activated when modelling ARCH effects in the mean process either as conditional volatility $(k = 1)$ or as conditional variance $(k = 2)$.

The return series mean component is modelled with an ARMA(m, n) process, of autoregressive order m and moving average order n $(m, n \in \mathbb{Z})$, defined as follows:

$$y_t = c + \sum_{i=1}^{m} \phi_i y_{t-i} + \sum_{j=1}^{n} \theta_j \varepsilon_{t-j} + \varepsilon_t \,, \tag{3}$$

where $y_t = r_t - \mu_t$ denotes a "demeaned" (i.e., mean-centred) process, c represents a constant related to the unconditional mean[58] and the (weak white noise) innovation process, $\varepsilon_t \sim (0, \sigma^2)$, is split into two terms to admit a GARCH-family variance equation. The latter process is defined as:

$$\varepsilon_t = \sigma_t z_t \,, \tag{4}$$

where z_t is a strong white noise process, assumed to be i.i.d. with mean zero and unit variance and distributed as a standardised, known distribution, $z_t \sim iid\, \mathcal{D}_\vartheta(0,1)$, z_t and σ_t are stochastically independent and σ_t, the conditional standard deviation of y_t at time t, models the conditional volatility dynamics through the GARCH-family model. In terms of maximum likelihood estimation, applied economists often demean data in order to reduce the parameter estimation set by the μ-variable (i.e., one less parameter to estimate by choosing to model y_t, as in (2) above). The technique is simply to estimate the population mean externally using the unbiased sample mean estimator of the population mean $(\bar{r}_t = n^{-1} \sum_{t=1}^{n} r_t \approx \mu_t)$. On the other hand, Bao (2015) showed that (a) for linear AR models, demeaning data does not seriously affect estimation of non-intercept parameters and (b) for a MA(1) model, it is not always advisable to demean the data if the model parameter is of direct interest and if the parameter is moderately large (i.e., $0.5 < \theta < 1$). In case of not demeaning, the data are passed straight to the GARCH routine and assumed to be zero-mean residuals. Of the

[58] It is generally suggested to include a constant term in the mean equation in order not to force the model towards the origin.

18 variables initially estimated as y_t, 6 were found not to require demeaning based on statistical insignificance of their mean estimates when estimated in the model per equation (3)[59]. The respective 6 variables skipped the conditional mean modelling and were passed directly to the GARCH routine as r_t.

The innovation series is influenced by its conditional distributional form:

$$z_t \sim iid\ \mathcal{D}_{\vartheta}(0,1)\,, \tag{5}$$

where $\mathcal{D}$ is the standardised, conditional distribution function of the series and the optional ϑ may contain additional distribution parameters to modify the skewness and kurtosis of the distribution.

3.3.2 Conditional Volatility Modelling

Several of the more popular GARCH-family specifications for conditional volatility are considered, although the standard GARCH(1,1) model consistently performed better in terms of parsimony and cohesiveness between information criteria and log-likelihood estimates. The following section references Ghalanos (2014a), since modelling is implemented using the `rugarch` package in `R` (Ghalanos, 2014b).

3.3.2.1 The Standard GARCH Model

The Bollerslev (1986) GARCH(p,q) generalisation of Engle's (1982) ARCH model is defined as:

$$\sigma_t^2 = \left(\omega + \sum_{k=1}^{m} \zeta_k \upsilon_{kt}\right) + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2\,, \tag{6}$$

where σ_t^2 denotes a strictly positive conditional variance process, $\omega > 0$ an intercept term representing the long-term average variance value, $\alpha_i \geq 0$ for $i = 1, \ldots, q$ controlling ARCH influences (i.e., lagged squared innovations), $\beta_j \geq 0$ for $j = 1, \ldots, p$ controlling GARCH influences (i.e., lagged conditional variances), ε_t^2 the squared innovations from the mean filtration process in equation (3) and $t \in \mathbb{Z}$. The ARCH(p) process is recovered by setting $m = p = 0$. The case $p = q = 1$ is the ubiquitous GARCH(1,1) model. The model admits m external, pre-lagged regressors υ_j. For simplicity, no external regressors are used in this study, so this variable is omitted in subsequent parameterisations.

The GARCH model is designed to capture volatility clustering, which, in turn, may be quantified by the following persistence parameter:

[59] The 6 variables: USDZAR, EURUSD, GSCI, MSCI.EM.USD, RBAS and JAYC10.

$$\hat{P} = \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j \,. \qquad (7)$$

The unconditional variance of the model is given by:

$$Var(\widehat{\varepsilon_t}) = \hat{\sigma}^2 = \frac{\hat{\omega}}{1 - \hat{P}}. \qquad (8)$$

3.3.2.2 The Integrated GARCH Model

The persistent variance IGARCH model (Engle & Bollerslev, 1986) is designed for modelling time series exhibiting permanent shocks in the process, where conditional volatility is considered integrated of order one. The model is identical to the GARCH(*p*,*q*) model, but is fit setting on $\hat{P} = 1$ in equation (7).

3.3.2.3 The ARCH-in-Mean Model

Motivated by the idea that returns on a risky asset should be positively related to its risk, Engle, et al. (1987) specified the conditional mean equation to depend on some function of its conditional volatility, or conditional variance.

Modifying the conditional mean in equation (2), the GARCH-M model may be represented as:

$$\mu_t = \xi\varphi(\sigma_t) + \mathcal{E}_t \; [= \xi\varphi(\sigma_t) + z_t\sigma_t] \,, \qquad (9)$$

where ξ describes the impact of risk on the conditional mean and the ARCH-in-mean term, $\varphi(\sigma_t)$, may be specified as, for example, conditional volatility $(\varphi(\sigma_t) = \sigma_t)$, conditional variance $(\varphi(\sigma_t) = \sigma_t^2)$ or simply as $\varphi(\sigma_t) = \log(\sigma_t)$, as in Engle, et al. (1987).

3.3.2.4 The Exponential GARCH Model

The EGARCH model was introduced by Nelson (1991) to capture asymmetric responses of volatility to shocks. Conditional volatility in the model depends on both the size and sign of lagged shocks, parameterised and through weighting the innovations as follows:

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i z_{t-i} + \sum_{i=1}^{q} \gamma_i [|z_{t-i}| - E(|z_{t-i}|)] + \sum_{j=1}^{p} \beta_j \ln\left(\sigma_{t-j}^2\right) \,, \qquad (10)$$

where lagged shocks are given by $z_t = \frac{\varepsilon_{t-i}}{\sigma_{t-j}}$, α_i allows for sign effects (i.e., asymmetry influences along different lags), γ_i captures size effects (i.e., volatility impacts induced by lagged shocks) where,

as opposed to the combination of α_i and β_j capturing volatility clustering in the GARCH model, the β_j in the EGARCH model are entirely responsible for capturing volatility clustering.

3.3.2.5 The GJR-GARCH Model

Glosten, et al. (1993) specified a Boolean indicator function, where leverage coefficients are activated by past negative innovations. The model is defined as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \left(\alpha_i \varepsilon_{t-i}^2 + \gamma_i I_{t-i} \varepsilon_{t-i}^2\right) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \,, \qquad (11)$$

where γ_i captures the leverage effect induced by asymmetric volatility and I is the indicator function evaluating to 1 for $\varepsilon \leq 0$ and zero otherwise (i.e., the model uses zero as the threshold to separate the impacts of past shocks[60]).

3.3.2.6 The Asymmetric Power ARCH Model

Ding, et al. (1993) introduced the A-PARCH model to capture the leverage effect, but with a power variable estimated from the data instead of a square on the lagged innovations and lagged conditional variances. The authors built the A-PARCH specification in response to their observation that sample autocorrelations of absolute daily returns (in the S&P 500 index for the period 1928 to 1991) were larger than those of the squared daily returns (for every lag up to at least 100 lags). The model is specified as:

$$\sigma_t^{\delta} = \omega + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^{\delta} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{\delta} \,, \qquad (12)$$

where power parameter $\delta > \mathbb{R}^+$, a Box-Cox (1964) transformation of the conditional volatility, is estimated directly from the data and leverage parameter $|\gamma_i| \leq 1$. The A-PARCH specification nests a number of ARCH-based sub-models for different power and leverage parameter settings.

3.3.2.7 The Threshold GARCH Model

Zakoïan (1994) introduced the threshold GARCH (TGARCH) model, the first-order specification of which is a variant of the GJR-GARCH model. TGARCH models volatility clustering by specifying conditional volatility (as opposed to conditional variance) as a function of the lagged positive and negative parts of the innovations. The model is defined as[61]:

[60] See, for example, Tsay (2012, p. 222).
[61] This specification is a reparameterised version of Francq and Zakoïan's (2010, p. 250) TGARCH model, where, there, the leverage term γ_i in equation (13) is specified as $(\alpha_{i,+} - \alpha_{i,-})$.

$$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}\,, \qquad (13)$$

where γ_i is the coefficient of the leverage term and is activated when $\varepsilon_{t-i} < 0$.

3.3.3 GARCH-family Conditional Distributions

The shape of the conditional distribution of the innovations z_t in equation (4) is assumed to be the same for future returns. This has implications for under- or overestimating future risks in the variable. In addition the distribution is required to admit standardisation through linear transformation of the innovations (i.e., centring about the returns using the conditional mean, $[y_t - \mu_t]$, and scaling on conditional volatility, $[\varepsilon_t/\sigma_t]$). After appropriate transformation, standardised innovations $z_t = \varepsilon_t/\sigma_t$ are assumed to be mean zero and unit variance and are modelled with (the now scaled version of) the same conditional distribution of y_t. Several models for the innovations z_t are considered.

3.3.3.1 The Normal Distribution

The default choice for innovation distribution $\mathcal{D}$ is the Gaussian probability function. A normally distributed random variable X has density given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,, \qquad (14)$$

where mean μ and variance σ^2 may both be time-varying (i.e., evaluated conditionally).

The standardised normal probability function for the innovations is given by:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,. \qquad (15)$$

3.3.3.2 The Student *t* Distribution

The standardised Student *t* distribution has density given by:

$$f(z|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{z^2}{(\nu-2)}\right)^{-\left(\frac{\nu+1}{2}\right)}\,, \qquad (16)$$

where $\nu > 2$ is the shape parameter measuring tail thickness and $\Gamma(\nu) = \int_0^\infty e^{-x} x^{\nu-1}\,dx$ is the gamma function.

3.3.3.3 The Generalised Error Distribution

The standardised GED has density given by:

$$f(z|\kappa) = \frac{\kappa}{\lambda_\kappa 2^{1+\kappa^{-1}}\Gamma(\kappa^{-1})} e^{-\frac{1}{2}\left|\frac{z}{\lambda_\kappa}\right|^\kappa}, \quad (17)$$

$$\lambda_\kappa = \left(2^{-2/\kappa}\frac{\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})}\right)^{1/2},$$

where $0 < \kappa \leq \infty$ is the shape parameter.

3.3.3.4 The Skewed Distributions

Fernández and Steel (1998) specified a general density function to admit skewness in any unimodal and symmetric distribution by changing the shape on each side of the mode. The density of a random variable z, given a shape parameter ξ, is hereby given as:

$$f(z|\xi) = \frac{2}{\xi + \xi^{-1}}\left[f(\xi z)H(-z) + f(\xi^{-1}z)H(z)\right], \quad (18)$$

where $0 < \xi < \infty$ describes the degree of asymmetry[62] and $H(z) = [1 + \text{sign}(z)]/2$ is the Heaviside unit step function. When $\xi = 1$, $f(z|\xi = 1) = f(z)$, yielding a symmetric distribution.

3.3.3.4.1 The Skewed Normal Distribution

The standardised skewed normal density[63], denoted by SN$(0,1,\xi)$, is given by:

$$f(z|\xi) = \frac{2}{(\xi + \xi^{-1})}\frac{1}{\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(z\xi^{-\text{sign}(z)}\right)^2\right]}, \quad (19)$$

where $\xi > 0$ is the skewness parameter, determining the direction and intensity of the skewness, and $\text{sign}(z)$ is the sign function evaluating to 1 if $z > 0$, -1 if $z < 0$ and 0 if $z = 0$ (i.e., the skewing mechanism scales $f(z)$ differently for negative and positive values).

3.3.3.4.2 The Skewed Student *t* Distribution

The standardised skewed Student *t* density[64], denoted by SKST$(0,1,\xi,\nu)$, is given by:

[62] $\xi < 1$ produces left skewness; $\xi > 1$ produces right skewness.
[63] See, for example, Li, et al. (2013).
[64] See, for example, Bao, Lee and Saltoğlu (2007) and Palmitesta and Provasi (2006).

$$f(z|\xi,\nu)=\left(\frac{2}{\xi+\xi^{-1}}\right)\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}\left[1+\frac{\xi^{-2\,\mathrm{sign}(z)}z^2}{\nu-2}\right]^{-\frac{(\nu+1)}{2}}, \quad (20)$$

where $\xi>0$ and $\nu>2$. If $\xi=1$, the distribution is symmetric about zero, with zero mean and unit variance.

3.3.3.4.3 The Skewed Generalised Error Distribution

The standardised skewed GED[65], denoted by SGED$(0,1,\xi,\kappa)$, is given by:

$$f(z|\xi,\kappa)=\left(\frac{2}{\xi+\xi^{-1}}\right)\Gamma^{\frac{1}{2}}\left(\frac{3}{2\kappa}\right)\Gamma^{-\frac{3}{2}}\left(\frac{1}{2\kappa}\right)\kappa e^{\left[-c^{\kappa}\xi^{-2\kappa\,\mathrm{sign}(z)}|z|^{2\kappa}\right]}, \quad (21)$$

$$c=\frac{\Gamma\left(\frac{3}{2\kappa}\right)}{\Gamma\left(\frac{1}{2\kappa}\right)},$$

where $\xi>0$ and $\kappa>0$. The distribution collapses to the skew normal when $\kappa=1$ and becomes the skew Laplace distribution when $\xi\neq1$ and $\kappa=1/2$.

3.3.3.5 The Generalised Hyperbolic Distribution

The generalised hyperbolic distribution, denoted by GHYP$(\lambda,\alpha,\beta,\delta,\mu)$, is given by:

$$f(x|\lambda,\alpha,\beta,\delta,\mu)=\frac{\left(\frac{\gamma}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\mathrm{K}_{\lambda_{\mathrm{B}}}(\delta\gamma)}\frac{\mathrm{K}_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\frac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha}\right)^{\frac{1}{2}-\lambda}}e^{\beta(x-\mu)}, \quad (22)$$

$$\gamma^2=\alpha^2-\beta^2,$$

where $x\in\mathbb{R}$, $\mathrm{K}_{\lambda_{\mathrm{B}}}(x)=\frac{1}{2}\int_0^{\infty}u^{\lambda_{\mathrm{B}}-1}e^{-x(u+u^{-1})/2}\,du$, $x>0$, is the modified Bessel function of the third kind[66] (with index parameter $\lambda_{\mathrm{B}}\in\mathbb{R}$), $\mu\in\mathbb{R}$ is the location parameter, $\delta\geq0$ the scale parameter, $\beta\in[-\alpha,\alpha]$ the skewness parameter, $\alpha\geq0$ the tail parameter and $\lambda\in\mathbb{R}$ the shape parameter.

The distribution GHYP$(x|\lambda,\alpha,\beta,\delta,\mu)$ is reparameterised to obtain the standardised GHYP$(z|\zeta,\rho)$ distribution. This is done by estimating $\zeta=\delta\sqrt{\alpha^2-\beta^2}$ and $\rho=\beta/\alpha$, setting the mean equal to zero and variance equal to one and demeaning and scaling the random variable x to obtain the standardised variable z (Ghalanos, Rossi & Urga, 2015).

[65] See, for example, Bao, Lee and Saltoğlu (2004) and Palmitesta and Provasi (2006).
[66] See, for example, Bibby and Sørensen (2003, p. 243) and Paolella (2007, pp. 315-320).

3.3.3.6 The Normal Inverse Gaussian Distribution

The normal inverse Gaussian (NIG) distribution is the sub-class distribution nested in the GHYP$(\lambda, \alpha, \beta, \delta, \mu)$ distribution and obtained when $\lambda = -1/2$. The density[67], denoted by NIG$(\alpha, \beta, \delta, \mu)$, is given by:

$$f(x|\alpha,\beta,\delta,\mu) = \frac{\alpha\delta}{\pi} e^{\delta\gamma} \frac{\mathrm{K}_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\sqrt{\delta^2+(x-\mu)^2}} e^{\beta(x-\mu)} \qquad (23)$$

where $x, \mu \in \mathbb{R},\ \alpha, \delta \geq 0,\ \beta \in (-\alpha, \alpha)$ and $\gamma^2 = \alpha^2 - \beta^2$. Standardisation of x is done using the distribution's associated expressions for the mean and variance.

3.3.3.7 The Generalised Hyperbolic Skew Student *t* Distribution

Aas and Haff (2006) specified the GHST distribution, a limiting case of the GHYP distribution, to admit one tail to decay at a polynomial rate and one at an exponential rate. The limiting case arises when $\alpha \to |\beta|$ and $\lambda = -\nu/2$, where ν is the shape (df) parameter of the Student *t* distribution. In terms of scaling, variance is only finite for $\nu > 4$. Skewness exists only for $\nu > 6$ and kurtosis only for $\nu > 8$ (i.e., the n-th moment exists when $\nu > 2n$). The GHST density[68] is given by:

$$f(x) = \frac{\delta^{\nu}}{\sqrt{\pi}2^{\frac{(\nu-1)}{2}}\Gamma\left(\frac{\nu}{2}\right)} \frac{\mathrm{K}_{\frac{(\nu+1)}{2}}\left(|\beta|\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\frac{\sqrt{\delta^2+(x-\mu)^2}}{|\beta|}\right)^{\frac{(\nu+1)}{2}}} e^{\beta(x-\mu)}, \qquad (24)$$

where $\alpha \to |\beta|,\ \lambda = -\nu/2 \Rightarrow \nu = -2\lambda,\ \nu > 0,\ \gamma = 0,\ \delta > 0$ and $\mu, \beta \in \mathbb{R}$. As with the previous generalised hyperbolic distributions, standardisation of x is done using the distribution's associated expressions for the mean and variance.

3.3.3.8 The Reparameterised Johnson's SU Distribution

Johnson (1949) introduced an "unbounded" distribution, or "system of frequency curves", using a particular transformation on a random variable. The unbounded distribution, called "the system S_U", is known commonly as Johnson's SU distribution and is appropriate for leptokurtic data. The distribution, denoted by $JSU_o(x|\mu, \sigma, \nu, \tau)$, accommodates the mean-variance-skewness-kurtosis parameter hyperspace, but with the mean and variance specified as functions of the scale (σ),

[67] See, for example, Bibby and Sørensen (2003, p. 218) and Paolella (2007, p. 325).
[68] See also, for example, Bibby and Sørensen (2003, p. 221) and Paolella (2007, p. 322).

skewness (ν) and kurtosis (τ) parameters[69] (i.e., $E(X) \neq \mu$ and $Var(X) \neq \sigma$). The reparameterisation of the original Johnson SU distribution (Rigby & Stasinopoulos, 2005, 2010), denoted by JSU$(x|\mu,\sigma,\nu,\tau)$, has been done so to set $E(X) = \mu$ and $Var(X) = \sigma$ for all values of the skew (ν) and shape (τ) parameters. The JSU density is given by:

$$f(x|\mu,\sigma,\nu,\tau) = \frac{\tau}{c\sigma}\frac{1}{(r^2+1)^{1/2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}, \qquad (25)$$

$$z = -\nu + \tau \sinh^{-1}(r) = -\nu + \tau \log\left(r + \sqrt{r^2+1}\right),$$

$$r = \frac{x - \left(\mu + c\sigma\omega^{\frac{1}{2}}\sinh\Omega\right)}{c\sigma},$$

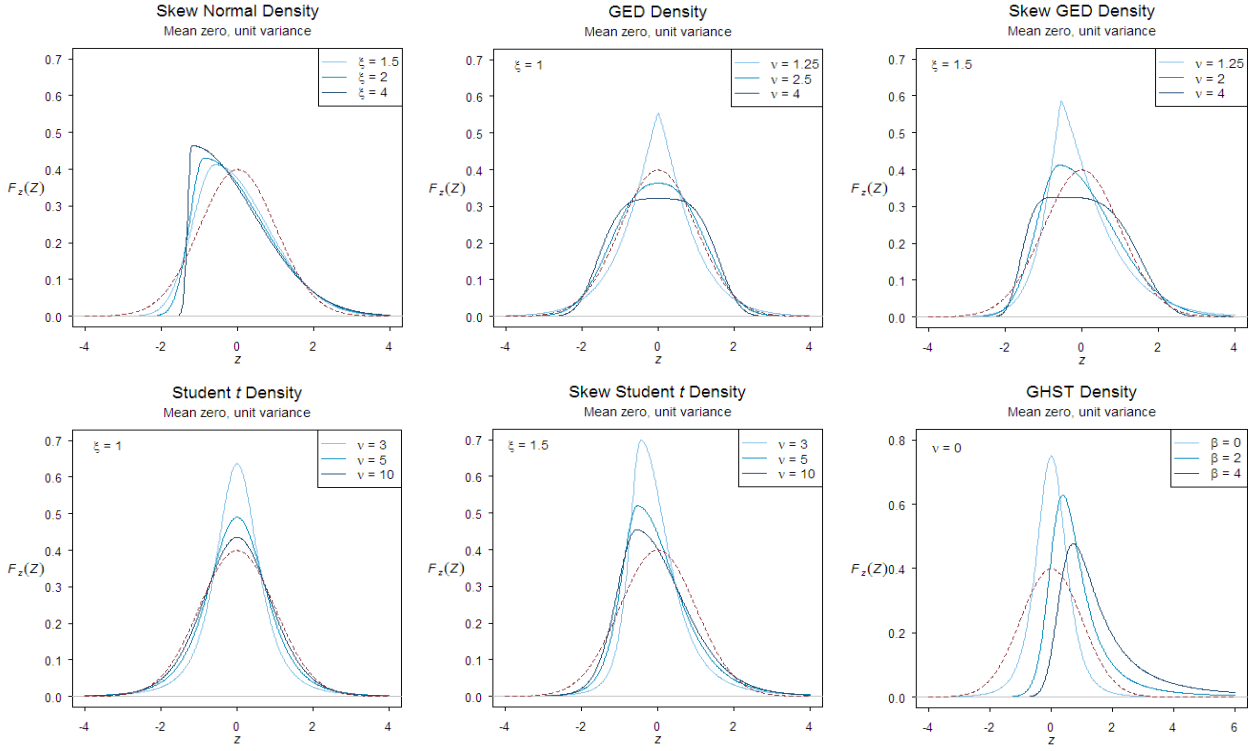
$$c = \left(\frac{1}{2}(\omega-1)[\omega\cosh(2\Omega)+1]\right)^{-1/2},$$

$$\omega = e^{(1/\tau^2)} \text{ and } \Omega = -\nu/\tau,$$

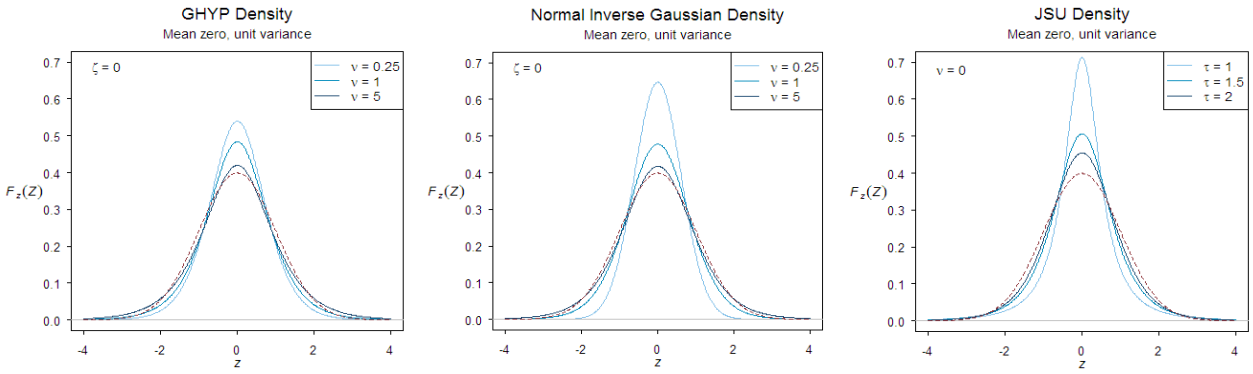
where $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$, $\tau > 0$ and $Z \sim N(0,1)$.

Figure 3.8 plots the standardised versions of the skew normal, GED, skew GED, Student *t*, skew Student *t*, GHST, GHYP ($\lambda = 1, \zeta = 0$), NIG ($\zeta = 0$) and JSU density functions for different parameter settings. The dashed lines in each plot mark the standard normal density function.

Figure 3.8: Standardised Conditional Distributions Used in the GARCH Models



[69] $E(X) = \mu - \sigma\omega^{\frac{1}{2}}\sinh(\nu/\tau)$ and $Var(X) = \sigma^2\frac{1}{2}(\omega-1)[\omega\cosh(2\nu/\tau)+1]$, where $\omega = e^{(1/\tau^2)}$ (Rigby & Stasinopoulos, 2010, p. 205).



3.3.4 ARMA-GARCH Model Estimation

The parameters of the ARMA-GARCH-family model are estimated jointly via quasi maximum likelihood estimation (QMLE), based on the Gaussian log-likelihood function. Assume that the time series $\{r_t\}$ from equation (3) is generated by a strictly stationary, non-anticipative solution of the ARMA(m,n)-GARCH(p,q) model[70], where the orders m,n,p,q are assumed known. The vector of parameters is given by:

$$\varphi = (\varrho^T, \psi^T) = (\mu_t, c, \phi_1, \dots, \phi_m, \theta_1, \dots, \theta_n, \psi^T)^T ,$$
$$\psi = \left(\psi_t, \dots, \psi_{p+q+1}\right)^T = \left(\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p\right)^T .$$

For $q \geq n$ (referring to the ARMA-GARCH lag lengths) and for any ϱ and ψ, the values of conditional moments $\tilde{\varepsilon}_t(\varrho)$, first, and $\tilde{\sigma}_t^2(\varphi)$, second, may be approximated[71], in the aforementioned order, by:

$$\begin{cases} \tilde{\varepsilon}_t = \tilde{\varepsilon}_t(\varrho) = y_t - c - \sum_{i=1}^{m} \phi_i y_{t-i} - \sum_{j=1}^{n} \theta_j \tilde{\varepsilon}_{t-j} \\ \tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\varphi) = \omega + \sum_{i=1}^{q} \alpha_i \tilde{\varepsilon}_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2 , \end{cases}$$

where the $\tilde{\sigma}_t^2$ are defined recursively for $t > 1$ and fed into the conditional Gaussian quasi-likelihood function, given by:

$$\mathcal{L}_n(\varphi) = \mathcal{L}_n(\varphi \in \Phi | \tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} e^{\left(-\frac{\tilde{\varepsilon}_t}{2\tilde{\sigma}_t^2}\right)} .$$

Whereas a joint probability density function is a function of the data given a set of parameters, $f(y_1, y_2, \dots, y_n|\Phi)$, maximum likelihood uses a likelihood function as a function of the parameters given a set of data, $\mathcal{L}_n(\Phi|y_1, y_2, \dots, y_n)$. To write the likelihood of the model, a distribution $\mathcal{D}_\vartheta$ must

[70] QMLE is not the focus of the dissertation, therefore estimation is illustrated for strictly the GARCH(p,q) model and without external regressors v_j.
[71] This section references Francq and Zakoïan (2010, pp. 150-151).

be specified for the i.i.d. variables z_t. However, QMLE does not make any assumption on the distribution $\mathcal{D}_\vartheta$. The Gaussian quasi-likelihood has been shown to coincide with the likelihood of distribution $\mathcal{D}_\vartheta$ when the z_t are distributed as standard Gaussian. The log of this function is maximised with respect to the parameters, essentially the same as finding the mode of the distribution.

The Gaussian log-likelihood is given by:

$$\tilde{\mathbf{I}}_n(\varphi) = n^{-1}\sum_{t=1}^{n}\tilde{\ell}_t\,, \qquad \tilde{\ell}_t = \tilde{\ell}_t(\phi) = \frac{\tilde{\varepsilon}_t^2(\varrho)}{\tilde{\sigma}_t^2(\varphi)} + \log\tilde{\sigma}_t^2(\varphi) = -2\log\big(\mathcal{L}_n(\varphi)\big)\,, \tag{26}$$

where φ is a subset of parameter space Φ.

A QMLE of φ is defined as any measurable solution $\hat{\varphi}_n$ to the following equation:

$$\hat{\varphi}_n = \arg\min_{\varphi\in\Phi}\tilde{\mathbf{I}}_n(\varphi) \equiv \arg\max_{\varphi\in\Phi}\,\mathcal{L}_n(\varphi)\,. \tag{27}$$

Put another way (Amendola & Francq, 2009, pp. 401-402), which highlights the importance of "standardisable" distributions in time series analysis, parameters are estimated as per equation (27), but with $\tilde{\ell}_t = [(Y_t - \tilde{m}_t)^2/\tilde{\sigma}_t^2 + \log\tilde{\sigma}_t^2]$, where the model is of the typical form $Y_t = m_t(\Phi) + \sigma_t(\Phi)z_t$, $z_t \sim iid\ \mathcal{D}_\vartheta(0,1)$. The parameter space is Φ, with $\tilde{m}_t(\Phi) \coloneqq \tilde{m}_t = E_\Phi(Y_t|Y_{t-1},\dots,Y_1)$ and $\tilde{\sigma}_t^2(\Phi) \coloneqq \tilde{\sigma}_t^2 = Var_\Phi(Y_t|Y_{t-1},\dots,Y_1)$.

3.3.5 ARMA-GARCH Model Selection

The procedure used to select the best fitting model per variable was initially implemented as follows:

- Fit the univariate time series with a mean and variance equation, the mean equation to be modelled under the ARMA(m,n) specification for $m,n \in \{0,1,2,3\}$.
- Fit each of the 16 ARMA(*m*,*n*) specifications with a variance equation from one of the following: GARCH(1,1), IGARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1), A-PARCH(1,1), TGARCH(1,1) and GARCH-In-Mean.
- Pair each composite structure with a conditional distribution: normal, skew normal, Student *t*, skew Student *t*, GED, skew GED, GHYP, NIG, GHST and JSU.
- Rank each model according to smallest Akaike and Bayesian information criterion scores (AIC and BIC, respectively) and largest log-likelihood (LLH) values. Search for cohesiveness, or agreement, in the ranking criteria.

The AIC and BIC statistics are calculated, respectively, as:

$$AIC = -\frac{2}{n}\ln(likelihood_{ML}) + \frac{2}{n}\times(number\ of\ parameters)$$

$$BIC = -\frac{2}{n}\ln(likelihood_{ML}) + \frac{2\log(n)}{n} \times (number\ of\ parameters)\ .$$

The AIC imposes a penalty equal to a sample size-related scaling of the number of model parameters on the log-likelihood value at the maximum. The BIC tends to select simpler models by imposing a greater penalty related to the magnitude of a function of the sample size. The increase in the "complexity penalty" between the AIC to BIC increases the tendency of the BIC to select simpler models. The BIC tends towards underparameterisation, a problem that may be a benefit when the purpose is true out-of-sample forecasting. Simpler models often perform better out-of-sample than do overparameterised models. The LLH has no complexity penalty, with emphasis placed on minimising innovation variances in-sample and tending towards overfit models with more parameters (which increases the risk of fitting noise and thereby forecasting poorly out-of-sample). If there is no cohesiveness among ranking criteria, the model which minimises the BIC is selected.

Meaningful discrepancies were found when cross-validating model fits produced by two different R software packages[72]. Simplifying the ARMA(m,n) specification to $m,n = 0$ or restricting the estimate bounds when m or $n = 1$ (e.g., $-0.9 \leq \phi_1 \leq 0.9$) produced concordant estimates. Fitting ARMA(0,0) structures as the mean equation implies there is no measurable or meaningful difference in the residuals from raw log returns and the residuals from respective fitted models. Since the difference between the raw return residuals and the fitted residuals depends on the mean equation, not the variance equation, this phenomenon may be expected. ARMA(0,0) models, therefore, do not explain anything about the level of return; rather, they are a preliminary pass-through filter aimed at presenting a weak white noise innovation series to the GARCH process. Capturing the variation in the conditional standard deviation is the modelling priority in this section of the dissertation. Further, for any given variable, the return equation has no explanatory power (returns are unpredictable), whereas the volatility equation does have explanatory power (volatility may be forecast with a degree of confidence[73]).

In terms of data granularity at a "coarse" monthly sampling frequency, it was found that the IGARCH and the larger-parameterised asymmetric GARCH models were failing to compete against the simpler GARCH(1,1) models. GARCH-family models tend to suit data sampled at daily, or higher, frequencies. The twenty years of monthly data used in the dissertation amounted to 242 data points per series. For sample sizes of $n < 700$, Ng and Lam (2006) showed that there may be two, or more, solutions to the conventional GARCH(1,1) model estimated by maximum likelihood. The authors recommended using samples of size $n > 1000$ for conventional GARCH(1,1) fitting. In light of the

[72] The `rugarch` (Ghalanos, 2014b) and `fGarch` (Wuertz & Chalabi, 2009) packages output conflicting parameter estimates in the conditional mean equations. In discussing the problem with the developer and maintainer of the `rugarch` package, the conflicting results were attributed to (1) spurious regression resulting from (2) over-parameterising the model to too little data.

[73] See, for example, Brownlees, Engle and Kelly (2011).

above, the procedure used to select the best fitting model per variable was adjusted to the following:

- Fit the univariate time series with a mean and variance equation, the mean equation to be modelled under the ARMA(m,n) specification, for $m,n \in \{0,1\}$, with optional restriction on the AR(1) coefficient when the tendency towards first-differencing was strong[74].
- Fit each of the ARMA(*m*,*n*) specifications with a variance equation from one of the following: GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) and GARCH-In-Mean.
- Pair each composite structure with a conditional distribution: normal, skew normal, Student *t*, skew Student *t*, GED, skew GED, GHYP, NIG, GHST and JSU.

Model fits were ranked according to cohesiveness among the AIC, BIC and LLH ranking criteria.

Continuing with the theme of analysing a representative variable, results for the ALSI return filtering follows.

The ARMA(0,0)-GARCH(1,1) model, fit to demeaned data[75], was selected. The conditional distribution for the ALSI GARCH model is the Student *t* distribution with shape parameter (robustly[76]) significant at the 5% level. Table 3.2 presents the parameter estimates and associated standard and robustified *p*-values.

Table 3.2: Optimal Parameter Estimates for the ALSI ARMA-GARCH Model

	Estimate	*p*-value (standard errors)	*p*-value (robust errors)
Mu	0.0159***	0.0000	0.0000
Omega	0.0003	0.2028	0.2243
Alpha_1	0.2342	0.1109	0.2666
Beta_1	0.6839***	0.0001	0.0031
Shape	7.2137***	0.0092	0.0374

*p < 0.1, **p < 0.05, ***p < 0.01, n = 242, based on standard errors.

Residuals from the fitted model are standardised and checked for model adequacy. They should form an approximately i.i.d. series, though with heavy tails, as well as an assumed realisation of a strict white noise process for use in EVT modelling. Standardised residual series are calculated as follows:

$$(z_{t-n+1}, \dots, z_t) = \left(\frac{r_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \dots, \frac{r_t - \hat{\mu}_t}{\hat{\sigma}_t}\right).$$

[74] The GMC1 Cash variable required a first-differencing (i.e., setting $\phi_1 = 1$ in equation (3)) to improve the modelling. However, doing so alienated the variable from the (un-differenced) set of portfolio variables in terms of modelling its co-relationship in the copula model. It was therefore decided to bound the AR coefficient and proceed with this caveat in mind. In addition, cash is a stalwart asset in a portfolio and its return series is not expected to co-vary meaningfully with other assets. The latter implies that, in terms of copula modelling of co-variation, cash may be considered a "non-critical" variable.
[75] That is, the estimated mean μ (with *p*-value approximately zero) was significant enough to include in the mean equation.
[76] Robust standard errors are based on asymptotically valid confidence intervals per White (1982).

The weighted Ljung-Box test (Fisher & Gallagher, 2012) for serial correlation, or independence, was applied to standardised residuals and standardised squared residuals produced in-sample from the fitted model[77]. The former (with df set to the number of AR and MA parameters) is used to test for ARMA effects and the latter (with df set to the number of ARCH and GARCH parameters) is used to test for ARCH/GARCH effects. Lag lengths vary as a function of the number of GARCH parameters in the model. If the null of no serial correlation in the data is rejected, then there is statistical evidence of the model not capturing ARMA and/or ARCH/GARCH effects (i.e., the model is misspecified). Table 3.3 reports *t*-values of the test with *p*-values in parentheses. The high *p*-values indicate little chance of serial correlation in the standardised residuals and standardised squared residuals at the lags tested.

Table 3.3: Weighted LB Test Results on Standardised Residuals and Standardised Squared Residuals

Lag	1	$2(p+q)+(p+q)-1$	$4(p+q)+(p+q)-1$
Standardised Residuals	0.0038 (0.9506)	0.0845 (0.9306)	0.8014 (0.9027)
Standardised Squared Residuals	0.5764 (0.4477)	1.0507 (0.8481)	1.6469 (0.9427)

*p < 0.1, **p < 0.05, ***p < 0.01, n = 242.

The weighted ARCH LM test for the null of an adequately fit ARCH process (i.e., no ARCH effects) was applied to the residuals of the fitted model (Fisher & Gallagher, 2012)[78]. Table 3.4 provides a summary report. The high *p*-values indicate statistically significant removal of ARCH effects by the model, at the respective lags. Test statistics are shown with *p*-values in parentheses.

Table 3.4: Weighted ARCH LM Test Results on Model Residuals

ARCH Lag	3	5	7
	0.3938 (0.5303)	0.7366 (0.8123)	1.0849 (0.8994)

*p < 0.1, **p < 0.05, ***p < 0.01, n = 242.

The Nyblom (1989) stability test evaluates in-sample individual and joint parameter stability. The test is a Lagrange multiplier test based on maximum likelihood scores for the null hypothesis of parameter stability (against a martingale process alternative[79]). The test statistic for the individual parameter Nyblom test is given by:

$$L_i = \frac{1}{n} V_i^{-1} \sum_{t=1}^{n} S_{it} \,,$$

where $i = 1, \dots, (k+1)$ is the number of exogenous variables, $V_i = \sum_{t=1}^{n} f_{it}^2 \;\; \forall i = 1, \dots, (k+1)$, the

[77] The use and interpretation of the *weighted* LB test statistics remain unchanged.
[78] The use and interpretation of the *weighted* ARCH LM test statistics remain unchanged.
[79] The martingale specification can cover a number of types of departure from parameter constancy, such as a random walk (i.e., slow random variation) or a change-point model (e.g., a single jump at an unknown time point in the series).

sum of f_t over time is calculated as $S_{it} = \sum_{t=1}^{n} f_{it}$ with $f_{it} = \begin{cases} x_{it}\hat{\varepsilon}_t\,, & i = 1, \ldots, k \\ \hat{\varepsilon}_t^2 - \hat{\sigma}\,, & i = k+1 \end{cases}$ (where f_{it} approximating zero implies parameter stability).

The test statistic for the joint parameter Nyblom test is given by:

$$L = \frac{1}{n\hat{\sigma}^2}\, tr\left[V^{-1}\sum_{t=1}^{n} S_t S_t'\right],$$

where H_0: parameters are constant $\Leftrightarrow \sigma_{\varepsilon_i}^2 = 0$ for all i, against $H_a: \sigma_{\varepsilon_i}^2 > 0$ for some i, $\sigma_{\varepsilon_i}^2$ is the variance at time t for parameter i, cumulative MLE scores $S_t = \sum_{t=1}^{n} f_t$, with $f_t = x_t\hat{\varepsilon}_t$, and $V = n^{-1}X'X$ for n observations and X is a matrix of variables from the regression model. For both tests, the null is rejected if the test statistic is greater than the respective asymptotic critical value (provided by the software).

From Table 3.5, it can be seen that, jointly, the model parameters are stable over the 20 years of monthly data. The mean, alpha$_1$ (i.e., ARCH) and beta$_1$ (i.e., GARCH) parameters exhibit stability over the sample period, while the omega (related to unconditional variance) and shape (related to tail behaviour) parameters exhibit instability at the 10% levels, at the monthly sampling frequency.

Table 3.5: Nyblom Test Results for Individual and Joint Parameter Stability

	Test Statistic	Asymptotic Critical Values		
		10%	5%	1%
Joint	1.1404	1.28	1.47	1.88
Mu	0.1360	0.35	0.47	0.75
Omega	0.3599*	0.35	0.47	0.75
Alpha$_1$	0.1144	0.35	0.47	0.75
Beta$_1$	0.2603	0.35	0.47	0.75
Shape	0.4077*	0.35	0.47	0.75

Decision rule: Reject H_0 if test statistic > critical value. Significance levels: *10%, **5%, ***1%, *n* = 242.

The sign bias test of Engle and Ng (1993) tests for different leverage effects in the standardised squared residuals and, thereby, for possible GARCH model misspecification. The test is administered through a regression of the standardised squared residuals from the fitted model on lagged positive and negative shocks (i.e., good news and bad news impacts), as follows:

$$\hat{z}_t^2 = c_0 + c_1 I_{\hat{\varepsilon}_{t-1}<0} + c_2 I_{\hat{\varepsilon}_{t-1}<0}\hat{\varepsilon}_{t-1} + c_3 I_{\hat{\varepsilon}_{t-1}\geq 0}\hat{\varepsilon}_{t-1} + \mu_t\,,$$

where $H_0: c_i = 0$ (for $i = 1,2.3$) and jointly $H_0: c_1 = c_2 = c_3 = 0$, evaluator function I assumes one if true (zero otherwise) and $\hat{\varepsilon}_t$ are the estimated residuals from the GARCH model. Rejection of the null hypothesis implies that an asymmetric GARCH-type structure may better model the presence of leverage effects.

Table 3.6: Sign Bias Test Results for Leverage Effects in the ALSI Model

	t-value	*p*-value
Sign Bias ($H_0: c_1 = 0$)	0.3264	0.7444
Negative Sign Bias ($H_0: c_2 = 0$)	0.2380	0.8121
Positive Sign Bias ($H_0: c_3 = 0$)	1.0342	0.3021
Joint Effect ($H_0: c_1 = c_2 = c_3 = 0$)	1.2010	0.7528

*p < 0.1, **p < 0.05, ***p < 0.01, n = 242.

As can be seen in Table 3.6, there is no statistical evidence (individually or jointly) supporting any sign or presence of asymmetric leverage effects in the ALSI model residuals.

The Pearson goodness-of-fit test is used to compare the empirical distribution of the standardised residuals to the theoretical distribution of the chosen conditional density. The standard test is adjusted, per Palm (1996), to admit non-i.i.d. observations: standardised residuals are re-classified based on magnitude (as opposed to value) and test statistics calculated as a function of the probability of observing a value smaller than the standardised residual. The default number of bins in the histograms that are tested are 20, 30, 40 and 50. The null is that the data follow the given distribution.

As can be seen in Table 3.7, the goodness-of-fit test fails to reject the Student *t* distribution as appropriate in the ALSI model (i.e., the chosen conditional density appears appropriate in modelling the standardised residuals).

Table 3.7: Adjusted Pearson Goodness-of-Fit Test Results of the Estimated Conditional Density in the ALSI Model

Number of Bins	Test Statistic	*p*-value
20	12.88	0.8449
30	23.04	0.7746
40	23.79	0.9738
50	32.79	0.9738

*p < 0.1, **p < 0.05, ***p < 0.01, n = 242.

In addition to statistical evaluation, results of the return filtering may be evaluated graphically. The left panel of Figure 3.9 displays the ALSI sample series plotted with upper and lower estimated 2-sigma conditional standard deviations superimposed. The right panel plots the estimated conditional standard deviation series $\hat{\sigma}_t$ superimposed over the absolute value of the log return series $|r_t|$.

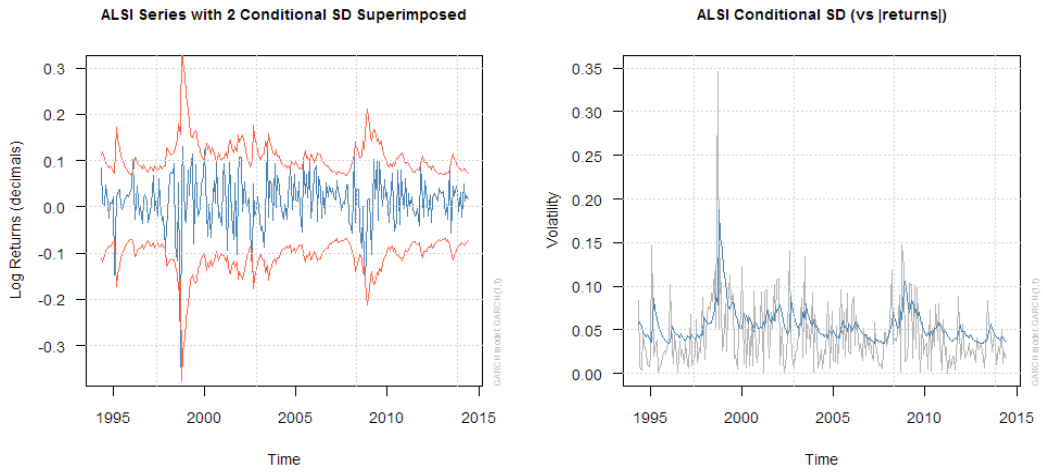


Figure 3.9: ALSI Returns with 2-Sigma Conditional Volatility Overlay and ALSI Absolute Returns with Conditional Volatility Overlay

Figure 3.10 plots the sample ACFs of the standardised residuals $\hat{\varepsilon}_t$ and squared standardised residuals $\hat{\varepsilon}_t^2$, respectively. There are no signs of autocorrelation in either series.

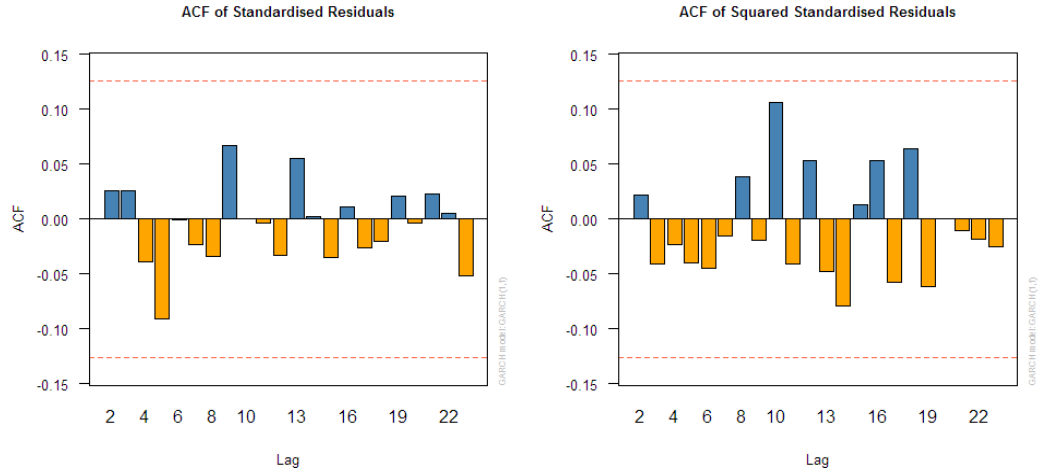


Figure 3.10: Sample ACF for Standardised Residuals and Standardised Squared Residuals

The left panel of Figure 3.11 displays the empirical histogram of the standardised residuals $\hat{\varepsilon}_t$ with fitted normal and Student *t* distributions superimposed. The Student *t* distribution is fit as this was the chosen conditional density for the ALSI model. The right panel plots the Student *t* QQ plot of the standardised residuals $\hat{\varepsilon}_t$.

The QQ plot illustrates graphically the prudence of using Extreme Value Theory as a subsequent and intermediate step before copula modelling. While the estimated GARCH model, with the selected fat-tailed density, is able to capture the bulk of the excess kurtosis found in the return series, there is still meaningful deviation observed in the tails of the distribution. The curve down at the left of the QQ plot is evidence of a heavier lower tail than the reference distribution. Contrastingly, the curve down at the right of the QQ plot is evidence of lighter-than-Student *t* upper tail density. EVT provides a formal framework for studying the statistical behaviour expected in the extreme areas of

the distribution. The EVT framework further admits asymmetric modelling through the separate fitting of each tail.

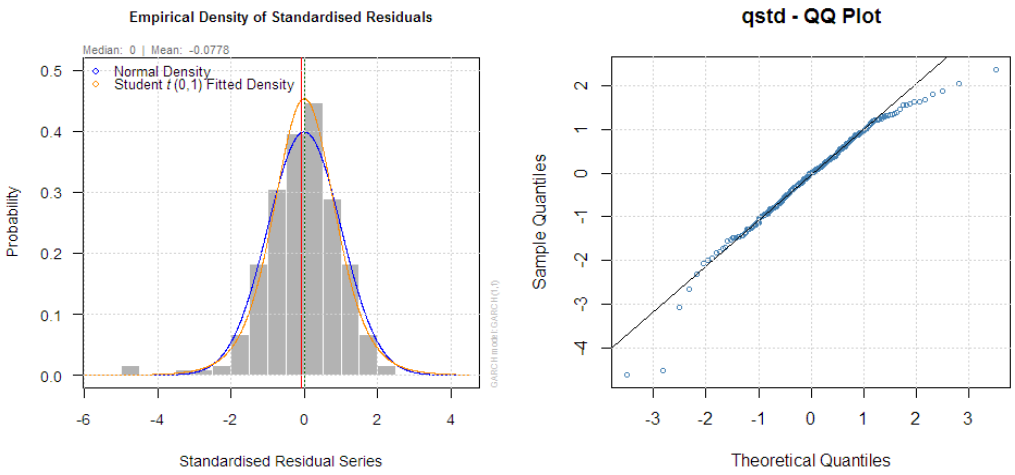


Figure 3.11: Histogram of Standardised Residuals with Normal and Student *t* Distributions and the Student *t* QQ Plot of Standardised Residuals

Table 3.8 displays the fitting results for all asset classes and risk factors consistent with the methodology of this section. Output includes the ARMA-GARCH model specification, conditional density family, estimated parameters with respective *p*-values of the *t*-test, log-likelihood estimates and information criteria estimates of the fitted time series models.

Table 3.8: Model Specification, Estimated Parameters, Log-Likelihood and Information Criteria of the Fitted Time Series Models

Index Name	Model Specification						Parameters										LLH	Info. Crit.	
	ARMA		GARCH			Distr.			ARMA		GARCH-(M)				Distribution				
	m	n	p	q	M			μ	ϕ	θ	$\xi^{\dagger}$	ω	α	β	ν	$\xi^{\ddagger}$	κ		AIC	BIC
ALSI	0	0	1	1		t	estimate	0.0159				0.0003	0.2343	0.6839	7.2137			376.33	-3.07	-2.99
							p-value	0.0000				0.2029	0.1109	0.0001	0.0092					
ALBI	0	0	1	1		t	estimate	0.0110				0.0001	0.2961	0.4766	5.1385			586.61	-4.81	-4.73
							p-value	0.0000				0.0125	0.0174	0.0001	0.0021					
MSCI.WRLD.ZAR	0	0	1	1		skew t	estimate	0.0129				0.0003	0.2726	0.6335		1.2060	6.0890	414.42	-3.38	-3.29
							p-value	0.0000				0.0381	0.0243	0.0000		0.0000	0.0147			
GMC1 Cash[§]	1	0	1	1	×	skew GED	estimate	0.0038	0.9000		6.0911	0.0000	0.1382	0.8544		0.6655	1.1696	1637.10	-13.47	-13.37
							p-value	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000		0.0000	0.0000			
GLOUS	0	0	1	1		GED	estimate	0.0041				0.0000	0.0657	0.8717	1.7251			625.64	-5.13	-5.06
							p-value	0.0004				0.2308	0.1451	0.0000	0.0000					
J253T	0	0	1	1		skew t	estimate	0.0161				0.0003	0.2134	0.6260	11.0473	0.9010		417.48	-3.40	-3.31
							p-value	0.0000				0.0717	0.0290	0.0000	0.1011	0.0000				
USDZAR	0	0	1	1		skew GED	estimate					0.0001	0.2231	0.7533		1.0747	1.1875	428.06	-3.50	-3.42
							p-value					0.0432	0.0046	0.0000		0.0000	0.0000			
EURUSD	0	0	1	1		GED	estimate					0.0001	0.0724	0.8513	1.5580			516.48	-4.24	-4.18
							p-value					0.2809	0.0774	0.0000	0.0000					
BRSPOT	0	0	1	1		t	estimate	0.0099				0.0005	0.1796	0.7755	8.2630			241.59	-1.96	-1.88
							p-value	0.0604				0.3023	0.0146	0.0000	0.0438					
GSCI	0	0	1	1		skew GED	estimate					0.0004	0.0642	0.8436		0.7343	1.4539	329.96	-2.69	-2.61
							p-value					0.0424	0.0914	0.0000		0.0000	0.0000			
GLFX	0	0	1	1		skew t	estimate	0.0046				0.0001	0.2363	0.7020	6.1742	1.2780		474.99	-3.88	-3.79
							p-value	0.0379				0.1169	0.0083	0.0000	0.0123	0.0000				
PLAT	0	0	1	1		t	estimate	0.0086				0.0007	0.2522	0.5792	5.6716			361.36	-2.95	-2.87
							p-value	0.0058				0.0465	0.0122	0.0000	0.0058					
MSCI.EM.USD	0	0	1	1		skew GED	estimate					0.0004	0.1305	0.8035		0.6174	1.7926	323.70	-2.63	-2.56
							p-value					0.0642	0.0098	0.0000		0.0000	0.0000			
FSPI	0	0	1	1		t	estimate	0.0102				0.0001	0.1993	0.7758	8.6394			435.89	-3.56	-3.49
							p-value	0.0000				0.1812	0.0045	0.0000	0.0417					
JPEMBI[§]	0	0	1	1	×	skew GED	estimate	-0.0066			0.4277	0.0000	0.0817	0.8954		0.7099	0.8608	507.51	-4.14	-4.06
							p-value	0.0000			0.0000	0.0000	0.0000	0.0000		0.0000	0.0000			
USALCI	0	0	1	1		t	estimate	0.0047				0.0000	0.0505	0.8490	10.8323			714.81	-5.87	-5.79
							p-value	0.0000				0.0000	0.0002	0.0000	0.0694					
RBAS[§]	1	0	1	1		GED	estimate		0.2542			0.0001	0.2665	0.6474			0.4687	552.83	-4.54	-4.48
							p-value		0.0000			0.0000	0.0010	0.0000			0.0000			
JAYC10[§]	1	1	1	1		skew GED	estimate		-0.7597	0.8219		0.0006	0.1780	0.4665		1.1689	1.2659	432.47	-3.52	-3.44
							p-value		0.0000	0.0000		0.0000	0.0180	0.0234		0.0000	0.0000			

[§] Variance targeting, where the long-run variance ω is imposed to be the sample variance, was used to stabilise and simplify the model. The technique results in the conditional variance converging towards the unconditional long-run variance $\hat{\sigma}^2$. In the GARCH(p,q) equation (6), the following replacement is made: $\omega = \left(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j\right)\sigma^2$, where the sample variance estimate $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n} \varepsilon_t^2$ is substituted for σ^2, before estimating the remaining parameters (Teräsvirta, 2009, p. 20). For JPEMBI the full sample variance was used; for RBAS and JAYC10, the variance of the last 5 years was used; for GMC1 Cash, the variance of the last 3 years was used.

[†] GARCH-In-Mean parameter.

[‡] Skewness parameter.

3.4 Extreme Value Theory Modelling

The estimation procedure of EVT relies on the assumption that the given data points are realisations of i.i.d. random variables. The ARMA-GARCH filtering process is a first-step towards producing sets of such variables. Although the assumption appears restrictive in terms of behaviour in financial data, it may nonetheless be satisfied for many stationary time series in the limit of exceedances over high thresholds (Carmona, 2014, p. 119). The data may indeed be assumed to be weakly dependent (McNeil, 1999, p. 6).

The BDS test (Brock, Dechert, Scheinkman & LeBaron, 1996) is a portmanteau test for the null hypothesis of i.i.d., designed for estimated residuals from fitted time series models. Before proceeding with EVT, standardised residuals are subjected to the test. The test uses a measure of the frequency with which temporal patterns are repeated in the data, called a correlation integral. A time series z_t for $t = 1, \dots, n$ is defined along its m-history as $z_t^m = (z_t, z_{t-1}, \dots, z_{t-m+1})$. The correlation integral, at embedding dimension m, is estimated by:

$$C_{m,\epsilon} = \frac{2}{n_m(n_m - 1)} \sum_{m \le s <} \sum_{t \le n} I(z_t^m, z_s^m; \epsilon) ,$$

where $n_m = n - m + 1$ and $I(z_t^m, z_s^m; \epsilon)$ is an indicator function evaluating to one if $|z_{t-i} - z_{s-i}| < \epsilon$ for $i = 0,1, \dots, m-1$ (zero otherwise). The correlation integral estimates the probability that any two m-dimensional points are within a distance ϵ of each other and may be understood as estimating the joint probability:

$$P(|z_t - z_s| < \epsilon, |z_{t-1} - z_{s-1}| < \epsilon, \dots, |z_{t-m+1} - z_{s-m+1}| < \epsilon) .$$

If z_t are i.i.d., the joint probability should equate to, in the limiting case:

$$C_{1,\epsilon}^m = P(|z_t - z_s| < \epsilon)^m .$$

The BDS statistic is then defined as follows:

$$V_{m,\epsilon} = \sqrt{n} \frac{C_{m,\epsilon} - C_{1,\epsilon}^m}{s_{m,\epsilon}},$$

where $s_{m,\epsilon}$ is a consistent estimate of the standard deviation of $\sqrt{n}\left(C_{m,\epsilon} - C_{1,\epsilon}^m\right)$. The null hypothesis of i.i.d. is rejected at the 5% significance level when $\left|V_{m,\epsilon}\right| > 1.96$. The epsilon points ϵ in the test should be set to between $\frac{1}{2}\hat{\sigma}_{z_t}$ and $\frac{3}{2}\hat{\sigma}_{z_t}$ (Wang, 2006, p. 115). For a small sample of 242 data points, the standardised residual series is embedded in m-space where $m = 2$. Table 3.9 displays results from the BDS test for i.i.d. in the GARCH-standardised residuals from the fitted ALSI model. The results indicate independence in the series.

Table 3.9: BDS Test Results for i.i.d. in the Standardised Residuals of the Fitted ALSI GARCH Model

ϵ for near points	Test Statistic	*p*-value
0.5032 = $\frac{1}{2}\hat{\sigma}_{z_t}$	1.1741	0.2404
1.0063 = $\hat{\sigma}_{z_t}$	0.6709	0.5023
1.5095 = $\frac{3}{2}\hat{\sigma}_{z_t}$	0.3245	0.7455

*p < 0.1, **p < 0.05, ***p < 0.01, n = 242, $\hat{\sigma}_{z_t} = 1.0063$. The GLFX variable rejected the null for $= \frac{1}{2}\hat{\sigma}_{z_t}$, possibly spuriously; however, GMC1 Cash, JPEMBI and RBAS strongly rejected the null at all epsilon points.

The marginal distribution for each variable is estimated using the respective standardised, uncorrelated residuals. A semi-parametric method is used, where the interior, or bulk, of the distribution is modelled with a non-parametric Gaussian kernel density estimator, while each tail of the distribution is modelled with a parametric generalised Pareto distribution.

3.4.1 The Kernel Density Estimator

Given a sequence of n i.i.d. observations $x_1, \dots, x_n$ from random variable X, with unknown true density $h(x)$ defined on $\mathbb{R}$, a kernel density estimator (KDE) of h is the function $\hat{h}_b$ defined by:

$$\hat{h}_b(x) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right), \qquad (28)$$

where $b > 0$ is the bandwidth, a smoothing parameter, and $K(x)$ is a kernel function, usually defined to be a smooth, symmetric-about-zero and unimodal PDF satisfying the conditions $K(x) \geq 0$ and $\int_{-\infty}^{\infty} K(x)dx = 1$. For consistency in the outside divisor (i.e., $1/nb$) with respect later in this chapter to the GPD tail estimators, the KDE in equation (28) is redefined using the scale notation $K_b(y) = b^{-1}K(y/b)$, transforming the KDE to:

$$\hat{h}_b(y) = n^{-1}\sum_{i=1}^{n} K_b(x - x_i)\,. \qquad (29)$$

A mean zero Gaussian PDF, given by $K(x) = \left(1/\sqrt{2\pi}\right)\exp(-x^2/2)$, is used as the kernel in this dissertation. Bandwidth is computed using the oversmoothed bandwidth selector, which, for the Gaussian KDE, is given by[81]:

$$b_{OS} = 1.14\frac{\hat{\sigma}}{n^{1/5}},$$

where $\hat{\sigma}$ is the standard deviation of the observations[82]. The form of the KDE used in this dissertation, therefore, becomes:

[81] See Wolters (2012, p. 48).

[82] For the ALSI GARCH model, $\hat{\sigma}_{z_t} \approx 1.0063$, yielding $b_{OS} = 1.14\frac{\hat{\sigma}}{n^{1/5}} = 1.14\frac{1.0063}{242^{1/5}} \approx 0.3827$.

$$\hat{h}_{b_{OS}}(y) = n^{-1}\sum_{i=1}^{n}\frac{1}{b_{OS}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-x_i}{b_{OS}}\right)^2\right]. \quad (30)$$

The kernel estimate converges in probability to the true density under the mild condition of decreasing the bandwidth as the sample size increases.

3.4.2 The Extreme Value Model

The Peaks-over-Threshold (POT) approach is the method of choice for modelling threshold exceedances. For a sequence of n i.i.d. observations $x_1, \dots, x_n$ from random variable X and threshold u, a fixed, real number in the support of X, the excess distribution over u is defined as the conditional excess distribution of $X - u$ given $X > u$. The corresponding CDF is given by:

$$F_u(y) = P(X - u \le y | X > u) = \frac{F(y+u) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}, \quad (31)$$

where $x \ge 0$, $y = x - u$ are the excesses and $0 \le y \le x_F - u$ for x_F the right endpoint of F. The excess CDF F_u describes the distribution of the excess losses over threshold u, given that the threshold is exceeded. The excess distribution F_u is also known as the residual life distribution function.

The mean of the excess distribution F_u, called the mean excess over threshold u, is given by:

$$e(u) = E(X - u | X > u),$$

and expresses the mean of F_u as a function of u. It describes the expected overshoot of a threshold, given that exceedance occurs. The mean excess function is also known as the mean residual life function. For the GPD, the mean excess function is given by:

$$e(u) = \frac{\beta(u)}{1-\xi} = \frac{\beta + \xi u}{1-\xi}, \quad (32)$$

where

$$\begin{cases} 0 \le u < \infty & 0 \le \xi < 1 \\ 0 \le u \le \frac{-\beta}{\xi} & \xi < 0 \end{cases}.$$

Note that in equation (32), the mean excess is linear in the threshold u. This property is useful in helping determine a threshold when estimating the shape and scale parameters for the GPD.

The POT approach is underpinned by the theorem of Balkema and de Haan (1974) and Pickands (1975).

Theorem 1 (Pickands-Balkema-de Haan). *The distribution of the block maxima M_n converge towards a GEV with shape parameter ξ if and only if the excess distribution $F_u(x)$ over a threshold u converges uniformly in x as u increases, toward a GPD with shape parameter ξ and a scale parameter possibly varying with u* (Carmona, 2014, p. 99).

In other words, if F_X is in the domain of attraction of the GEV distribution $M_n(\cdot;\xi)$, the excess distribution $F_u(x)$ of the sample of exceedances may be approximated by a GPD with shape parameter ξ (independent of the threshold) and a scale parameter $\beta = \beta(u)$, which may depend on u. Formally,

$$F_u(x) \approx G_{\xi,\beta,u}(x), \qquad x \to \infty ,$$

where $G_{\xi,\beta}(x)$ is a generalised Pareto distribution defined as follows:

$$G_{\xi,\beta,u}(x) = G_{\xi,\beta}(x-u) = P(X < x | X > u) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-u}{\beta}\right)\right]_+^{-1/\xi} & \xi \neq 0 , \\ 1 - \exp\left[-\left(\frac{x-u}{\beta}\right)\right]_+ & \xi = 0 , \end{cases} \qquad (33)$$

where

$$\begin{cases} x \geq 0 & \xi \geq 0 \\ 0 \leq x \leq \frac{-\beta}{\xi} & \xi < 0 , \end{cases}$$

with $[y]_+ = \max(y, 0)$ and $\xi \in \mathbb{R}$ and $\beta > 0$, the shape and scale parameters, respectively. The shape parameter determines the size of the tail of the distribution, as well as controls the magnitude and frequency of the extreme value occurrences. The parameter is therefore key in tail extrapolations of the GPD, as follows:

- $\xi = 0 \Rightarrow$ exponential tail decay (and in the limit $\xi \to 0$);
- $\xi > 0 \Rightarrow$ heavier tail decay than exponential (e.g., power law decay or reparameterised Pareto distribution with shape $\alpha = 1/\xi$);
- $\xi < 0 \Rightarrow$ short tail (e.g., Pareto type II distribution) with finite upper end point $u - \beta/\xi$.

Figure 3.12 plots GPD CDFs and PDFs for different shape parameter settings: $\xi = 0$ for exponential, $\xi = 0.5$ for Pareto and $\xi =$ -0.5 for Pareto type II functions. In all cases $\beta = 1$ and $u = 0$.

In fact, for a large class of distributions, as $u \to x_F$, the excess distribution F_u approaches a GPD[83]. This fact is comforting when choosing a Gaussian kernel to approximate the interior of the distribution

[83] The class of distributions includes all the common, continuous distributions typically used in actuarial and statistical science, such as the Gaussian, lognormal, exponential, *F*, gamma, loggamma, Burr, Cauchy, Student *t*, uniform, beta, etc. (McNeil, 1996, pp. 6-7).

and splice into the two GPD tails.

The generalised Pareto probability density function is given by:

$$g_{\xi,\beta,u}(x) = \begin{cases} \left(\frac{1}{\beta}\right)\left[1+\xi\left(\frac{x-u}{\beta}\right)\right]^{-(1/\xi+1)} & \xi \neq 0\,, \\ \left(\frac{1}{\beta}\right)\exp\left[-\left(\frac{x-u}{\beta}\right)\right] & \xi = 0\,, \end{cases} \qquad (34)$$

where

$$\begin{cases} u \leq x < \infty & \xi \geq 0 \\ u \leq x < u - \frac{\beta}{\xi} & \xi < 0\,. \end{cases}$$

with $\xi \in \mathbb{R}$ and $\beta > 0$.

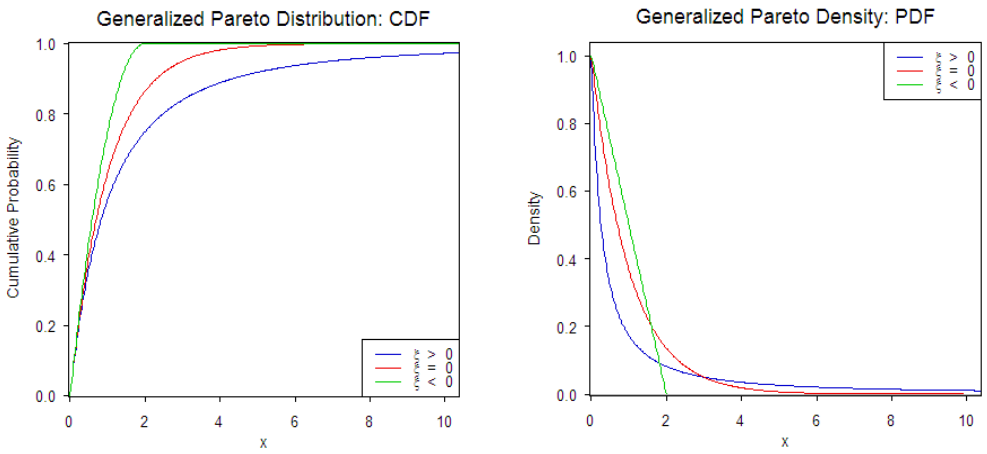


Figure 3.12: GPD CDFs and PDFs

In the semi-parametric distribution setting, the location parameter in equation (33) is set to the threshold u. This setting determines the beginning of the tail and implies that only the shape and scale parameters are fitted. Selecting a threshold necessarily involves a bias-variance trade-off. Choosing too high a threshold reduces the number of data points to which the GPD is fit, increasing variance and decreasing bias in parameter estimates. Conversely, choosing too low a threshold admits data that are not necessarily approximated well by a GPD, thereby increasing bias in parameter estimates in exchange for reduced variance in the model fit.

Using equation (31), theorem 1 and setting an empirical estimate for F_u to $\frac{n-k}{n}$ (McNeil, et al., 2005, p. 283), where n is the total number of observations and k is the number of excesses over the threshold, the CDF for the completed model is defined similarly to that of McNeil and Frey (2000, p. 296) as follows:

$$\hat{F}_Z(z) = \begin{cases} \frac{k_L}{n}\left[1 + \xi^{(k_L)}\left(\frac{\left|z - z_{(n-k_L)}\right|}{\beta^{(k_L)}}\right)\right]_+^{-1/\xi^{(k_L)}}, & z < z_{(n-k_L)} \\ H(z|Z,b)\,, & z_{(n-k_L)} \le z \le z_{(k_U+1)} \\ 1 - \frac{k_U}{n}\left[1 + \xi^{(k_U)}\left(\frac{z - z_{(k_U+1)}}{\beta^{(k_U)}}\right)\right]_+^{-1/\xi^{(k_U)}}, & z > z_{(k_U+1)} \end{cases} \quad (35)$$

where $z_{(1)} \geq z_{(2)} \geq \cdots \geq z_{(n)}$ represents the ordered standardised GARCH residuals, k_L/n is the sample proportion less than a fixed, lower threshold k_L and $(1 - k_U/n)$, the sample proportion greater than a fixed, upper threshold k_U. The sample data for the lower tail is the set of excesses $\{z_{(n)} - z_{(n-k_L)}, \ldots, z_{(n-k_L+1)} - z_{(n-k_L)}\}$, fit with a $G_{\xi^{(k_L)},\beta^{(k_L)}}(z - z_{(n-k_L)})$, and sample data for the upper tail is the set of excesses $\{z_{(1)} - z_{(k_U+1)}, \ldots, z_{(k_U)} - z_{(k_U+1)}\}$, fit with a $G_{\xi^{(k_U)},\beta^{(k_U)}}(z - z_{(k_U+1)})$. The remaining, interior portion of the data is fit with $H(\cdot\,|Z,b)$, the distribution function obtained by integrating the kernel in equation (30).

3.4.3 Threshold Selection and Model Estimation

The interior empirical density is computed non-parametrically using the bandwidth estimate as input to the Gaussian kernel's standard deviation parameter and the respective kernel function to distribute density smoothly over the local neighbourhood of each data point.

Each tail is fit separately by maximum likelihood based on the log of the GPD density, given by:

$$\begin{aligned} \ln L(\xi,\beta|Z_1,\ldots,Z_k) &= \sum_{j=1}^{k} \ln g_{\xi,\beta}(Z_j) \\ &= \sum_{j=1}^{k} \ln\frac{1}{\beta} + \sum_{j=1}^{k} \ln\left[1 + \xi\left(\frac{z_j - u}{\beta}\right)\right]^{-(1/\xi+1)} \\ &= -k\ln\beta - \left(1 + \frac{1}{\xi}\right)\sum_{j=1}^{k} \ln\left[1 + \xi\left(\frac{z_j - u}{\beta}\right)\right] \qquad \xi \neq 0 \end{aligned} \quad (36)$$

and for $\xi = 0$, the log-likelihood function is similarly obtained as:

$$\ln L(\beta|Z_1,\ldots,Z_k) = \sum_{j=1}^{k} \ln g_{\xi,\beta}(Z_j) = -k\ln\beta - \sum_{j=1}^{k}\left(\frac{z_j - u}{\beta}\right) \qquad \xi = 0\,. \quad (37)$$

where the conditions are consistent with those for equation (35).

A threshold needs to be chosen before model fitting may commence. The mean residual life plot has been suggested as a possible guide in threshold selection. Although notoriously difficult to interpret,

ideally, the mean excess above a threshold u is linear. Additional help in choosing a threshold is sourced from a variety of simulation studies on the topic, such as those from Chavez-Demoulin and Embrechts (2004), Chavez-Demoulin, et al. (2014), Embrechts, et al. (1997) and Nyström and Skoglund (2002b). This paper sets the threshold at 10% on each side of the (ordered) data and investigates these regions graphically to reinforce judgment. Figure 3.13 shows the mean residual life (MRL) plots for lower and upper tail ALSI GARCH-fit standardised residuals over the respective maximum ranges of threshold values. An upward trend in a MRL plot is characteristic of heavy-tailed behaviour. In particular, a straight line with positive gradient above some threshold is a sign of Pareto behaviour in the tail. A downward trend implies thin-tailed behaviour and a line with zero gradient implies an exponential tail. The initial negative-gradient to zero-gradient to positive-gradient transition in the left panel can be interpreted as thin-tailed behaviour gradually becoming exponential-tailed behaviour and, finally, becoming heavy-tailed as the threshold value moves beyond, roughly, 1.5 (or -1.5 in unsorted data). The 10% lower tail threshold evaluates to -1.286. The threshold is shown as a positive number in the figure since the GPD is fit to positive values (requiring a sign switch in the ordered data when this is not done automatically in the software). The 10% upper tail threshold is 1.197 (right panel). The right panel shows much of thin-tailed behaviour in the ALSI upper tail residuals, which, eventually, beyond a threshold of roughly 1.2, approaches exponential-tailed behaviour. The figures are plotted with approximate 95% confidence bands around the mean excesses superimposed, which further highlights the instability in mean excess estimates across varying levels of threshold values. As is evident, a precise choice for threshold value cannot be deduced from this kind of plot.

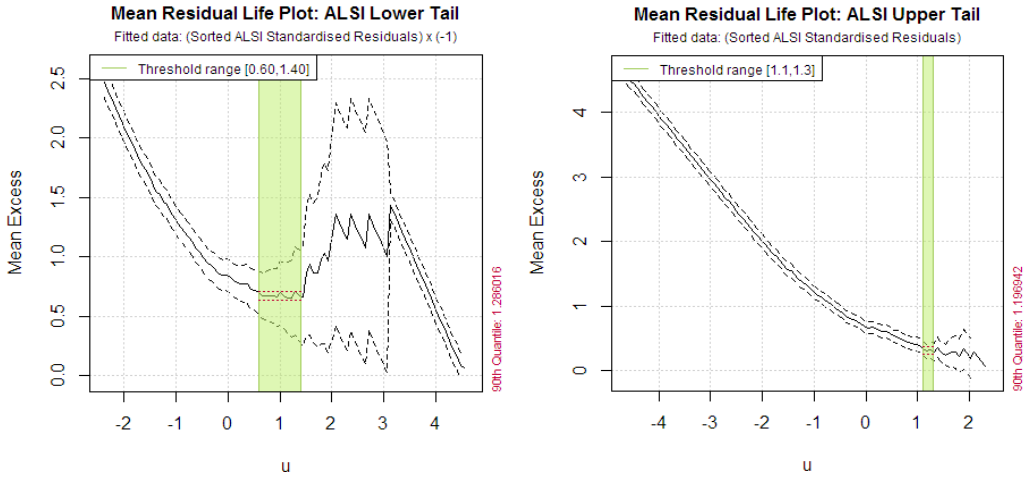


Figure 3.13: Mean Residual Life Plots for the Lower and Upper Tails of Standardised ALSI GARCH Residuals

Figure 3.14 plots maximum likelihood estimates of the scale and shape parameters of the GPD, with 95% confidence bands, along a zoomed-in range of threshold values in the proximity of the 10% lower and 10% upper tail thresholds. The threshold ranges have been narrowed to envelope the 10% lower (left panel) and 10% upper (right panel) threshold values. The graphs further highlight the degree of variability that exists in modelling sparse tail data. Nonetheless, it is interesting to note the

positive and increasing slope on the lower tail shape parameter estimate (lower left panel), which points to an increasingly heavy-tail estimate as a function of an increasing threshold. The upper tail shape parameter estimate points to an exponential tail decay.

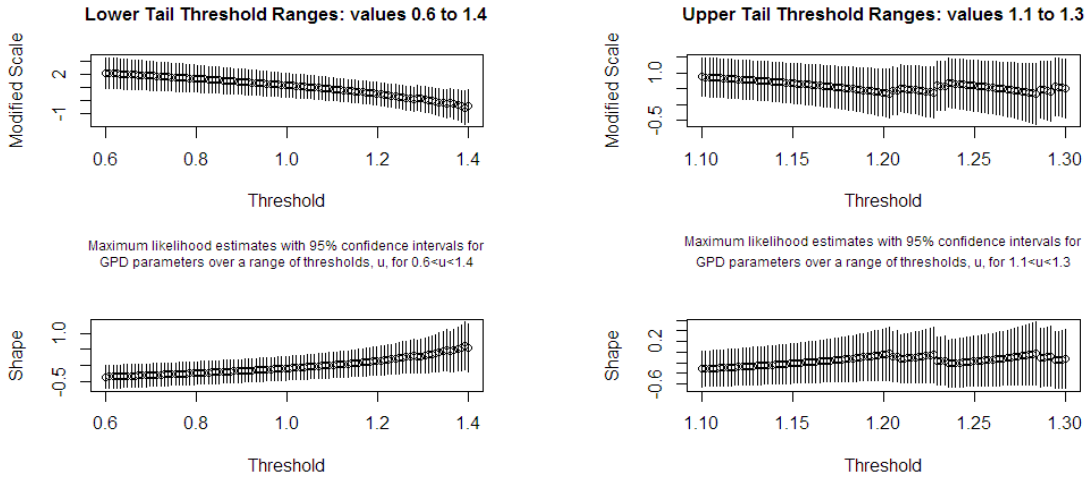


Figure 3.14: Maximum Likelihood Estimates for Scale and Shape Parameters of the GPD

Figure 3.15 is used to illustrate the intuitive effects of the chosen 10% threshold placement. The left panel plots the ordered residuals with the 10% lower and upper cut-off points. The right panel plots the raw residual data with the 10% lower and upper cut-off points. Clearly, the lower 10% of exceedances exhibit a greater range of magnitudes than do the upper 10% of exceedances. This variability in the exceedance data naturally transmits through to parameter estimation.

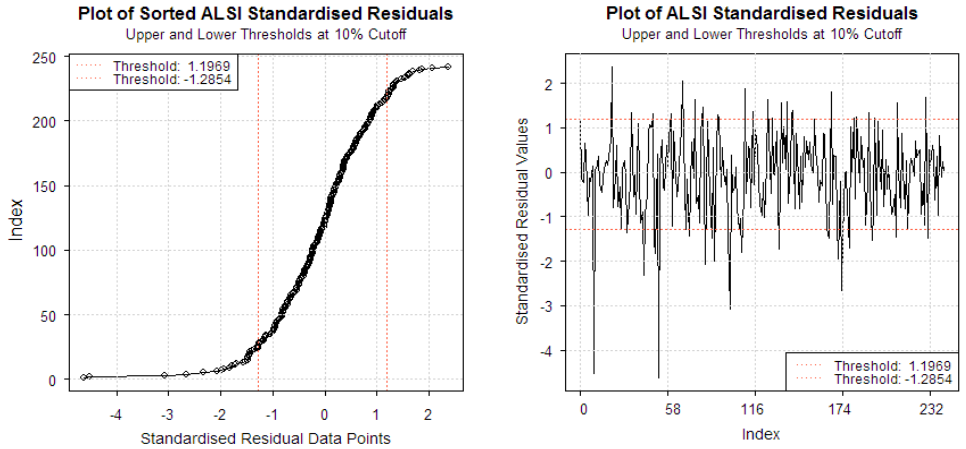


Figure 3.15: Ordered ALSI Standardised Residuals and Raw Residuals with 10% Lower and Upper Tail Threshold Points

Figures 3.16 and 3.17 contain four popular GPD diagnostic plots for the lower and upper tail fits, respectively. In the top left panels, the theoretical GPD distribution of excesses is plotted and superimposed on points plotted at empirical estimates of the excess probabilities for each exceedance. The top right panels plot the lower and upper tails of the fitted GPD, respectively, together with the empirical tails given by the actual data points. The bottom left panels show

scatterplots of each tail's GPD residuals with a locally-weighted polynomial regression smoother (i.e., a fitted ordinary least-squares line) of the residuals superimposed. To note here is the steepness of the regression line in both plots, signalling a size effect in the extreme quantile data. A consequence of the size effect is difficulty in determining the extreme tail bound[84], making estimates unstable with time. The bottom right panels plot a QQ plot for the threshold data. The standard exponential is the reference distribution in the QQ plots, with $\xi = 0$, where the theoretical quantiles are plotted as the straight red line. If $\xi \neq 0$, the reference distribution is the generalised Pareto with that value of ξ. The approximate linear QQ plot in both graphs indicate that the residuals from the GPD follow the reference distribution; the upper tail more so than the lower tail. The diagnostic plots show that the tails are well fit by the GPD model.

Lower tail diagnostics in Figure 3.16: the top left panel plots a GPD fit to 23 exceedances of the threshold u = -1.286. The top right panel plots the GPD fit to the tail of the underlying distribution. The bottom left panel shows a scatterplot of the GPD residuals. The bottom right panel shows a QQ plot of GPD residuals against the exponential reference distribution.

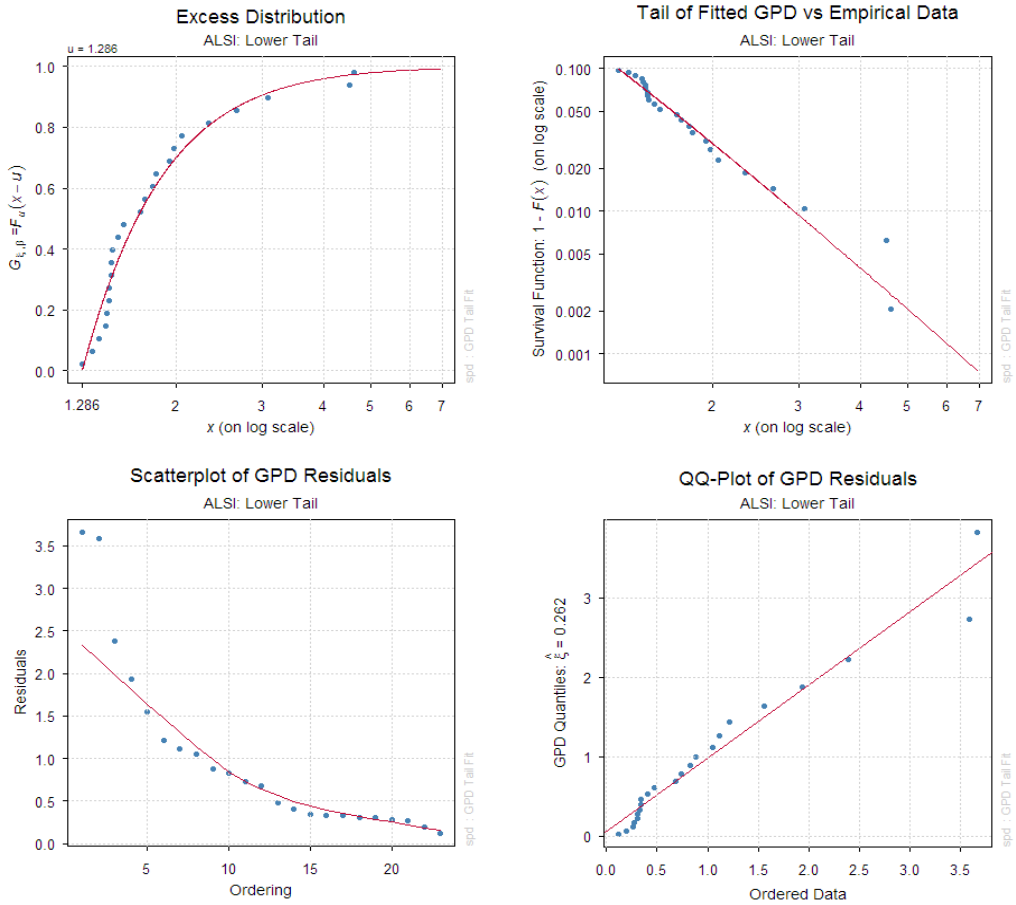


Figure 3.16: GPD Diagnostic Plots for the ALSI Lower Tail Fit

Upper tail diagnostics in Figure 3.17: the top left panel plots a GPD fit to 23 exceedances of the

[84] $\xi > 0$ indicates fat-tailed behaviour and possible difficulty determining boundedness; $\xi < 0$ indicates bounded tails.

threshold $u = 1.197$. The top right panel plots the GPD fit to the tail of the underlying distribution. The bottom left panel shows a scatterplot of the GPD residuals. The bottom right panel shows a QQ plot of GPD residuals against the exponential reference distribution.

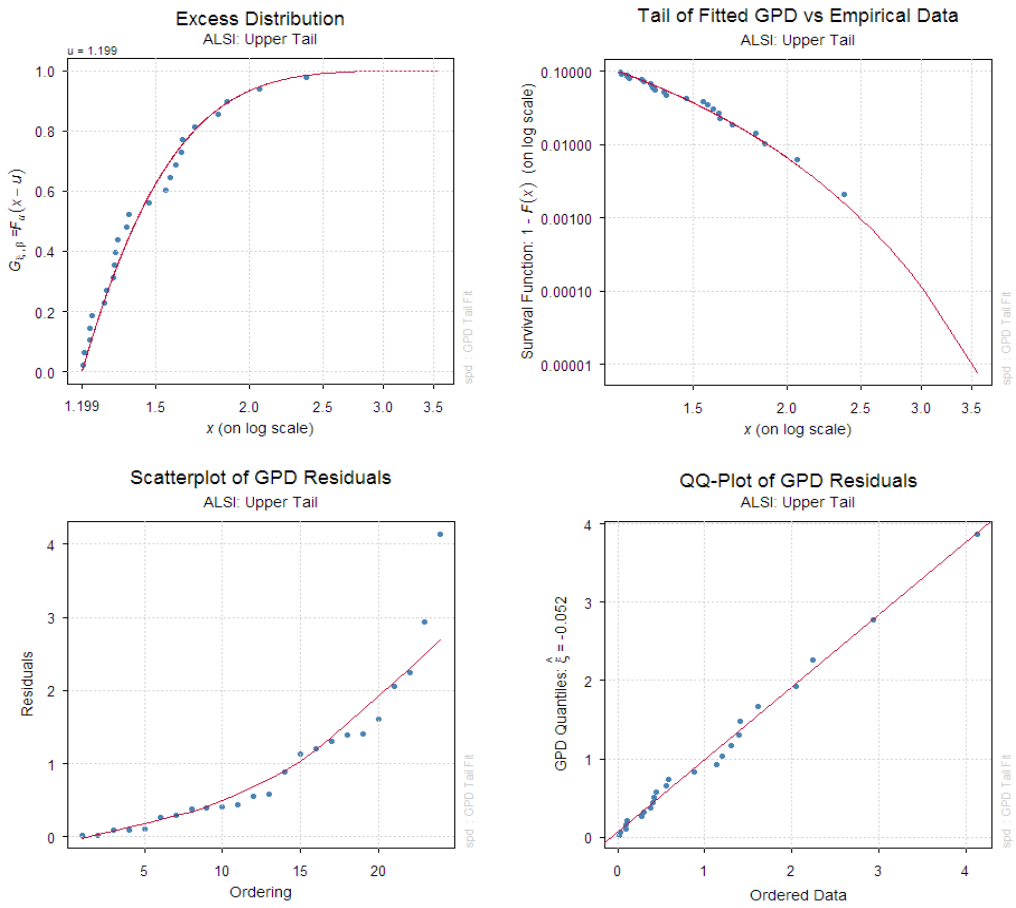


Figure 3.17: GPD Diagnostic Plots for the ALSI Upper Tail Fit

3.4.4 The Semi-Parametric Distribution

The final step in the EVT modelling section is creating the completed semi-parametric distribution. Each variable is represented by:

1. A lower tail fitted with generalised Pareto density estimates for shape and scale;
2. An interior represented by density estimates along an equally-spaced grid spanning the range of the data (and extended into the tails by the support of the kernel);
3. An upper tail fitted with generalised Pareto density estimates for shape and scale.

Figure 3.18 shows the combined non-parametric interior distribution and parametric tail GPDs for the ALSI standardised residual series superimposed over their empirical counterparts. Table 3.10 displays the fitted results for all asset classes and risk factors consistent with the methodology of this section. Output includes the GPD lower tail and upper tail estimated threshold values and parameter estimates (standard errors in parentheses) of the return filtered standardised residuals.

Estimates were obtained using the spd package in R (Ghalanos, 2014c).

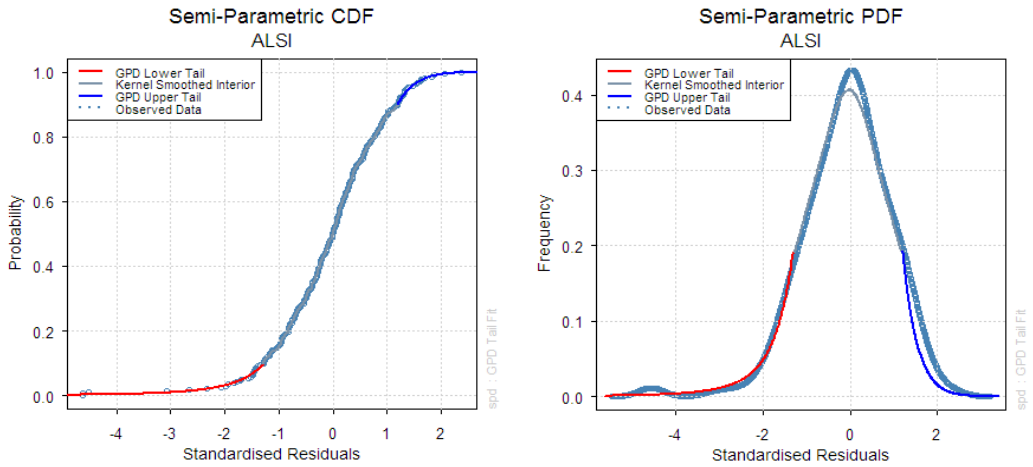


Figure 3.18: Semi-Parametric CDF and PDF for ALSI Standardised Residuals

Table 3.10: GPD Threshold Value and Parameter Estimates

Index Name		GPD Parameters							
		Lower Tail				Upper Tail			
		k_L	$\xi^{(k_L)}$	$\beta^{(k_L)}$	Heavy?	k_U	$\xi^{(k_U)}$	$\beta^{(k_U)}$	Heavy?
ALSI	estimate	1.2863	0.2620	0.5420	✓	1.1979	-0.0561	0.3186	×
	std. error		(0.2638)	(0.1798)			(0.2463)	(0.1018)	
ALBI	estimate	1.2213	0.0657	0.7653	✓	1.2223	-0.1861	0.4870	×
	std. error		(0.2321)	(0.2385)			(0.2640)	(0.1615)	
MSCI.WRLD.ZAR	estimate	1.0509	-0.4749	0.9465	×	1.0680	-0.1691	0.8982	×
	std. error		(0.2831)	(0.3193)			(0.1861)	(0.2462)	
GMC1 Cash	estimate	0.3473	-0.6201	1.4153	×	1.3985	0.2789	0.4606	✓
	std. error		(0.2209)	(0.3890)			(0.2689)	(0.1533)	
GLOUS	estimate	1.0986	-0.7506	1.1732	×	1.2286	-0.2283	0.7475	×
	std. error		(0.2973)	(0.3844)			(0.2523)	(0.2401)	
J253T	estimate	1.2068	-0.2325	0.8586	×	1.2226	0.1817	0.3551	✓
	std. error		(0.2817)	(0.2980)			(0.2382)	(0.1106)	
USDZAR	estimate	0.9121	-0.6600	0.8078	×	1.2941	-0.0344	0.9952	×
	std. error		(0.2202)	(0.2213)			(0.2192)	(0.2980)	
EURUSD	estimate	1.3262	-0.1650	0.6748	×	1.2410	-0.4945	0.7752	×
	std. error		(0.2048)	(0.1958)			(0.2326)	(0.2268)	
BRSPOT	estimate	1.3674	0.0702	0.5602	✓	1.1151	-0.1217	0.5387	×
	std. error		(0.2030)	(0.1628)			(0.1901)	(0.1497)	
GSCI	estimate	1.1097	0.1534	0.5824	✓	1.1966	-0.8014	0.7836	×
	std. error		(0.2233)	(0.1770)			(0.3604)	(0.2983)	
GLFX	estimate	1.0683	-0.1909	0.5945	×	1.2540	0.2733	0.4721	✓
	std. error		(0.4178)	(0.2766)			(0.2747)	(0.1594)	
PLAT	estimate	1.1034	-0.1656	0.9610	×	1.1089	-0.3287	0.7509	×
	std. error		(0.2255)	(0.2933)			(0.2620)	(0.2443)	
MSCI.EM.USD	estimate	1.3343	-0.0127	0.7157	×	1.2109	-0.4111	0.3631	×
	std. error		(0.2681)	(0.2430)			(0.2901)	(0.1251)	
FSPI	estimate	1.4959	0.0628	0.5978	✓	1.0488	-0.0734	0.2600	×
	std. error		(0.2024)	(0.1735)			(0.2706)	(0.0880)	
JPEMBI	estimate	1.1585	0.5103†	0.5906	✓	0.9734	-0.4721	0.4769	×
	std. error		(0.3127)	(0.2129)			(0.2111)	(0.1323)	
USALCI	estimate	1.2686	-0.3722	0.8071	×	1.1229	0.0222	0.5661	✓
	std. error		(0.2723)	(0.2696)			(0.1737)	(0.1517)	
RBAS	estimate	1.2962	-0.1484	1.7023	×	0.9817	-0.0095	1.1338	×
	std. error		(0.2436)	(0.5434)			(0.2961)	(0.4078)	
JAYC10	estimate	1.2621	-0.2298	0.7521	×	1.1868	0.0362	0.7947	✓
	std. error		(0.1746)	(0.2013)			(0.2259)	(0.2419)	

† As noted in Inanoglu and Ulman (2009, p. 16), a GPD has an infinite variance if its shape parameter is greater than or equal to 0.5 ($\xi \geq 0.5$) and an infinite mean for shape parameter greater than or equal to 1 ($\xi \geq 1$).

3.5 Dependence Modelling

In a portfolio of multiple financial instruments, all the information on the stochastic behaviour of the portfolio is fully described by the joint probability distribution. The multivariate (MV) distribution of a portfolio may be fully specified by the separate marginal distributions of the variables and by their copula (or dependence structure). Marginal distributions were modelled in Chapter 3.4, resulting in common continuous distributions $\hat{F}_{Z_i}$ for the $i = 1, \dots, 18$ variables. This section fits an MV *t* copula to the univariate marginals.

3.5.1 Copulas

A univariate distribution function F_X, which maps from the domain of a random variable X to the "grade" of the random variable X (i.e., a mapping from $\mathbb{R}$ to $\mathbb{I} \in [0,1]$), has the following basic properties:

1. F_X is non-decreasing;
2. $F_X(-\infty) = 0$;
3. $F_X(\infty) = 1$.

An arbitrarily-distributed random variable X may be transformed by its own CDF (known as a probability transform) to obtain the grade of X: $U \equiv F_X(X)$. The distribution of the grade of X is uniform on the unit interval, regardless of the original distribution: $U \equiv F_X(X) \sim U_{[0,1]}$[86]. The result also works backwards. A standard uniform random variable U may be fed into the inverse CDF F_X^{-1} to obtain a random variable X (known as a quantile or inverse transform): $X \equiv F_X^{-1}(U) \sim f_X$[87]. In addition, any continuous distribution F_X may be chosen, thus admitting the completed semi-parametric distribution form of Chapter 3.4.

In the following, let $\boldsymbol{X} = (X_1, \dots, X_d)$ be a vector of random variables $X_1, \dots, X_d$ described by joint density function f and joint cumulative distribution function F. Further, let $f_1, \dots, f_d$ denote the corresponding marginal density functions and $F_1, \dots, F_d$ the strictly increasing and continuous marginal distribution functions of $X_1, \dots, X_d$. A joint multivariate distribution function F, which maps from $\mathbb{R}^d$ to $\mathbb{I} = [0,1]$, has the properties:

1. F is d-increasing (see point 1 below);

[86] It may be noted that the CDF of a uniform distribution $F_U(u) = P(U \leq u) = P(F_X(X) \leq u) = P\left(X \leq F_X^{-1}(u)\right) = F_X\left(F_X^{-1}(u)\right) = u$.

[87] Some caveats do apply. For example, the normal inverse CDF is not defined for the uniform boundary values $u = 0$ or $u = 1$.

2. F is grounded (see point 2 below);
3. $F(\infty) = 1$;
4. $F\left(\infty, \dots, \infty, x_j, \infty, \dots, \infty\right) = F_j\left(x_j\right)$ for $j = 1, \dots, d$.

A d-dimensional copula $C(u_1, \dots, u_d)$ is a MV uniform distribution function of the univariate grades, is defined on the unit hypercube $[0,1]^d$, and has d univariate margins distributed standard uniformly on the closed unit interval $\mathbb{I} = [0,1]$. Formally, $C: [0,1]^d \mapsto [0,1]$, where $C(\boldsymbol{u}) = C(u_1, \dots, u_d)$ is a d-dimensional copula if:

1. $C(\boldsymbol{u}) \geq 0 \; \forall \; \boldsymbol{u} \in \mathbb{I}^d$ (i.e., $C(\boldsymbol{u})$ is d-increasing in each component u_i for $i = 1, \dots, d$);
2. $C(u_1, \dots, u_i, \dots, u_d) = 0$ if any $u_i = 0$ for $i = 1, \dots, d$ (i.e., the function is grounded);
3. $C(1, \dots, 1, u_i, 1 \dots, 1) = u_i \; \forall \; i \in \{1, \dots, d\}, \; u_i \in [0,1]$.

Sklar's theorem (Sklar, 1959) admits a copula function to characterise multivariate dependence. This function is the link which "glues together" the univariate distributions to create a multivariate distribution.

Theorem 2 (Sklar's Theorem). *Let F be a d-dimensional distribution function with margins $F_1, \dots, F_d$. Then there exists a d-dimensional copula C such that for all $\boldsymbol{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$,*

$$F(x_1, \dots, x_d) = C\left(F_1(x_1), \dots, F_d(x_d)\right). \tag{38}$$

If the margins $F_1, \dots, F_d$ are continuous, then C is unique. Conversely, if C is a d-dimensional copula and $F_1, \dots, F_d$ are univariate distribution functions, then the function F defined in equation (38) is a d-dimensional distribution function with margins $F_1, \dots, F_d$.

For the d-dimensional random vector $\boldsymbol{X} = (X_1, \dots, X_d)'$, the copula C of joint distribution function F in equation (38) may be extracted as follows:

$$C(\boldsymbol{u}) := C(u_1, \dots, u_d) = F\left(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\right), \tag{39}$$

where $F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)$ denote the generalised inverse functions, or quantile transforms, of the margins, defined by $F_i^{-1}(u) = \inf\{x \in \mathbb{R}^1; \; F_i(x) \geq u\}$.

Assuming a joint multivariate PDF $f_{\boldsymbol{X}}$ of vectors $\boldsymbol{X}$, Sklar admitted the following procedure:

1. Extract all the d marginal distributions $X_j \sim f_{X_j}$, where $j = 1, \dots, d$, by:

$$f_{X_j}\left(x_j\right) = \int_{\mathbb{R}^{d-1}} f_{\boldsymbol{X}}\left(x_1, \dots, x_d\right) dx_1 \cdots dx_{j-1} \, dx_{j+1} \cdots dx_d \, .$$

2. Compute the marginal CDFs by $F_{X_j}\left(x_j\right) = \int_{-\infty}^{x} f_{X_j}(z) dz$.

3. Feed each CDF F_{X_j} with the corresponding entry of the vector $\boldsymbol{X}$ (i.e., the random variable X_j) to obtain the grades with uniform distribution on the unit interval: $U_j = F_{X_j}(X_j) \sim U_{[0,1]}$. This probability transform removes the idiosyncratic information contained in each marginal distribution. The resulting entries of $\boldsymbol{U} \equiv (U_1, \dots, U_d)$
    a. are *not* independent
    b. represent the pure joint information amongst the X_j's.
4. Estimate the joint distribution $f_{\boldsymbol{U}}$ (the copula function) on the grades of arbitrary distribution $f_{\boldsymbol{X}}$. The joint distribution $f_{\boldsymbol{U}}$ of the grades is, by 3(a), not uniform on its domain, the unit hypercube $[0,1]^d$:

$$\begin{pmatrix} U_1 \equiv F_{X_1} \\ \vdots \\ U_d \equiv F_{X_d} \end{pmatrix} \sim f_{\boldsymbol{U}} \, .$$

Sklar's "joint = copula + marginals" decomposition may formally be defined as follows (Meucci, 2011):

$$\underbrace{f_{\boldsymbol{X}}\left(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d)\right)}_{\text{joint}} = \underbrace{f_{\boldsymbol{U}}(u_1, \dots u_d)}_{\text{pure joint}} \times \underbrace{\left[f_{X_1}\left(F_{X_1}^{-1}(u_1)\right) \times \cdots \times f_{X_d}\left(F_{X_d}^{-1}(u_d)\right)\right]}_{\text{pure marginal}} \qquad (40)$$

where

$$\begin{aligned} f_{\boldsymbol{X}}\left(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d)\right) &= f_{\boldsymbol{X}}(x_1, \dots, x_d) \\ &= \frac{\partial^d \left[C\left(F_1(x_1), \dots, F_d(x_d)\right)\right]}{\partial F_1(x_1) \cdots \partial F_d(x_d)} \cdot \prod_{i=1}^{d} f_i(x_i) \\ &= c\left(F_1(x_1), \dots, F_d(x_d)\right) \cdot \prod_{i=1}^{d} f_i(x_i) \end{aligned} \qquad (41)$$

with $f_i(x_i) = \mathrm{d}F_j(x_j)/\mathrm{d}x_j$ and c, the density of the copula used to calibrate its parameters to the data, follows as:

$$\begin{aligned} c\left(F_1(x_1), \dots, F_d(x_d)\right) &= c(u_1, \dots, u_d) \\ &= \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \cdots \partial u_d} \sim f_{\boldsymbol{U}} \, . \end{aligned} \qquad (42)$$

Equation (41) is the canonical copula representation in copula theory (care of Sklar's decomposition theorem) for the MV density function.

In practice, steps 1 – 4 may be adjusted and manipulated to accommodate the copula fitting. Analytical derivation is typically not possible, necessitating the use of numerical techniques (e.g., Monte Carlo simulation) in the majority of practical applications.

3.5.1.1 The Standard Student *t* Copula

An elliptical copula is the copula that, by Sklar's theorem, derives from its corresponding elliptical distribution. Elliptical distributions are those whose multivariate densities $f_{\boldsymbol{X}}$ depend on $\mathbf{x}$ only through the quadratic form $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$. The MV Student *t* copula is the copula structure extracted from the MV Student *t* distribution.

The d-dimensional random vector $\boldsymbol{X}=(X_1,\dots,X_d)'$, is said to have a (non-singular) MV Student *t* distribution with ν degrees of freedom, mean vector $\boldsymbol{\mu}$ and positive-definite dispersion or scatter matrix Σ, denoted $\boldsymbol{X}\sim t_d(\nu,\boldsymbol{\mu},\Sigma)$, if its density is given by (Demarta & McNeil, 2005, p. 2):

$$f_{t_d(\nu,\boldsymbol{\mu},\Sigma)}(\mathbf{x})=\frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d|\Sigma|}}\left(1+\frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}}, \qquad (43)$$

where $\Gamma(\cdot)$ is the gamma function and $|\cdot|$ denotes the determinant function. Note that in this standard parameterisation, the expectation vector $E(\boldsymbol{X})=\boldsymbol{\mu}$ is only defined for $\nu>1$ and the covariance matrix $cov(\boldsymbol{X})=\frac{\nu}{\nu-2}\Sigma\ (\neq\Sigma)$ is only defined for $\nu>2$. Also, for $\nu>2$, the shape parameter matrix Σ can be interpreted as the linear correlation matrix. It is not necessarily the shape parameter matrix Σ, equivalently in $\nu>2$ the linear correlation matrix Σ, that determines the extent of tail dependence, but rather the df parameter ν that controls dependence and, thus, the tendency to exhibit extreme co-movements.

A key property of the copula is that it remains invariant under any application of strictly increasing, or co-monotonic transformations (vs. counter-monotonic decreasing transforms), of the components of the random vector $\boldsymbol{X}$. A series of co-monotonic transformations $Y_j\equiv g_j(X_j)$ of the entries of $\boldsymbol{X}$, where $g_j(x)\equiv F_{Y_j}^{-1}\left(F_{X_j}(x)\right)$ for arbitrary inverse CDFs $F_{Y_j}^{-1}$, do not alter the copula structure of $\boldsymbol{X}$:

$$\begin{pmatrix}X_1\\ \vdots\\ X_d\end{pmatrix}\sim f_{\boldsymbol{X}} \quad \xrightarrow{g_1,\dots,g_d} \quad \begin{pmatrix}Y_1\equiv F_{Y_1}^{-1}\left(F_{X_1}(X_1)\right)\\ \vdots\\ Y_d\equiv F_{Y_d}^{-1}\left(F_{X_d}(X_d)\right)\end{pmatrix}\sim f_{\boldsymbol{Y}}$$

$$F_{X_1},\dots,F_{X_d} \qquad\qquad F_{Y_1},\dots,F_{Y_d}$$

$$\begin{pmatrix}U_1\\ \vdots\\ U_d\end{pmatrix}\sim f_{\boldsymbol{U}}\,.$$

The co-monotonic invariance property implies that the copula remains invariant under a standardisation of the marginal distributions. Hence, the copula of a $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ is identical to that of a $t_d(\nu, \mathbf{0}, P)$ distribution, where P is the correlation matrix implied by the dispersion matrix Σ. The corresponding standardised MV Student *t* distribution, with ν df and P a symmetric, positive-definite correlation matrix with $\mathrm{diag}(P) = \mathbf{1}$, is given by (Bouyé, et al., 2000, p. 16):

$$T_{\nu,P}(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d|P|}} \left(1 + \frac{\mathbf{x}' P^{-1} \mathbf{x}}{\nu}\right)^{-\frac{\nu+d}{2}} d\mathbf{x}\,. \qquad (44)$$

The unique MV Student *t* copula is thus given by (Demarta & McNeil, 2005, p. 3):

$$C(u_1, \dots, u_d; \nu, P) = C^t_{\nu,P}(\boldsymbol{u}) = T_{\nu,P}\left(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)\right) = P(t_\nu(X_1) \le u_1, \dots, t_\nu(X_d) \le u_d)$$

$$= \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_d)} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d|P|}} \left(1 + \frac{\mathbf{x}' P^{-1} \mathbf{x}}{\nu}\right)^{-\frac{\nu+d}{2}} d\mathbf{x}\,, \qquad (45)$$

where t_ν^{-1} denotes the quantile function of a standard univariate Student t_ν distribution. For estimation purposes, the corresponding *t* copula density may be calculated per equation (39) and has the form:

$$c(u_1, \dots, u_d; \nu, P) = c^t_{\nu,P}(\boldsymbol{u}) = \frac{f_{\nu,P}\left(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)\right)}{\prod_{i=1}^d f_\nu\left(t_\nu^{-1}(u_i)\right)}$$

$$= |P|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\nu+\mathrm{d}}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{d-1}}{\Gamma\left(\frac{\nu+1}{2}\right)^d} \left(1 + \frac{\boldsymbol{\varsigma}' P^{-1} \boldsymbol{\varsigma}}{\nu}\right)^{-\frac{\nu+d}{2}} \prod_{i=1}^d \left(1 + \frac{\varsigma_i^2}{\nu}\right)^{\frac{\nu+1}{2}}, \qquad (46)$$

where $f_{\nu,P}$ is the joint density of a $t_d(\nu, \mathbf{0}, P)$-distributed random vector, f_ν is the density of the univariate standard Student *t* distribution with ν df and $\boldsymbol{\varsigma}_i = t_\nu^{-1}(u_i)$ (Bouyé, et al., 2000, p. 16).

3.5.1.2 The Meta Student *t* Copula with SPD-distributed Margins

The traditional way to construct multivariate distributions suffers from the restriction that the margins are usually of the same type (i.e., the corresponding random variables are a linear affine transformation of each other). The second part of Sklar's decomposition in equation (38) admits a converse process: given any copula C and univariate distribution functions $F_1, \dots, F_d$, the d-dimensional distribution function F defined by (38) is a MV distribution function with margins $F_1, \dots, F_d$. In what follows the margins $F_1, \dots, F_d$ are assumed to be continuous, implying F has a unique representation in terms of the copula C and the margins of F. This converse statement a very powerful technique for constructing multivariate distributions with arbitrary margins and copulas.

Rewriting the first line of equation (45) as:

$$C_{\nu,P}^{t}(\boldsymbol{u}) := F_{T_{\nu,P}}\left(F_{t_\nu}^{-1}(u_1), \dots, F_{t_\nu}^{-1}(u_d)\right), \tag{47}$$

where the $F_{t_\nu}^{-1}$ denote the quantile functions of the standard univariate Student *t* distribution with ν df, it is clear that the univariate df parameter is both static across the univariate distributions and consistent with the df parameter of the MV *t* distribution function $F_{T_{\nu,P}}$. What is not immediately clear in (47) is that the probability transformed $U_j = F_{X_j}(X_j)$ are also each transformed by $F_{X_j}(X_j) = T_\nu(X_j)$, where the "local" degrees of freedom parameters are equal to the "global" df parameter from $F_{T_{\nu,P}}$. In fact, the strict underlying assumption supporting the Student *t* copula may be conveyed in the more transparent setting:

$$C_{\nu,P}^{t}(\boldsymbol{u}) := T_{\nu,P}^{d}\left(t_\nu^{-1}\left(T_\nu(Y_1)\right), \dots, t_\nu^{-1}\left(T_\nu(Y_d)\right)\right), \tag{48}$$

where $T_{\nu,P}^{d}$ is the MV *t* distribution function, t_ν^{-1} the univariate Student *t* quantile function and T_ν the univariate Student *t* distribution function, all with the same, static global degrees of freedom estimate ν. In addition, the d-dimensional random vector $\boldsymbol{Y} = (Y_1, \dots, Y_d)'$ in (48) is (strictly) assumed to comprise centred Student *t* distributed random variables, with ν degrees of freedom, given by:

$$\boldsymbol{Y} = RA\boldsymbol{Z}, \tag{49}$$

where $\boldsymbol{Z} \sim N_{\boldsymbol{d}}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $A \in \mathbb{R}^{d \times d}$ is a $(d \times d)$ matrix, $R = \sqrt{\nu}/\sqrt{S}$, with $S \sim \chi_\nu^2$ (a chi-square distribution with $\nu > 0$ df), and $\boldsymbol{Z}$ and R, statistically independent random variables. In other words, the MV t distribution with ν df is constructed with the $C_{\nu,P}^{t}$ t copula function, which distributes a $[0,1]^d$ "probability layer" that binds together the d univariate Student *t* margins with the same degrees of freedom parameter ν.

By admitting separate modelling of marginals, Sklar's theorem enables a greater degree of flexibility in MV distribution construction. Univariate Student *t* distributions with different df parameters $\nu_1, \dots, \nu_d$ may be combined in the $C_{\nu,P}^{t}$ copula function to construct meta t_ν distribution functions. The density of the meta t_ν copula with *t*-distributed margins may be calculated from (47) and has the form (Ageeva, 2011, p. 14):

$$c_{\nu,P}^{\text{meta}-t}(\boldsymbol{u}) = |P|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\nu+\mathrm{d}}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{d-1}}{\Gamma\left(\frac{\nu+1}{2}\right)^{d}} \left(1 + \frac{\boldsymbol{z}' P^{-1} \boldsymbol{z}}{\nu}\right)^{-\frac{\nu+d}{2}} \prod_{i=1}^{d} \left(1 + \frac{\boldsymbol{z}_i^2}{\nu_i}\right)^{\frac{\nu_i+1}{2}}, \tag{50}$$

where

$$\boldsymbol{z} = \left(F_{t_{\nu_1}}^{-1}(u_1), \dots, F_{t_{\nu_d}}^{-1}(u_d)\right)' = \left(t_{\nu_1}^{-1}\left(T_{\nu_1}(Y_1)\right), \dots, t_{\nu_d}^{-1}\left(T_{\nu_d}(Y_d)\right)\right)'.$$

The form (50) introduces a class of new distributions termed, by Fang, et al. (2002), "multivariate

asymmetric *t*-distributions", which "enjoy certain symmetry, but the marginal degrees of freedom are different".

This dissertation uses the meta t_ν copula, with SPD-distributed margins, as the "density weighting function" in the multivariate structure to construct the combined assets and risk factors process. The function is given by:

$$c_{\nu,P}^{\text{meta}-t}(\boldsymbol{u}) = T_{\nu,P}^{d}\left(t_\nu^{-1}\left(F_{SPD_1}(x_1)\right), \dots, t_\nu^{-1}\left(F_{SPD_d}(x_d)\right)\right), \tag{51}$$

where $F_{SPD_i}(x_i)$ denotes the probability transform function for variable $i = 1, \dots, d$ where F_{SPD_i} is the semi-parametric form of the distribution function given by equation (35). The density $c_{\nu,P}^{\text{meta}-t}$ in (51) has the same functional form as that of equation (46), but with:

$$\boldsymbol{\varsigma}_i = t_\nu^{-1}\left(F_{SPD_i}(X_i)\right),$$

$$\boldsymbol{\varsigma} = \left(t_\nu^{-1}\left(F_{SPD_1}(X_1)\right), \dots, t_\nu^{-1}\left(F_{SPD_d}(X_d)\right)\right)'$$

and ν denoting the global degrees of freedom parameter.

The copula density in (51) is used to construct a d-dimensional "asymmetric" meta *t* distribution[88], with the joint density specified as:

$$f(x_1, \dots, x_d) = T_{\nu,P}^{d}\left(t_\nu^{-1}\left(F_{SPD_1}(x_1)\right), \dots, t_\nu^{-1}\left(F_{SPD_d}(x_d)\right)\right)\prod_{i=1}^{d} t_\nu(x_i), \tag{52}$$

where f_{SPD_i} denotes the PDF corresponding to the semi-parametric distribution of variable $i = 1, \dots, d$ and the t_ν in the product term denotes the univariate Student *t* density function operating on x_i constructed as $x_i = t_\nu^{-1}\left(F_{SPD_i}(x_i)\right)$ (i.e., constructed from the pseudo-copula data with the global df parameter).

The left panel in Figure 3.19 plots a standard Student *t* copula density surface fit on univariate *t* margins of the ALSI and ALBI log returns. The copula density surface fit is $c^{t}_{\hat{\nu}=5.659,\hat{\rho}_\tau=0.255}$. A similar surface could be constructed for the transformed uniform variates. The right panel plots corresponding Monte Carlo simulated copula density output for the two variables.

Figure 3.20 maps the surface of the meta Student *t* distribution and density constructed using the meta *t* copula fit to ARMA-GARCH-EVT filtered standardised ALSI and ALBI residuals (i.e., in the bivariate case) with Kendall's tau rank correlation parameter ρ_τ. The bivariate meta *t* distribution

[88] Asymmetric in the sense that, since the parameters in the marginal semi-parametric distributions govern asymmetric tail probabilities in the marginal distributions, the tail probabilities for the new class of meta-elliptical MV distributions constructed in this manner may be different in different directions (Fang, et al., 2002, p. 12).

surface is built using copula distribution $C^{t}_{\hat{\nu}=5.659,\hat{\rho}_{\tau}=0.255}$ (left). The bivariate meta *t* density surface is built using copula density $c^{t}_{\hat{\nu}=5.659,\hat{\rho}_{\tau}=0.255}$ (right).

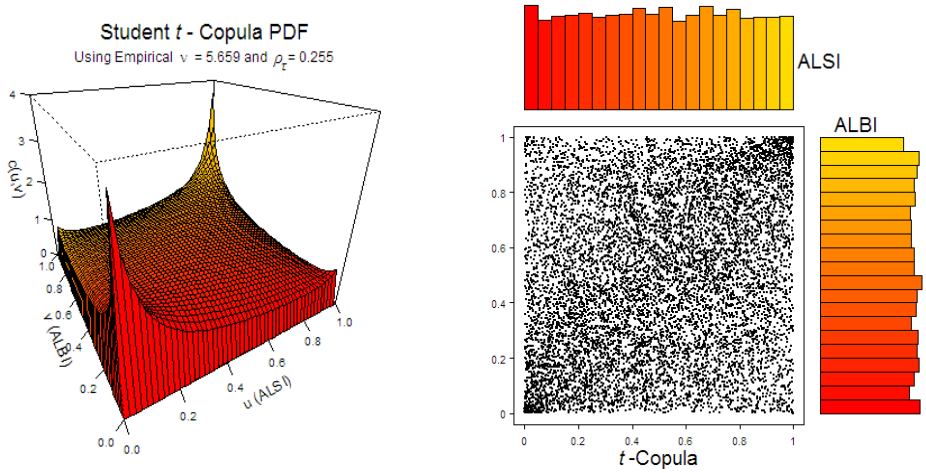


Figure 3.19: Empirical Standard Bivariate Student *t* Copula Density Surface and Corresponding Simulated Sample

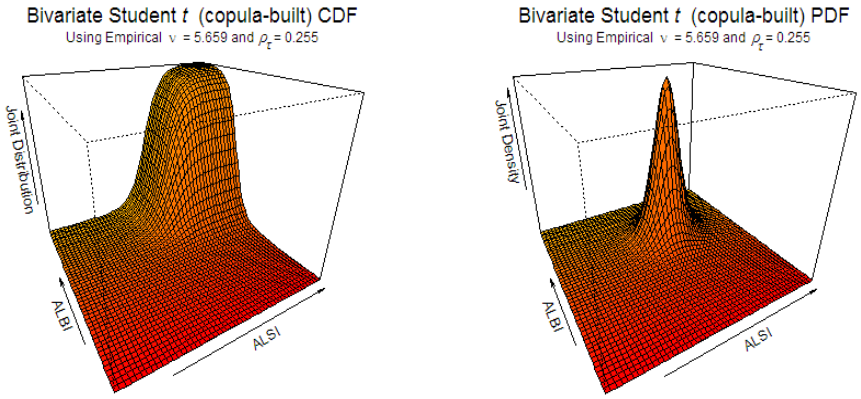


Figure 3.20: Bivariate Meta *t* Distribution and Density Surfaces

Figure 3.21 illustrates the practical differences between modelling within the Gaussian framework versus a heavier-tailed framework. The Gaussian model lacks a tail parameter to allocate density in the extremes and the dispersion is fairly symmetric as a function of the linear correlation matrix (or linear coefficient in the bivariate case) and symmetric Gaussian margins. The (symmetric) Student *t* copula allocates probability in both upper and lower tails. This is true even for the case where correlation is zero between variables. The figure plots 10,000 simulated points, each from the two multivariate distributions, with the empirical ALSI and ALBI data superimposed. In the left panel, a standard bivariate Gaussian distribution is constructed with Gaussian-fitted margins and a Gaussian copula, with correlation parameter $\hat{\rho} = 0.483$. In the right panel, a bivariate meta *t* distribution is constructed with SPD-fitted margins and a Student *t* copula $c^{t}_{\hat{\nu}=5.659,\hat{\rho}_{\tau}=0.255}$. Consistent with the claim in Fang, et al. (2002), it appears the meta *t* distribution does indeed exhibit a degree of distributional asymmetry, likely induced by admitting asymmetry via the EVT-based marginal models. The positive lower tail dependence, observed in the lower left region of the graph, may be interpreted as the

variables exhibiting similar behaviour in the lower tail. It represents an allocation of probability to both variables jointly producing losses in the same time period and, in terms of portfolio risk, reflects the opposite of diversification (i.e., a reduction in diversification gains). Accurate modelling of, in particular, positive lower tail dependence allows for potential improvement in diversification.

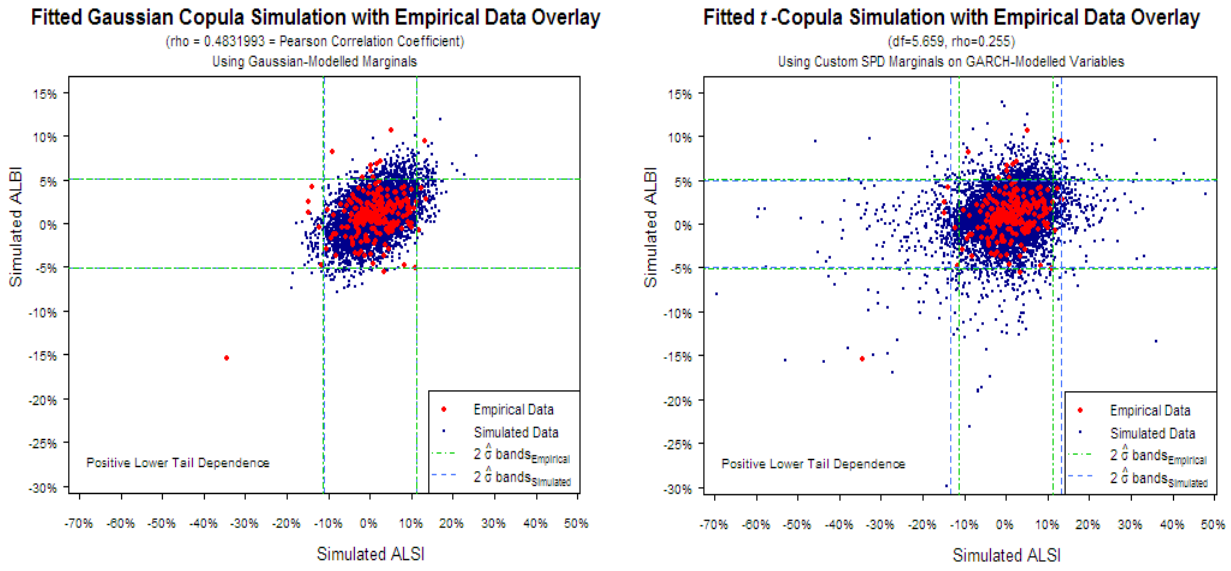


Figure 3.21: Simulation from a Bivariate Gaussian and Bivariate Meta *t* Distribution

3.5.2 Maximum Likelihood Based Estimation of the Student *t* Copula

Several methods of estimating copula parameters exist. However, this dissertation uses the canonical maximum likelihood (CML) method due to the method not requiring any a priori assumption on the distributional form of the margins. Typically, CML relies on the empirical distribution function (ECDF) to approximate the unknown parametric margins and the associated empirical probability integral transforms to obtain uniform margins required for copula fitting. Consequently, only the copula parameters need estimating, which significantly reduces complexity and computational burden in higher-dimensional modelling. This dissertation uses EVT-based margins instead of ECDFs due to the latter not able to extrapolate beyond the data.

Mashal and Zeevi (2002, p. 43) suggested an algorithm, similar to the following, to estimate the parameters ν and P of the Student *t* copula via the CML method.

Algorithm 1 (CML Estimation):

1. Transform the standardised residual data $\boldsymbol{Z} = (Z_{t,1}, \dots, Z_{t,d})$ for $t = 1, \dots, n$ to a "pseudo-sample" of uniform variates $\widehat{\boldsymbol{U}}^t = (\widehat{U}_1^t, \dots, \widehat{U}_d^t) = \left(\hat{F}_{Z_1}(Z_{t,1}), \dots, \hat{F}_{Z_d}(Z_{t,d})\right)$ using the respective probability transform functions (i.e., the SPD CDFs) from equation (35);

2. Estimate the correlation matrix $\hat{P}$ using a transformation of the non-parametric Kendall's tau estimated for each pair of bivariate margins of the copula:

$$\hat{P}_{ij} = \hat{\rho}_\tau(X_i, X_j) = \frac{2}{\pi}\arcsin(\hat{\rho}_{ij}) \approx \hat{\rho}_\tau(U_i, U_j), \qquad i, j = 1, \dots, d \tag{53}$$

where X_i, X_j represent a pair of original data vectors, $\hat{\rho}_{ij}$ denotes Kendall's tau for rank correlations and U_i, U_j represent a pair of probability-transformed vectors (see Figure 3.22);

3. Perform a numerical search for the remaining degrees of freedom parameter ν of the copula:

$$\hat{\nu} = \underset{\nu\in(2,\infty]}{\arg\max} \sum_{t=1}^{n} \ln\left(c\left(\widehat{\boldsymbol{U}}^t; \nu, \hat{P}\right)\right). \tag{54}$$

Kendall's tau, in equation (53), is a robust and efficient bivariate measure of dependence (or rank correlation) that is, further, invariant under co-monotonic transformations of the margins. The estimator may be defined as:

$$\begin{aligned}\rho_\tau(X_1, X_2) &= 4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x_1, x_2)dF(x_1, x_2) - 1 \\ &= 4\int_0^1 \int_0^1 C(u_1, u_2)dC(u_1, u_2) - 1 .\end{aligned} \tag{55}$$

As can be seen in equation (55), Kendall's tau measures dependency between the CDFs of the random variables X_1 and X_2 and does not depend on the actual random variables. The measure may be contrasted with the linear correlation measure, defined as:

$$\begin{aligned}\varrho(X_1, X_2) &= \frac{1}{\sigma_{X_1}\sigma_{X_2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[F(x_1, x_2) - F_1(x_1)F_2(x_2)]\,dx_1 dx_2 \\ &= \frac{1}{\sigma_{X_1}\sigma_{X_2}}\int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2]\,dF_1^{-1}(u_1)dF_2^{-1}(u_2),\end{aligned} \tag{56}$$

where σ_{X_i} denotes the standard deviation of random variable X_i. As can be seen in equation (56), dependency in the copula form of ϱ is a function of data-dependent distorting factors, the volatilities σ_{X_i}.

Figure 3.22 plots Kendall's τ estimates evaluated, firstly, on the 18 original log return data (dark grey dots) and, secondly, on uniform data transformed by the SPD CDFs fitted and applied to the 18 respective GARCH-standardised variables (green dots). The number of bivariate pairs evaluated are $d(d-1)/2 = 18(17)/2 = 153$ pairs. Reference point 2 of Algorithm 1, above.

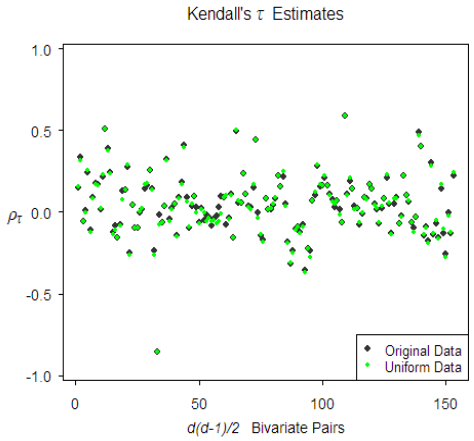


Figure 3.22: Kendall's Tau Estimates on Empirical Returns and Transformed Uniform Copula-Data

Under full MLE, the LLH function for the standard *t* copula in equation (46) is represented as:

$$\ln L(\nu, P; \hat{\boldsymbol{u}}) = \sum_{t=1}^{n} \ln f_{\nu,P}\left(t_\nu^{-1}(\hat{u}_{t,1}), \dots, t_\nu^{-1}(\hat{u}_{t,d})\right) - \sum_{t=1}^{n}\sum_{i=1}^{d} \ln f_\nu\left(t_\nu^{-1}(\hat{u}_{t,i})\right), \qquad (57)$$

where the second summand depends on the degrees of freedom parameter ν (Inanoglu & Ulman, 2009, p. 27).

For any the meta *t* distribution, the LLH equation for the meta *t* copula would change significantly. For example, consider:

$$\ln L(\nu, P; \hat{\boldsymbol{u}}) = \sum_{t=1}^{n} \ln f_{\nu,P}\left(g_1^{-1}(\hat{u}_{t,1}), \dots, g_d^{-1}(\hat{u}_{t,d})\right) - \sum_{t=1}^{n}\sum_{i=1}^{d} \ln g_i\left(g_i^{-1}(\hat{u}_{t,i})\right), \qquad (58)$$

where the marginal density functions g_i could be represented by any arbitrary, continuous univariate distribution. In equation (58), the second summand does not depend on the degrees of freedom parameter ν.

Inanoglu and Ulman (2009) noted the above differences in "full" MLE between the standard and meta *t* distribution estimators. The first summands on the right hand side of LLH functions (57) and (58) are essentially similar, since each sums the logarithms of a joint Student *t* distribution (although evaluated at different arguments over the domain). The second summand introduces a significant difference on the LLH estimation of parameter ν where, in equation (58), the second summand has no influence on the ν-estimator.

CML estimation admits a "hybrid" technique (with respect to standard *t* copula LLH estimation), where the issue of the second summand is addressed. Since the margins are estimated externally, only the first summand in the meta *t* copula depends on ν. The authors noted that "the same software functions cannot be employed to fit a standard *t* copula and to fit a meta *t* distribution (copula)" (p. 28). However, they also noted that, with the goal of simulation, sample data could be fit to a standard

copula rather than a meta-copula and the respective meta-copula simulated data obtained by appropriate quantile transforms on the standard copula-simulated probabilities (p. 31). The reasoning is as follows:

1. Sklar's theorem applied to
    a. standard copulas may be simplified as: $C(u_1, u_2) = C\big(F(x_1), F(x_2)\big) = F(x_1, x_2)$.
    b. meta-copulas (i.e., with arbitrary margins) as: $C(u_1, u_2) = C\big(G_1(y_1), G_2(y_2)\big) = F(y_1, y_2)$.
2. In terms of simulation, sample data from a meta-copula can be generated by applying the quantile transforms of the meta-copula's arbitrary margins to the probabilities u_i simulated from a standard copula.[89]
3. Fitting a standard *t* copula requires estimating parameters of a copula with identical Student *t* margins. The authors showed that non-parametric estimation of the margins by the empirical CDF (ECDF) imitates the "*t*-copula-*t*-margins" case.
4. By applying each ECDF to its corresponding sample data vector, the resulting matrix "contains a set of identical probabilities, but with different ordering in the respective column" (p. 31). The effect is that the non-parametric ECDF matrix emulates drawings from a set of identical *t* marginals.
5. The non-parametric pseudo-sample may be passed to a standard *t* copula fitting routine which solves for equation (57) and the df parameter and correlation parameters should provide appropriate standard *t* copula estimators.
6. Simulated log return data from the corresponding meta-copula may be obtained from the simulated probabilities of the standard copula (with reference to point 2 in the reasoning, above) by applying the appropriate quantile transform for each marginal distribution.

Similarly, it may be argued in this dissertation that the use of SPD estimation of the margins may serve a similar purpose to that of the use of ECDFs.

The LLH for the standard *t* copula is:

$$\ln L^{standard}(\nu, P; \widehat{\boldsymbol{u}}) = n \ln\left[\Gamma\left(\frac{\nu+d}{2}\right)\right] - n \ln\left[\Gamma\left(\frac{\nu}{2}\right)\right] - n \ln\left[|P|^{\frac{1}{2}}\right] - \left(\frac{\nu+d}{2}\right)\sum_{t=1}^{n} \ln\left(1 + \frac{\boldsymbol{z}_t' P^{-1} \boldsymbol{z}_t}{\nu}\right)$$
$$- \left\{ dn \ln\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - dn \ln\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \left(\frac{\nu+1}{2}\right)\sum_{t=1}^{n}\sum_{i=1}^{d} \ln\left(1 + \frac{\boldsymbol{z}_{it}'}{\nu}\right)\right\}, \qquad (59)$$

where

[89] Let random variables $X_1 \sim F_1$ and $Y_1 \sim G_1$, for F_1, G_1 continuous distribution functions, have C copula function. For another pair of continuous distribution functions F_2, G_2, set $X_2 = F_2^{-1}\big(F_1(X_1)\big)$ and $Y_2 = G_2^{-1}\big(G_1(Y_1)\big)$. Then (a) $X_2 \sim F_2$ and $Y_2 \sim G_2$ and (b) the copula of X_2 and Y_2 is C (Inanoglu & Ulman, 2009, pp. 30-31). In the framework of this dissertation, denote by F_2, G_2 a pair described by the SPD functional form in equation (35).

$$\boldsymbol{z} = \left(F_{t_\nu}^{-1}(u_1), \ldots, F_{t_\nu}^{-1}(u_d)\right)'.$$

The software routine[90] used to calibrate the meta *t* copula in this study optimises the standard copula LLH in equation (59). The fitted *t* copula is used to construct the MV *t* distribution function, with the density given in equation (52). Simulated log returns from the standard copula-fitted MV *t* distribution are then:

1. Transformed back to uniform (probability) data by the univariate Student *t* distribution.
2. Transformed forward to quantile data (i.e., log returns) by the respective SPD transforms, hereby forming simulated returns from the meta *t* distribution.

Equivalently, simulated probabilities from the fitted standard *t* copula (as opposed to the distribution) may directly be transformed to quantile data per step 6 above.

Note that there is no guarantee of positive definiteness in the componentwise transformed (53) matrix of Kendall's tau rank correlation coefficients[91]. Positive definiteness is a desired quality in practice, since it admits tractable matrix operations, such as inversions and decompositions (e.g., Cholesky decomposition). If $\hat{P}$ is not positive definite, there exist adjustment techniques to transform the matrix such that it is positive definite and close to the original matrix. The eigenvalue method is one such technique[92]. McNeil, et al. (2005, p. 231) suggested an algorithm to implement the eigenvalue method to obtain a positive definite matrix that is close to $\hat{P}$ and which is implemented, if necessary, in the software routine.

Algorithm 2 (Eigenvalue Method). Let P^* be a so-called *pseudo*-correlation matrix (i.e., a symmetric matrix of pairwise correlation estimates with unit diagonal entries and off-diagonal entries in $[-1, 1]$) that is not positive semi-definite.

1. Calculate the spectral decomposition $P^* = GLG'$, where L is the matrix of eigenvalues and G is an orthogonal matrix whose columns are eigenvectors of P^*.
2. Replace all negative eigenvalues in L by small values $\delta > 0$ (for positive definiteness) or zeros (for positive semi-definiteness) to obtain $\tilde{L}$.
3. Calculate $\tilde{P} = G\tilde{L}G'$, which will be symmetric and positive definite (or positive semi-definite), but not a correlation matrix, since its diagonals will not necessarily equal one.

[90] The function chosen to fit the standard *t* copula is the `fitCopula` function from the `copula` package in `R` (Hofert, Kojadinovic, Maechler & Yan, 2014).
[91] A correlation matrix is a symmetric, positive definite matrix with unit diagonal entries and off-diagonal entries in $[-1, 1]$. A matrix P is positive definite if $x^t P x > 0 \; \forall \; x \neq 0$ and positive semi-definite if $x^t P x \geq 0 \; \forall \; x \neq 0$.
[92] For symmetric matrices, positive definiteness is equivalent to having all eigenvalues positive; positive semi-definiteness is equivalent to having all eigenvalues non-negative.

4. Return the correlation matrix $P^{\dagger} = \wp(\tilde{P})$, where $\wp$ denotes the correlation matrix operator defined as:

$$\wp(\tilde{P}_{ij}) := \frac{\tilde{P}_{ij}}{\sqrt{\tilde{P}_{ii}\tilde{P}_{jj}}} \qquad i, j = 1, \dots, d.$$

Finally, it may be noted that an "unstructured" dispersion structure is chosen to characterise the structure of the dispersion matrix in the copula model. Other options, considered inadmissible or too restrictive in the current setting, include: autoregressive of order 1 (diminishing correlation), exchangeable (correlation does not vary) and Toeplitz (correlations between a number of different variables are the same). The imposition of the "unstructured" structure admits unique correlation estimates for each pair of fitted variables. The four options, illustrated for dimension $d = 3$, may be shown as follows:

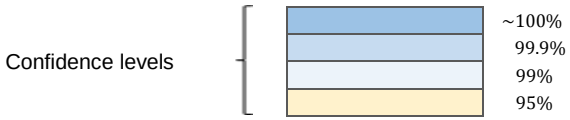
$$\begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}_{\text{unstructured}}, \begin{pmatrix} 1 & \rho_1 & \rho_1^2 \\ \rho_1 & 1 & \rho_1 \\ \rho_1^2 & \rho_1 & 1 \end{pmatrix}_{AR(1)}, \begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_1 & 1 \end{pmatrix}_{\text{exchangeable}} \text{and} \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix}_{\text{Toeplitz}}.$$

This chapter concludes with output in Table 3.11. The table displays CML estimates of the MV Student *t* copula fit to SPD-transformed uniform data. The degrees of freedom parameter estimate and dispersion matrix parameter estimates are given, with corresponding confidence levels highlighted. The degrees of freedom parameter $\hat{\nu} = 49.0149$ is estimated with a high degree of confidence. Although heavier joint tail dependence than Gaussian copula (where $\nu = \infty$), the estimate is not very low. The two primary reasons suspected for this are:

1. Generally lighter upper tails in each of the variables acts to lower the aggregate joint dependence estimate by way of the symmetric structure imposed on the data by the Student *t* copula (i.e., by design, aggregate light upper tail dependence acts to lighten, or mitigate, aggregate heavier lower tail dependence).
2. Consistent with the philosophy of diversification, the degree of co-variation in a portfolio is expected to decrease as the number of assets in the portfolio increases. As a result, the degree of joint dependence may be expected to lighten as more assets and risk factors to the portfolio are added to the model.

Table 3.11: CML Estimates of the Fitted Student *t* Copula Degrees of Freedom Parameter and Dispersion Matrix Parameters

The CML estimated standard *t* copula degrees of freedom parameter $\hat{v} = 49.0149$ (confidence level: 99.9%)																		
$\hat{P}_{\text{standard } t}$	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}	U_{12}	U_{13}	U_{14}	U_{15}	U_{16}	U_{17}	U_{18}
U_1	1	0.2156	0.4538	-0.1109	-0.0146	0.3594	-0.1752	0.1459	0.2533	0.2615	0.0780	0.3815	0.7180	0.5472	0.4072	-0.1824	-0.1874	-0.2173
U_2	0.2156	1	-0.2047	0.0727	0.2509	0.4714	-0.4502	0.1429	-0.1330	-0.1602	0.0587	0.0615	0.2865	0.2367	0.4000	0.2466	-0.4656	-0.9694
U_3	0.4538	-0.2047	1	-0.1241	-0.1344	-0.0066	0.4938	-0.0776	-0.0068	0.0571	-0.2213	0.1053	0.2331	0.5547	0.0980	-0.1756	0.1324	0.2242
U_4	-0.1109	0.0727	-0.1241	1	0.0096	-0.1068	0.0346	-0.0270	-0.0706	-0.0414	-0.0680	-0.0762	-0.1222	-0.1190	-0.0413	0.1131	0.2354	-0.0263
U_5	-0.0146	0.2509	-0.1344	0.0096	1	0.1869	-0.2762	0.7123	0.1214	0.1012	0.3885	0.2088	0.0821	0.0134	0.2808	0.6667	-0.0937	-0.2682
U_6	0.3594	0.4714	-0.0066	-0.1068	0.1869	1	-0.3454	0.1726	0.0480	0.0384	0.1363	0.1503	0.3986	0.2736	0.3608	0.0752	-0.3655	-0.5153
U_7	-0.1752	-0.4502	0.4938	0.0346	-0.2762	-0.3454	1	-0.4133	-0.2180	-0.1658	-0.1817	-0.1794	-0.5363	-0.3056	-0.4113	0.0234	0.2364	0.4952
U_8	0.1459	0.1429	-0.0776	-0.0270	0.7123	0.1726	-0.4133	1	0.2761	0.2696	0.3742	0.3183	0.2397	0.1197	0.1552	0.1782	-0.0727	-0.1841
U_9	0.2533	-0.1330	-0.0068	-0.0706	0.1214	0.0480	-0.2180	0.2761	1	0.7952	0.1793	0.3383	0.2478	0.0465	0.0820	-0.0998	-0.0383	0.1122
U_{10}	0.2615	-0.1602	0.0571	-0.0414	0.1012	0.0384	-0.1658	0.2696	0.7952	1	0.1482	0.3179	0.2540	0.0913	0.0633	-0.1244	0.0468	0.1211
U_{11}	0.0780	0.0587	-0.2213	-0.0680	0.3885	0.1363	-0.1817	0.3742	0.1793	0.1482	1	0.3961	0.1157	-0.1659	0.1336	0.1525	-0.1351	-0.0992
U_{12}	0.3815	0.0615	0.1053	-0.0762	0.2088	0.1503	-0.1794	0.3183	0.3383	0.3179	0.3961	1	0.3763	0.1750	0.2085	-0.0611	-0.1901	-0.0667
U_{13}	0.7180	0.2865	0.2331	-0.1222	0.0821	0.3986	-0.5363	0.2397	0.2478	0.2540	0.1157	0.3763	1	0.6585	0.6103	-0.1774	-0.1880	-0.3178
U_{14}	0.5472	0.2367	0.5547	-0.1190	0.0134	0.2736	-0.3056	0.1197	0.0465	0.0913	-0.1659	0.1750	0.6585	1	0.4613	-0.1725	-0.0726	-0.2404
U_{15}	0.4072	0.4000	0.0980	-0.0413	0.2808	0.3608	-0.4113	0.1552	0.0820	0.0633	0.1336	0.2085	0.6103	0.4613	1	0.2493	-0.1518	-0.4245
U_{16}	-0.1824	0.2466	-0.1756	0.1131	0.6667	0.0752	0.0234	0.1782	-0.0998	-0.1244	0.1525	-0.0611	-0.1774	-0.1725	0.2493	1	-0.0878	-0.2424
U_{17}	-0.1874	-0.4656	0.1324	0.2354	-0.0937	-0.3655	0.2364	-0.0727	-0.0383	0.0468	-0.1351	-0.1901	-0.1880	-0.0726	-0.1518	-0.0878	1	0.4254
U_{18}	-0.2173	-0.9694	0.2242	-0.0263	-0.2682	-0.5153	0.4952	-0.1841	0.1122	0.1211	-0.0992	-0.0667	-0.3178	-0.2404	-0.4245	-0.2424	0.4254	1



3.5.3 Simulation

3.5.3.1 Simulation from the Meta Student *t* Distribution

Recall the centred Student t distributed random vector $\boldsymbol{Y} = (Y_1, \dots, Y_d)'$. It is well known that the centred MV t distribution belongs to the class of multivariate normal variance mixture distributions and can be defined by the stochastic representation per equation (49), $\boldsymbol{Y} = RA\boldsymbol{Z}$. From Sklar's theorem, the distribution constructed in this study is characterised by:

- Its marginal CDFs $SPD_1, \dots, SPD_d$ per equation (35);
- Its copula $c_{\nu,P}^{\text{meta}-t}$ per equation (51).

Given the MV t stochastic (mixture) representation and Sklar's characterisation, a simulation algorithm (one of a number of possible approaches) may be implemented to generate random samples $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(d)}$ from a MV *t* distribution. Denote the dimension of this distribution by d and by $b_j^{(i)}$ the j-th component of $\mathbf{b}^{(i)}$, for $j = 1, \dots, n$ number of observations to be generated (Carmona, 2014, p. 162). The implemented simulation algorithm follows from Cherubini, et al. (2004, p. 181), but is detailed to be consistent with the software implementation of the algorithm:

1. Find the Cholesky factor $A \in \mathbb{R}^{d\times d}$ of $\hat{P}_{\text{standard } t}$, an upper triangular matrix such that $A'A = \hat{P}$ (equivalently a lower triangular matrix $L \in \mathbb{R}^{d\times d}$ such that $LL' = \hat{P}$).
2. Simulate d i.i.d. column vectors $\mathbf{z} \in \mathbb{R}^{n\times d} = (z_1, z_2, \dots, z_d)'$ from a standard normal $N(0,1)$ distribution.
3. Simulate n random variates S from a $\chi^2_{\hat{\nu}}$ distribution with copula estimated $\hat{\nu}$, independent of $\mathbf{z}$[93].
4. Set $R = \sqrt{\hat{\nu}}/\sqrt{S}$.
5. Set $\boldsymbol{X}_{(n\times d)} = \boldsymbol{Z}_{(n\times d)}A_{(d\times d)}$.
6. Set $\boldsymbol{Y} = R\boldsymbol{X}$.
7. Set $u_i = T_{\hat{\nu}}(y_i)$ for $i = 1, \dots, d$, where $T_{\hat{\nu}}$ denotes the univariate Student *t* distribution function.
8. $\boldsymbol{U} = (u_1, \dots, u_d)' = \left(T_{\hat{\nu}}(Y_1), \dots, T_{\hat{\nu}}(Y_d)\right)'$.
9. Set $b_i = t_{\hat{\nu}}^{-1}(u_i)$ for $i = 1, \dots, d$, where $t_{\hat{\nu}}^{-1}$ denotes the univariate Student *t* quantile function.

[93] The standard procedure to generate a sequence of χ^2 distributed random variables uses the relationship between a χ^2 and gamma distribution. If random variables $X_1, \dots, X_n \sim N(0,1)$ and are independent, then $Y = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$, where $X \sim Y$ denotes that random variables X and Y have the same distribution. The PDF of the χ^2 distribution, $f_{\chi^2_\nu}(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1}e^{-x/2}$ for $x \geq 0$ (zero otherwise) and degrees of freedom parameter $\nu \geq 0$, is a special case of the gamma PDF $\Gamma(\alpha, \beta)$, given by $f_{\Gamma(\alpha,\beta)}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta}$ for shape $\alpha > 0$, scale $\beta > 0$ and $x \geq 0$ (zero otherwise). The relationship $\chi^2_\nu \sim \Gamma(\nu/2, 2)$ is used to generate χ^2_ν variables from the $\Gamma(\nu/2, 2)$ distribution.

10. Return a d-dimensional matrix of simulated standardised residuals $\boldsymbol{B} = (b_1, \dots, b_d)'$.

With reference to Inanoglu and Ulman (2009, p. 31), the following transformation backwards and forwards serves to introduce a degree of idiosyncrasy into the meta *t* distribution:

11. Set $\boldsymbol{U} = \left(T_{\hat{\nu}}(b_1), \dots, T_{\hat{\nu}}(b_d)\right)' = (u_1, u_2, \dots, u_d)'$ to retrieve the output from Step 8.
12. Set $\check{z}_i = f_{SPD_i}^{-1}(u_i)$ for $i = 1, \dots, d$, where $f_{SPD_i}^{-1}$ denotes the univariate SPD quantile function.
13. Return the desired d-dimensional matrix of simulated standardised residuals $\check{\boldsymbol{Z}} = (\check{z}_1, \dots, \check{z}_d)'$.

The Cholesky factor of the dispersion matrix (i.e., the dependence structure associated with the copula) induces row-wise rank correlation in the simulated variables matrix. This preserves the non-linear MV co-variation, including that of joint tail events. The upshot is that simulated outcomes are independent in time (i.e., each column represents a univariate i.i.d. stochastic process when viewed in isolation), but are dependent at any point in time (i.e., each row shares the rank correlation induced by the copula).

The left panel of Figure 3.23 plots simulated standardised residuals extracted from the first column (i.e., the column corresponding to the ALSI variable) of two different distributions, the MV standard *t* distribution and the MV meta *t* distribution: $\mathbf{b}^{(1)} \subset \boldsymbol{B}$ (black dots) and $\check{\boldsymbol{z}}^{(1)} \subset \check{\boldsymbol{Z}}$ (blue dots), respectively. The MV standard *t* distribution is constructed with a standard *t* copula and standard *t* margins. The MV meta *t* distribution is constructed with a standard *t* copula and SPD margins. There is a large degree of overlap between the two simulations in the interior portion of the distributions (between the two standard deviation bands). As one moves into the tails, a marked difference is observed. Above the upper $2\hat{\sigma}$ threshold, there appears to be a commensurate, or only slightly different, amount of density allocated by each MV distribution. Below the $2\hat{\sigma}$ threshold, however, a significant degree of asymmetry materialises: the SPD_1 extrapolates further into the lower tail than does the t_ν distribution. The right panel of Figure 3.23 plots the first 242 (i.e., the sample set size) simulated standardised residuals from the meta *t* distribution $\check{\boldsymbol{z}}^{(1)} \subset \check{\boldsymbol{Z}}$ (blue dots) over the "empirical" ALSI GARCH-fit standardised residuals (red dots). A high degree of cohesiveness between the data points is observed. This cohesiveness, or consistency, is embodied all through the data points simulated from the meta *t* distribution.

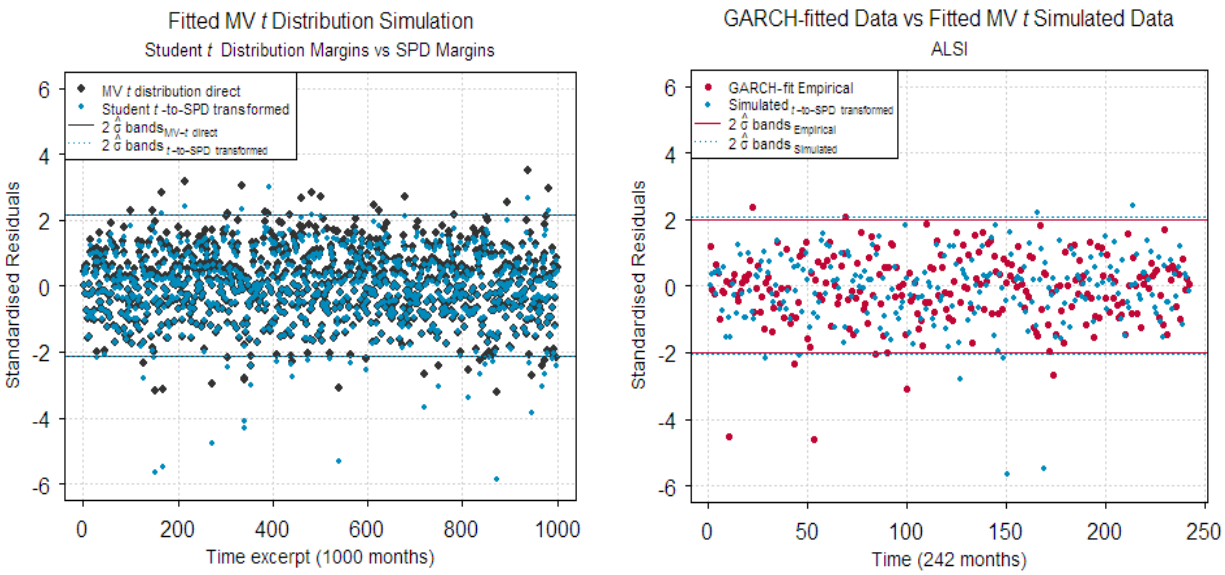


Figure 3.23. Comparison of Standardised Residuals Simulated from Standard and Meta *t* Distributions

Each variable's SPD quantile function $f_{SPD_i}^{-1}$ transforms the respective copula-simulated uniform vector to a vector of standardised residuals consistent with those standardised residuals obtained from the ARMA-GARCH filtering process of Chapter 3.3. The resulting simulated i.i.d. noise process defines a customised sample density used as input to the univariate simulation model when reintroducing the autocorrelation and heteroskedasticity characteristics observed in each variable's original returns.

3.5.3.2 Simulation from the Univariate ARMA-GARCH Models

Univariate, idiosyncratic autocorrelation and heteroskedasticity is reintroduced to each variable through a final round of simulation. The $f_{SPD_i}^{-1}$ transformed vector $\breve{\boldsymbol{z}}^{(i)} \subset \breve{\boldsymbol{Z}}$ of copula-dependent innovations forms an innovation density sample which is input to the i-th fitted GARCH model ($i = 1, \ldots, d$). The innovation density mimics the corresponding conditional density fitted in the GARCH model[94], while preserving the correlation induced by the copula. An n number of monthly log return data is then simulated, where $n_{\text{univariate}} = n_{\text{multivariate}} = 1{,}200{,}000$.

Typically, a Monte Carlo simulation involves generating independent, random draws from a fitted probabilistic model. For time series models, the random draw is represented by a realisation of a sample path of defined length n, which may be generated any number of times m. In this section, m is set to one so that the corresponding copula-simulated innovations may be used.

[94] This assumption was confirmed by fitting each copula-simulated vector (i.e., each simulated variable) with the distribution corresponding to that variable's GARCH-model, conditional distribution (i.e., those from column 7 in Table 3.8). Shape and skewness estimates across all variables confirmed that the copula-simulated innovation densities do indeed mimic their corresponding conditional densities in the GARCH models (i.e., copula-simulated innovation shape and skewness estimates were close matches to their GARCH-fitted counterparts from columns 16, 17 and 18 in Table 3.8).

Consider the typical GARCH process used in this dissertation in terms of simulating for the period ahead:

$$y_{t+1} = \mu + \varepsilon_{t+1}\,, \tag{60}$$

where

$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\,, \tag{61}$$

$$z_{\forall t} \sim iid\ \mathcal{D}_{\vartheta}(0,1)\,, \tag{62}$$

and

$$\sigma_{t+1}^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2\,. \tag{63}$$

The models are fit on information up to time t. The simulation uses the model parameter estimates to simulate $t+1$. The subsequent conditional variance $\hat{\sigma}_{t+1}^2$ is generated recursively, using the fitted conditional variance model. A starting value for the ε_t^2 term in equation (63) is set to the model's last known estimated value $\hat{\varepsilon}_t^2$ at time t. Similarly, the last observation in the dataset is used as the initial value for y_t in equation (60). The starting value $\hat{\sigma}_t^2$ is output from the fitted model and model estimates at $t-1$ by $\hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2$. This initialising step obtains $\hat{\sigma}_{t+1}^2$. At time $t+2$, however, a value for ε_{t+1}^2 cannot be derived from the estimated model (which stops at time t). At this point, instead of simulating a random draw z_{t+1} from the model's $iid\ \mathcal{D}_{\vartheta}(0,1)$ distribution (as is typically done), the z_{t+1} value is replaced with the associated $\check{z}_{t+1}$ value from the copula-simulated innovation series. An estimate $\hat{\varepsilon}_{t+2}^2$ is obtained, per equation (61), by multiplying the copula-simulated value $\check{z}_{t+1}$ by the estimated $\hat{\sigma}_{t+1}$ derived in the initialising step. In this manner and from this point forward in the simulation, all subsequent realisations of the GARCH process[95] are recursively generated by:

$$\hat{\sigma}_{t+i}^2 = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_{t+j}^2 + \hat{\beta}_1 \hat{\sigma}_{t+j}^2\,, \tag{64}$$

for $i = 3, \ldots, n$ evaluated in lockstep with $j = 2, \ldots, n-1$. At each step, the estimated $\hat{\varepsilon}_{t+i}$ value feeds through to equation (60), producing a full n-length simulated return series.

The matrix of simulated monthly log returns is obtained and aggregated (in this case annualised) for use in conditioning on anticipated economic scenarios. The forecast horizon is set to yearly, given the granularity of the data. Higher-frequency data will, naturally, lend itself better to shorter forecast horizons. The gain from this current section of work is that non-linear co-variation relationships have been preserved among the variables in the resulting simulated matrix and the matrix is ready for conditioning on scenarios and portfolio optimisation.

[95] Monte Carlo simulations from other GARCH-family models are implemented similarly, using the model's corresponding conditional variance equation and copula-simulated standardised residuals.

3.6 Portfolio Optimisation

Malevergne and Sornette (2006, p. 275) noted: “The process of decision-making under conditions of deep uncertainty requires first to consider ensembles of scenarios, then to seek robust and adaptive strategies, and finally to combine machine and human capabilities interactively. Outstanding questions involve the compromise between near-term objectives and long-term sustainability, and the characterisation of irreducible risks and of ‘surprises’.”

The Monte Carlo simulation engine of Chapter 3.5.3 enables practitioners to comprehensively “consider ensembles of scenarios”, after each of which the portfolio may be optimised, response plans crafted and/or specific risks hedged.

3.6.1 Scenario Generation

Ensembles of scenarios may be constructed using subjective views and/or expert forecasts for next-period returns on any set of modelled variables. For example, Table 3.12 demonstrates the use of range forecasts for key variables believed to be highly relevant to the portfolio over the forecast horizon. Range forecasts are more realistic than point forecasts in the sense that they admit an aggregate view of the future. Expert opinions may be aggregated and may span a realistic range of anticipated returns for a number of important variables. For example, a portfolio or risk manager may value the opinions of several oil market experts, each of which have different return expectations for the oil price over the forecast period. These opinions may be aggregated to form a range of returns on the variable and, combined with ranges of returns on other key variables, may be used to define a future economic scenario. In the case of Table 3.12, point returns on select variables were extracted from actual (but out-of-sample) returns experienced over the period 30 April 2014 to 30 April 2015. A range encapsulating each variable’s point return was constructed and used, as if one had perfect foresight for the variables. Furthermore, any number of scenarios may be defined, including hypothetical scenarios not contained in the history of the data. The modelling process admits extrapolation into the tails both individually and jointly and, thereby, admits the evaluation of extreme moves in any variable of choice (or combinations of extreme moves in different variables). Worth noting is that wider range forecasts are preferred, in that they yield more possible joint scenarios when evaluating multiple variables.

Although the model in this dissertation is constructed using asset classes and return drivers, it may be built on or extended laterally to include, for example, relative valuation measures derived from financial statements[96] or technical indicators derived from generic price, volume or open interest

[96] Fundamental business measures, such as the degree of operating leverage (% change in EBIT/% change in Sales), price-to-cash flow, price-to-book and price-to-earnings ratios.

activity[97].

Table 3.12: Risk Factor Range Forecasts for Out-of-Sample Year-Ahead Horizon

INPUT VARIABLES (i.e., Return Drivers)	EXPECTED SCENARIOS (i.e., Range Forecasts)
These are predetermined risk factors expected to play a key role over the forecast horizon	These are subjective or expert views that provide a range basis for returns over the forecast horizon
USDZAR	Rand depreciates between +10% to +15%
EURUSD	Dollar strengthens between -15% to -30%
Domestic 10-yr Govt Bond	SA govt bond yields decrease 0% to -10%
Oil Price	Oil price falls -35% to -55%
GSCI (commodities basket)	Global commodities prices fall -30% to -50%
Gold	Gold price falls 0% to -20%
Platinum	Platinum price falls -15% to -30%
Emerging Market Equities	EM equities remain stable -5% to +15%
Developed Market Equities	S&P 500 advances +5% to +15%

Under the assumption of perfect foresight, the next step evaluates the joint effect of the forward-looking scenario on asset classes in the portfolio. Table 3.13 illustrates the 2014-to-2015 out-of-sample joint scenario (corresponding to that in Table 3.12) to which a hypothetical portfolio of asset classes will be optimised. A basic, domestic, diversified portfolio (i.e., the output) is optimised, given the pre-specified backdrop scenario (i.e., the input).

Table 3.13: Risk Factor Variables Describing a Forward-Looking Scenario and Portfolio Variables to be Optimised

INPUT VARIABLES (i.e., Return Drivers)	OUTPUT VARIABLES (i.e., Asset Classes)
Range forecasts over subsequent period	Asset classes to be conditioned and optimised
Rand depreciates (+10% to +15%)	FTSE/JSE All Share Index
Dollar strengthens (-15% to -30%)	FTSE/JSE All Bond Index
SA10-yr yields fall (0% to -10%)	MSCI World Equity Index
Oil price falls (-35% to -55%)	Domestic Money-market Cash Index
Commodities prices fall (-30% to -50%)	
Gold price falls (0% to -20%)	
Platinum price falls (-15% to -30%)	
EM equities stable (-5% to +15%)	
S&P 500 advances (+5% to +15%)	

[97] Technical measures, such as the accumulation/distribution ratio, on-balance volume measure and relative strength index.

The matrix of simulated returns in Chapter 3.5.3.2 is filtered, row-wise and jointly, through the forecast ranges describing the anticipated scenario. For each iteration where the input variables' range forecasts are collectively met, the corresponding set of portfolio asset class returns is extracted. Out of the matrix of 100,000 simulated annual returns, 138 yearly observations jointly corresponded to the input scenario and were extracted (refer to Figure 2.13). This process builds a conditioned matrix of returns, with each resulting output variable expected to exhibit different distributional characteristics to the respective variable's historical profile. Figure 3.24 highlights an example of the distributional differences. The left panel plots a fitted JSU distribution to empirical ALSI (annualised) data over a JSU distribution fit to annualised ALSI conditioned data from the 2014-2015 scenario. In this instance, the left tail (resp. right tail) of the historical data is fatter (resp. thinner) than that of the conditioned data. The right panel plots a fitted JSU distribution to annualised empirical JSE Listed Property data over the conditioned variable's annualised JSU-fitted data. In this instance, the converse is true: the right tail (resp. left tail) of the historical data is fatter (resp. thinner) than that of the conditioned data. The graphs illustrate the differences in not only the bulk and skewness of the returns, but also in the upper and lower tails. Similarly, in the multivariate case, the conditioned matrix of returns reflects a new co-dependence pattern. These new characteristics lead to portfolios that generate a different efficient frontier than those arising when optimising purely on historical data.

In practice, the models used up to this point would be refit on a regular basis so as to include new information contained in recent data. Doing so may materially benefit investors since, for example, GARCH models are designed to account for changes in volatility levels as the changes occur.

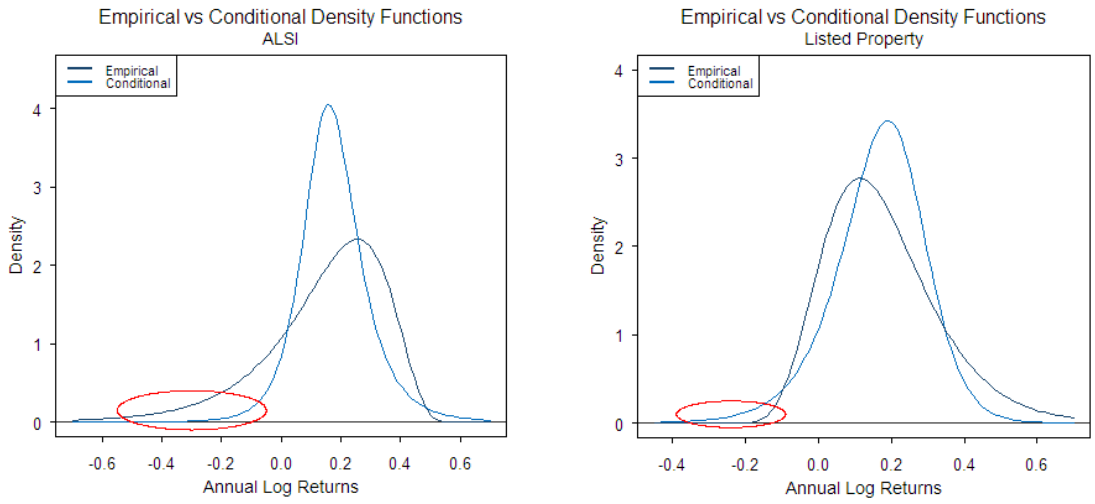


Figure 3.24: Annualised Historical and Scenario-Conditioned ALSI and Listed Property Data

3.6.2 Portfolio Optimisation

Markowitz (1952) established the (implicitly Gaussian) framework and fundamental principles on which portfolio optimisation is based: risk, return and correlation. Combined, these principles

singularly unlock the key proposition that is diversification, itself arguably the most important tenet in portfolio theory. This section compares Markowitz's historical mean-variance optimisation results to the simulation-based results from the previous section optimised using mean-CVaR.

3.6.2.1 Mean-Variance Optimisation

This section references Fabozzi, et al. (2007, pp. 24-26) and Merton (1972).

The objective of the portfolio optimisation problem in the MVO approach is formulated as follows: given a set of financial assets, characterised by their expected return and their covariances, find the optimal weight of each asset such that, combined, the resulting portfolio reflects the lowest covariance risk per given target return. What emerges as a solution to the problem is an "efficient frontier", which is the set of all feasible portfolios that offers the highest rate of portfolio return across given levels of portfolio risk. The emergence of a frontier implies that, for each level of risk, there is exactly one achievable combination of asset weights offering the highest rate of return.

A portfolio is an amount of wealth invested in a set of N weights w_i corresponding to a group of assets A_i, for $i = 1,2, \dots, N$. Weights represent percentages[98] in an N-vector $\boldsymbol{w} = (w_1, \dots, w_N)'$, where:

$$\sum_{i=1}^{N} w_i = 1 \,, \tag{65}$$

implying that the portfolio is fully invested.

Define the return on A_i by R_i. Portfolio investments may be described by:

$$\sum_{i=1}^{N} w_i R_i \; = R \,. \tag{66}$$

In practice, the portfolio is characterised by expected returns:

$$\hat{\mu}_P = E(R) = E\left(\sum_{i=1}^{N} w_i R_i\right) = \sum_{i=1}^{N} w_i \hat{\mu}_i \;, \tag{67}$$

where $\hat{\mu}_i = E(R_i)$ is the expected return value of A_i. Written compactly, $\hat{\mu}_P = \boldsymbol{w}'\widehat{\boldsymbol{\mu}}$, for $\widehat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_N)'$.

An estimate $\widehat{\boldsymbol{\Sigma}}$ of the covariance matrix of the portfolio is given by:

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1N} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{N1} & \cdots & \hat{\sigma}_{NN} \end{pmatrix}, \tag{68}$$

where $\hat{\sigma}_{ij} = \hat{\sigma}_{ji}$ denotes the covariance between returns on assets A_i and A_j such that $\hat{\sigma}_{ii} = \hat{\sigma}_i^2$, $\hat{\sigma}_{ij} =$

[98] Implying the N assets are infinitely divisible.

$\hat{\rho}_{ij}\hat{\sigma}_i\hat{\sigma}_j$ and $\hat{\rho}_{ij}$ is the coefficient of Pearson's linear correlation between returns on assets A_i and A_j.

The variance-covariance risk of the portfolio is given by:

$$\hat{\sigma}_P^2 = E(|R - \hat{\mu}_P|^2) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \hat{\sigma}_{ij} = \boldsymbol{w}'\hat{\boldsymbol{\Sigma}}\boldsymbol{w}\,. \qquad (69)$$

The problem of portfolio selection is a constrained minimisation problem, formulated as follows:

$$\boldsymbol{w} = \arg\min_{\boldsymbol{w}} \boldsymbol{w}'\hat{\boldsymbol{\Sigma}}\boldsymbol{w}\,, \qquad (70)$$

subject to constraints:

$$\boldsymbol{w}'\hat{\boldsymbol{\mu}} = \bar{r}$$

$$\boldsymbol{w}'\mathbf{1} = 1\,,$$

where $\bar{r}$ is the portfolio target return and $\mathbf{1} = (1,1,\ldots,1)'$.

The problem of minimising the variance-covariance risk $\hat{\sigma}_P^2$ for a given target return $\bar{r}$ and fully invested portfolio $\boldsymbol{w}'\mathbf{1} = 1$ is a quadratic programming problem with linear constraints[99].

For a given target return, the weight vector in equation (70) has a unique solution:

$$\boldsymbol{w}^* = \bar{r}\boldsymbol{w}_0^* + \boldsymbol{w}_1^*\,, \qquad (71)$$

where

$$\boldsymbol{w}_0^* = \frac{1}{ab - c^2}\left(b\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} - c\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\right),$$

with

$$\boldsymbol{w}_1^* = \frac{1}{ab - c^2}\left(c\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} - a\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\right),$$

$$a = \hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\,,$$

$$b = \mathbf{1}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\,,$$

$$c = \mathbf{1}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\,.$$

The corresponding standard deviation $\hat{\sigma}_P$ for the optimal portfolio with weights $\boldsymbol{w}^*$ is calculated as:

$$\hat{\sigma}_P = \sqrt{\frac{1}{ab - c^2}(b\bar{r}^2 - 2c\bar{r} + a)}\,, \qquad (72)$$

where

[99] There is an alternative formulation, where return is maximised for given levels of risk. In the latter case, the problem becomes one of optimising a linear objective function with quadratic constraints and requires a different class of solver, with more complexity in the algorithm.

$$\bar{r} = (\boldsymbol{w}^*)'\hat{\boldsymbol{\mu}}\,. \qquad (73)$$

Equations (72) and (73) are used to construct a hyperbola with a locus of points in the $(\hat{\sigma}_P, \bar{r})$-space. The efficient frontier (or "efficient set") is the set of portfolios along the border of the hyperbola. The "feasible set" is that set of mean-variance portfolios inside the hyperbola (i.e., sub-optimal portfolios). At the one end of the efficient set is the portfolio of minimum variance; at the other end is the portfolio of maximum expected return. Figure 3.25 displays an efficient frontier and feasible set constructed using the MVO method. At the left end of the efficient set is the portfolio of minimum variance; at the right end is the portfolio of maximum target return.

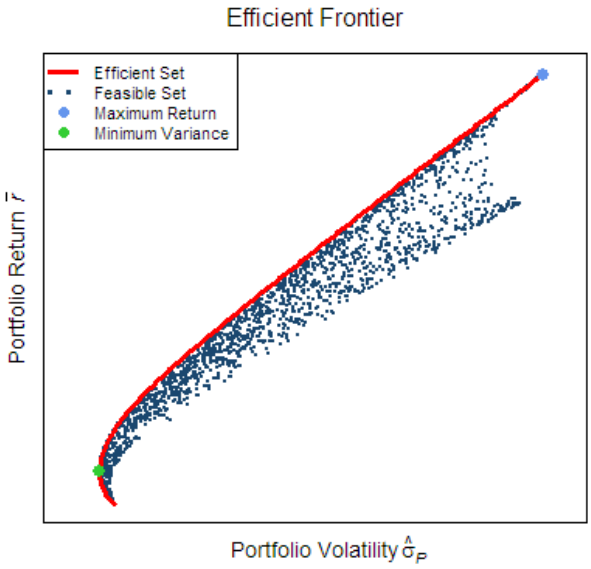


Figure 3.25: Efficient Frontier and Feasible Set

3.6.2.1 Mean-CVaR Optimisation

This section references Rockafellar and Uryasev (2000) and Krokhmal, Palmquist and Uryasev (2002).

The mean-CVaR optimisation problem is formulated using CVaR as the proxy for risk instead of the variance of returns.

Consider a portfolio of assets with random returns. Denote by $\boldsymbol{w}$ the portfolio decision vector of weights, with $\boldsymbol{W} \in \mathbb{R}$ the set of available portfolios, and by $\boldsymbol{r}$, a random events vector in $\mathbb{R}$. For each $\boldsymbol{w}$, the loss $f(\boldsymbol{w}, \boldsymbol{r})$ is a random variable with distribution in $\mathbb{R}$ induced by the distribution of $\boldsymbol{r}$.[100] The random vector $\boldsymbol{r}$ is assumed to have a probability density function, denoted by $p(\boldsymbol{r})$. The probability of $f(\boldsymbol{w}, \boldsymbol{r})$ not exceeding a threshold value α is then given by:

[100] Note that a negative value for the loss $f(\boldsymbol{w}, \boldsymbol{r})$ constitutes a gain.

$$\Psi(\boldsymbol{w}, \alpha) = \int_{f(\boldsymbol{w},\boldsymbol{r}) \leq \alpha} p(\boldsymbol{r})\, d\boldsymbol{r}\,. \tag{74}$$

As a function of α and fixed decision vector $\boldsymbol{w}$, $\Psi(\boldsymbol{w}, \alpha)$ is the CDF for the loss associated with $\boldsymbol{w}$. The function $\Psi(\boldsymbol{w}, \alpha)$ is non-decreasing with respect to α and is assumed to be everywhere continuous with respect to α.

Then, for a given confidence level $\beta \in (0,1)$,[101] the β-VaR for the loss random variable associated with portfolio $\boldsymbol{w}$ over the specified holding period is defined as:

$$\alpha_\beta(\boldsymbol{w}) = \min\{\alpha \in \mathbb{R} : \Psi(\boldsymbol{w}, \alpha) \geq \beta\}\,. \tag{75}$$

Similarly, the β-CVaR associated with portfolio $\boldsymbol{w}$ is defined as:

$$\phi_\beta(\boldsymbol{w}) = \frac{1}{1-\beta} \int_{f(\boldsymbol{w},\boldsymbol{r}) \geq \alpha_\beta(\boldsymbol{w})} f(\boldsymbol{w}, \boldsymbol{r})\, p(\boldsymbol{r})\, d\boldsymbol{r}\,. \tag{76}$$

Equation (75) defines β-VaR as the percentile α of the loss distribution, where the β-VaR of a portfolio is the lowest amount α such that, with probability β, the loss $f(\boldsymbol{w}, \boldsymbol{r})$ is less than or equal to α. Equation (76) defines β-CVaR as the conditional expectation of the loss associated with $\boldsymbol{w}$, given that the loss is greater than or equal to the VaR percentile $\alpha_\beta(\boldsymbol{w})$ (i.e., the probability of $f(\boldsymbol{w}, \boldsymbol{r}) \geq \alpha_\beta(\boldsymbol{w})$ is equal to $1 - \beta$). In this setting, it is difficult to optimise equation (76).

The key to optimisation problems with CVaR constraints is to re-characterise $\alpha_\beta(\boldsymbol{w})$ and $\phi_\beta(\boldsymbol{w})$ in terms of a function F_β on $\boldsymbol{W} \times \mathbb{R}$, defined as follows:

$$F_\beta(\boldsymbol{w}, \alpha) = \alpha + \frac{1}{1-\beta} \int_{\boldsymbol{r} \in \mathbb{R}} [f(\boldsymbol{w}, \boldsymbol{r}) - \alpha]^+\, p(\boldsymbol{r})\, d\boldsymbol{r}\,, \tag{77}$$

where $[t]^+ = \max(t, 0)$.

The function $F_\beta(\boldsymbol{w}, \alpha)$ has the following important properties that make it computationally tractable:

1. $F_\beta(\boldsymbol{w}, \alpha)$ is convex and continuously differentiable with respect to α, making the function easy to minimise numerically.
2. β-VaR is a minimum point of function $F_\beta(\boldsymbol{w}, \alpha)$ with respect to α, implying that β-CVaR can be calculated without having to first calculate β-VaR (on which the β-CVaR definition depends), thus simplifying the optimisation routine.

[101] Typical values are $\beta \in \{0.9, 0.95, 0.99\}$. This dissertation uses β set to 0.95 (i.e., 95%).

3. The minimum value of the function $F_\beta(\boldsymbol{w}, \alpha)$ is the β-CVaR value $\phi_\beta(\boldsymbol{w})$, in the sense that $\min_{\boldsymbol{w} \in \boldsymbol{W}} \phi_\beta(\boldsymbol{w}) = \min_{(\boldsymbol{w}, \alpha) \in \boldsymbol{W} \times \mathbb{R}} F_\beta(\boldsymbol{w}, \alpha)$.

As a consequence, the vector $\boldsymbol{w}$ that yields the minimum β-CVaR can be determined by optimising $F_\beta(\boldsymbol{w}, \alpha)$ with respect to the weights $\boldsymbol{w}$ and α.

In order to reduce complexity in the optimisation routine, the integral in $F_\beta(\boldsymbol{w}, \alpha)$ may be approximated by sampling the probability distribution of $\boldsymbol{r}$ according to its density $p(\boldsymbol{r})$. If the sampling generates a collection of vectors, say $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_S$, which may represent some historical returns or Monte Carlo returns, then a corresponding approximation to equation (77) may be formulated as:

$$\tilde{F}_\beta(\boldsymbol{w}, \alpha) = \alpha + \frac{1}{S(1-\beta)} \sum_{s=1}^{S} [f(\boldsymbol{w}, \boldsymbol{r}_s) - \alpha]^+ . \qquad (78)$$

The discretised function $\tilde{F}_\beta(\boldsymbol{w}, \alpha)$ is convex and piecewise linear with respect to α and, although no longer differentiable with respect to α in the approximation form, it can readily be minimised.

Typically, the loss function $f(\boldsymbol{w}, \boldsymbol{r})$ is convex with respect to $\boldsymbol{w}$, producing convexity in the approximation expression $\tilde{F}_\beta(\boldsymbol{w}, \alpha)$.[102] The problem of minimising $\tilde{F}_\beta(\boldsymbol{w}, \alpha)$ over $\boldsymbol{W} \in \mathbb{R}$, therefore, is one of convex programming. However, $\tilde{F}_\beta(\boldsymbol{w}, \alpha)$ may be linearised and the resulting linear expression (subject to a set of linear constraints) may be solved by linear programming algorithms. Replace $[f(\boldsymbol{w}, \boldsymbol{r}_s) - \alpha]^+$ in $\tilde{F}_\beta(\boldsymbol{w}, \alpha)$ by artificial real variables z_s for $s = 1, \ldots, S$ to obtain the linear function:

$$\alpha + \frac{1}{S(1-\beta)} \sum_{s=1}^{S} z_s , \qquad (79)$$

subject to constraints:

$$z_s \geq f(\boldsymbol{w}, \boldsymbol{r}_s) - \alpha$$

$$z_s \geq 0 .$$

The mean-CVaR portfolio selection problem may be solved by minimising β-CVaR in its linearised form with respect to portfolio $\boldsymbol{w}$, and is formulated as follows:

$$\boldsymbol{w} = \arg\min_{\boldsymbol{w}} \mathrm{CVaR}_\beta(\boldsymbol{w}) , \qquad (80)$$

subject to constraints:

$$\boldsymbol{w}' \hat{\boldsymbol{\mu}} = \bar{r}$$

$$\boldsymbol{w}' \mathbf{1} = 1 .$$

[102] As well as convexity in $F_\beta(\boldsymbol{w}, \alpha)$ with respect to $(\boldsymbol{w}, \alpha)$ and convexity in $\phi_\beta(\boldsymbol{w})$ with respect to $\boldsymbol{w}$.

The loss $f(\boldsymbol{w},\boldsymbol{r})$ does not depend on $\boldsymbol{r}$ having a Gaussian distribution, as is implied by the use of estimate $\hat{\boldsymbol{\Sigma}}$ in the formulation (70). The formulation in (80), by strictly focussing on mean tail losses, fully admits non-normal distributions into the framework. Worth noting, however, is that for elliptically distributed asset returns, optimising the mean-CVaR problem will yield the same set of asset weights as for the mean-variance Markowitz portfolio.

3.6.3 Optimisation Results

This section uses the `fPortfolio` package in `R` (Wuertz, Setz & Chalabi, 2014). The aim is to compare the mean-variance and mean-CVaR methods. The simulated, conditioned set of returns from Chapter 3.6.1 are optimised to obtain an efficient frontier using the mean-CVaR procedure explained above. In order to compare the historical and scenario-based methods, an efficient frontier is constructed using Markowitz's MVO on the historical data. Two other constraints, in addition to those of (65) and (67) are input to the optimisation routines. The first is that, simply, all portfolios are "long only", where assets comprising portfolios may only be purchased (i.e., no borrowing of assets to sell short is considered). This constraint characterises the majority of global investment mandates. The second is that offshore investments are constrained to a maximum of 25% of the portfolio. This reflects the fact that South African investors are restricted to allocating no more than 25% of their investments beyond the borders of South Africa.

Figure 3.26 displays the efficient frontier plots of optimal mean-variance portfolios optimised using 20 years of historical returns (red) and optimal mean-CVaR portfolios optimised on the 138 years of simulated returns conditioned on the 2014-2015 input scenario (blue). Both routines assume a risk-free rate of 5.75% (i.e., the 3-month Treasury bill rate at the beginning of the forecast horizon) from which the "tangency portfolio" of each frontier is calculated. The tangency portfolio yields the optimal risky portfolio among the efficient set in terms of maximising the Sharpe ratio for a given risk-free rate. It is the point on the frontier that maximises the quantity:

$$\arg\max_{\boldsymbol{w}} g(\boldsymbol{w}) = \frac{\boldsymbol{w}'\hat{\boldsymbol{\mu}} - r_f}{\sqrt{\boldsymbol{w}'\hat{\boldsymbol{\Sigma}}\boldsymbol{w}}}, \tag{81}$$

subject to constraints $\boldsymbol{w}'\hat{\boldsymbol{\mu}} = \bar{r}$ and $\boldsymbol{w}'\mathbf{1} = 1$. Note: for ease of comparison, instead of plotting CVaR as the x-axis risk measure, the portfolio volatility levels associated with the mean-CVaR frontier returns were extracted and the locus points plotted in the $(\hat{\sigma}_P, \bar{r})$-space.

The tangency portfolios in Figure 3.26 are used as a reference to describe some evident differences between the two frontiers. For example, investing in the tangency portfolio constructed using historical MVO, an investor may target an annual return of 10.8% by choosing to accept a corresponding expected volatility of returns of 2.9%. Constructing the tangency portfolio using the scenario-based simulation and mean-CVaR optimisation, an investor may expect an annual return

of 14.5% for assuming a commensurate 2.9% level of volatility risk. Alternatively, an investor in the tangency portfolio on the simulation-based frontier may expect an annual return of 13.2% with a volatility level of 2.1%, whereas the composition of the MVO-constructed portfolio targeting the same 13.2% return would reflect a much greater level of expected volatility, at 9.4%.

Figures 3.27 and 3.28 display the compositional differences between portfolios along the historical MVO and the simulated, conditional mean-CVaR frontiers. Optimal allocation weights, showing explicitly how to construct a portfolio on the frontier, are given below each figure for increasing levels of target risk (top line of x-axis) or target return (bottom line of x-axis). The effect of the 25% offshore constraint is apparent in Figure 3.28: beyond the target return level of 14.67%, portfolio weight allocated to the MSCI World Equity Index is capped at 25%. This cap is accompanied by a rapid increase in allocation to the domestic equity All Share Index to compensate for higher levels of targeted returns.

Below each figure is a table reflecting target returns, volatility levels and CVaR levels corresponding to 10 equidistant optimal portfolios. For each portfolio, the associated weight allocations used to construct the portfolio are given. The vertical red (resp. gold) lines represent the location of the minimum variance (resp. CVaR) portfolios. Note that all output (returns, volatility levels, CVaR levels and weights) is in percentage format.

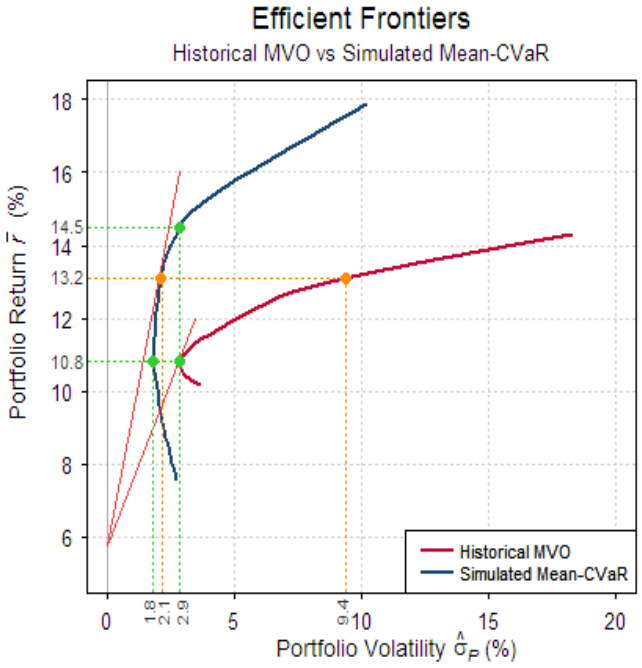
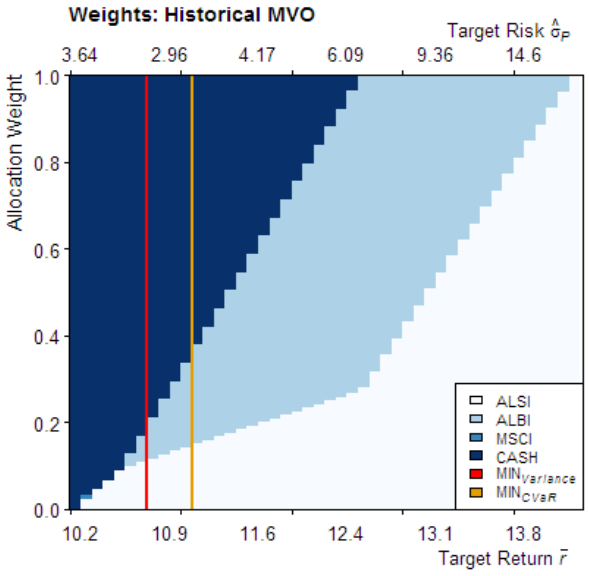


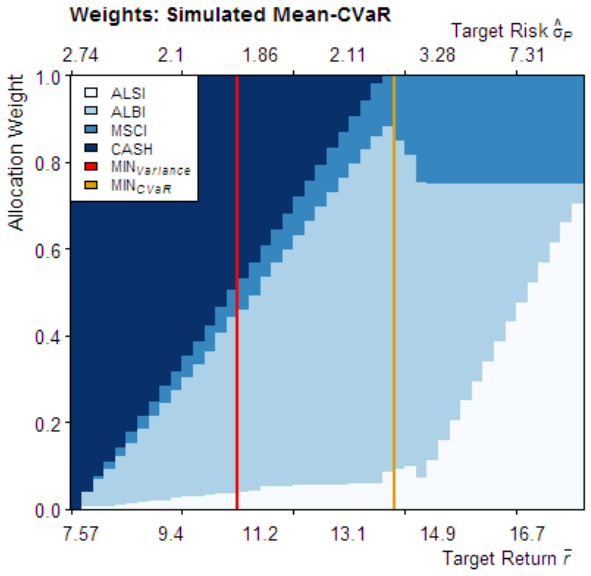
Figure 3.26: Efficient Frontiers of Historical MVO and Simulation-Based Mean-CVaR Optimisation

Figure 3.27: Optimal MVO Portfolios Optimised on Historical Returns



Portfolio	1	2	3	4	5	6	7	8	9	10
$\bar{r}$	10.2	10.8	11.1	11.5	12	12.5	12.9	13.4	13.8	14.3
$\hat{\sigma}_P$	3.6	2.9	3.1	4	5.1	6.3	8.2	11.2	14.6	18.3
CVaR	-5.2	-5.6	-6.4	-5.6	-3.9	-1.7	3.5	12.2	20.8	29.5
	Weights									
ALSI	0	10	14.1	18.3	22.4	26.6	43	62	81	100
ALBI	0	2.8	19.5	36.3	53	69.8	57	38	19	0
MSCI	0	0	0	0	0	0	0	0	0	0
CASH	100	87.3	66.3	45.4	24.5	3.6	0	0	0	0

Figure 3.28: Optimal Mean-CVaR Portfolios Optimised on Forward-Looking Simulated Returns



Portfolio	1	2	3	4	5	6	7	8	9	10
$\bar{r}$	7.6	8.7	9.9	11	12.2	13.3	14.4	15.6	16.7	17.9
$\hat{\sigma}_P$	2.7	2.3	2	1.9	1.9	2.2	2.8	4.6	7.3	10.2
CVaR	-5.6	-6.8	-7.6	-8.4	-8.9	-9.4	-9.7	-7.9	-4.7	-0.7
	Weights									
ALSI	0	1.2	2.8	4	5.4	5.7	10	24.9	47.6	70.4
ALBI	0	14	27.4	41.9	57.8	73.7	71.7	50.1	27.4	4.6
MSCI	0	2.5	4.9	7.2	8.7	11.1	18.3	25	25	25
CASH	100	82.3	64.8	46.9	28.2	9.5	0	0	0	0

In order to evaluate mean-variance and mean-CVaR optimisations comparatively, both methods are applied to the same set of 138 conditioned 2014-2015 input scenario observations. The resulting efficient frontiers are plotted in Figure 3.29 in the $(\hat{\sigma}_P, \bar{r})$-space and Figure 3.30 in the $(\mathrm{CVaR}_{\beta=0.95}, \bar{r})$-space. Note that negative CVaR values on the x-axis express negative losses (i.e., positive returns). For example, in the latter figure, $\mathrm{CVaR}_{0.95}$ = -9.8% translates to the average annual return of the ±7 worst scenarios (5% x 138 years) being equal to +9.8%. Figure 3.29 shows that, in the target return region of 8%-14%, the optimal MVO portfolios have slightly lower return volatilities $\hat{\sigma}_P$ than do the optimal mean-CVaR portfolios. In Figure 3.30, over the same target return region of 8%-14%, the optimal mean-CVaR portfolios reflect slightly lower downside risks (CVaR) than do the optimal MVO portfolios. The potential risk advantages (either downside risk or symmetric risk) conferred by either technique dissipate beyond a target return of +14%.

Figure 3.29: Optimal Mean-Variance and Mean-CVaR Portfolios Plotted in the $(\hat{\sigma}_P, \bar{r})$-Space

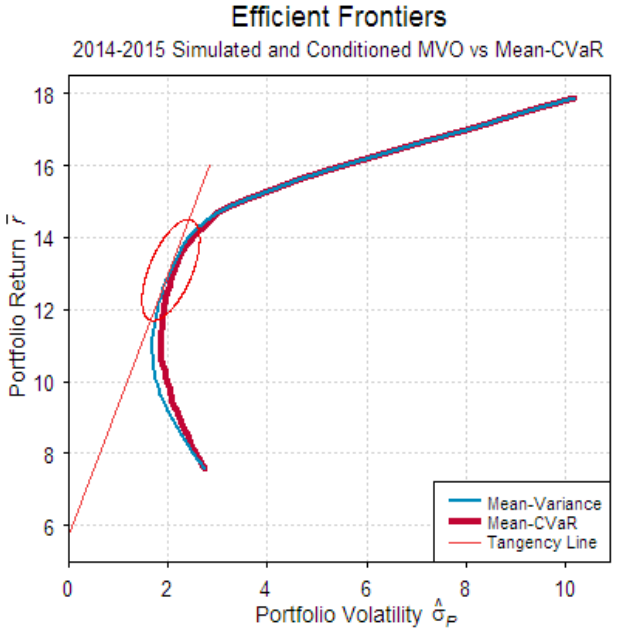
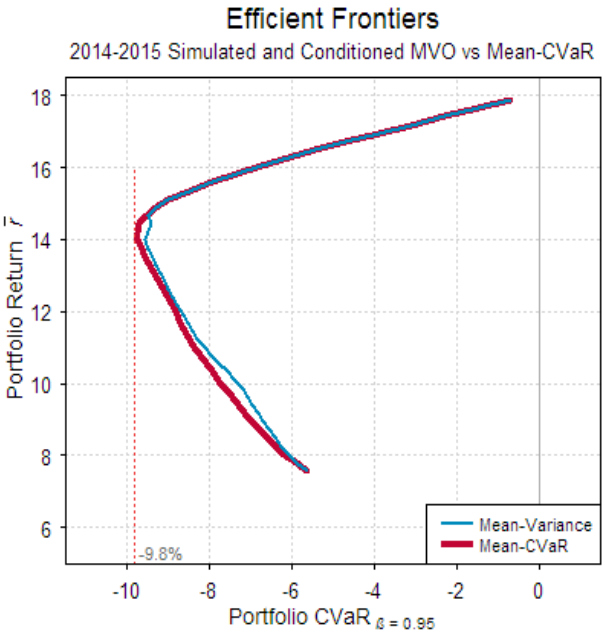
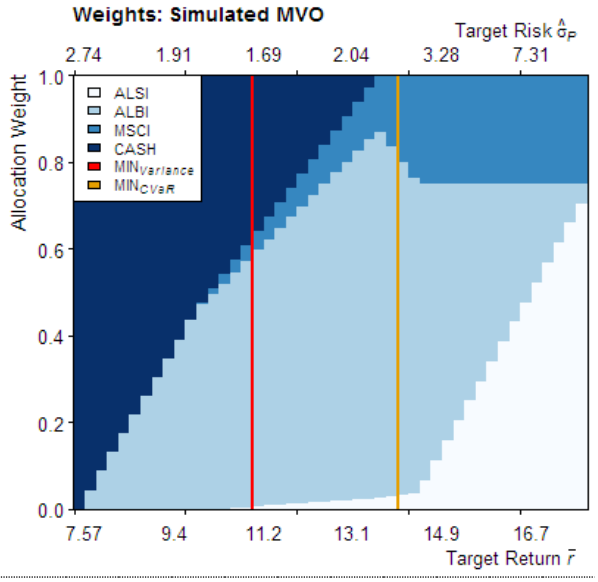


Figure 3.30: Optimal Mean-Variance and Mean-CVaR Portfolios Plotted in the $(\mathrm{CVaR}_{\beta=0.95}, \bar{r})$-Space



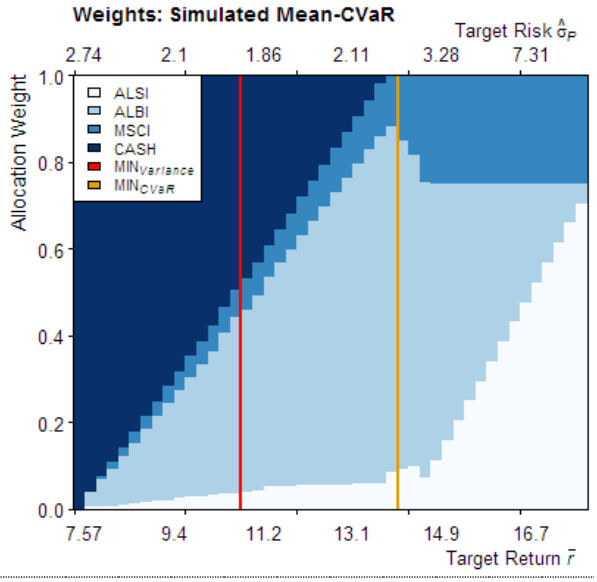
Figures 3.31 and 3.32 display, respectively, the corresponding optimal MVO and mean-CVaR portfolio asset weights, with statistics tables listed below each figure. Again, all output is in percentage format. Optimal mean-CVaR portfolios 2 – 7 in Figure 3.32 have slightly higher return volatilities $\hat{\sigma}_P$ than do the corresponding optimal MVO portfolios in Figure 3.31. In exchange, however, the optimal mean-CVaR portfolios have lower downside risks (CVaR). It appears that what the mean-CVaR portfolios sacrifice in higher volatility, they gain in lower risk of tail losses. In the words of Wang and Zheng (2010, p. 20): "the cost of protecting against downside risk [through mean-CVaR portfolio optimisation] is only slightly higher variance in the portfolio's returns". The red highlighted column in Figure 3.31 reflects the asset allocation weights corresponding to the minimum variance portfolio. The minimum CVaR portfolio weights are highlighted in the gold column of Figure 3.32. The composition and risk-return profile of each risk-minimising portfolio is very different, illustrating the importance in choice of risk measure by risk-averse investors. The optimal mean-CVaR portfolio allocates greater weights to domestic asset classes ALSI and ALBI, as well as to offshore equities, than the optimal mean-variance portfolio. On the other hand, there is no cash reserve in the mean-CVaR portfolio, versus a 39.4% cash allocation in the mean-variance portfolio. A preferred minimum cash allocation constraint, however, may be fed into the optimisation routine and optimal portfolios re-evaluated. Such a constraint will leave a consistent amount of cash in the portfolio.

Figure 3.31: Weights of Optimal MVO Portfolios Optimised on Forward-Looking Simulated Returns



Portfolio	1	2	3	4	5	6	7	8	9	10
$\bar{r}$	7.6	8.7	9.9	11	12.2	13.3	14.4	15.6	16.7	17.9
$\hat{\sigma}_P$	2.7	2.2	1.8	1.7	1.8	2.1	2.7	4.6	7.3	10.2
CVaR	-5.6	-6.6	-7.2	-8.1	-8.8	-9.3	-9.4	-7.9	-4.7	-0.7
Weights										
ALSI	0	0	0	0.4	1.3	2.1	3.6	24.9	47.6	70.4
ALBI	0	21.7	43.4	56.6	68.5	80.4	72.7	50.1	27.4	4.6
MSCI	0	0	0	3.6	7.4	11.2	23.7	25	25	25
CASH	100	78.3	56.6	39.4	22.8	6.2	0	0	0	0

Figure 3.32: Weights of Optimal Mean-CVaR Portfolios Optimised on Forward-Looking Simulated Returns



Portfolio	1	2	3	4	5	6	7	8	9	10
$\bar{r}$	7.6	8.7	9.9	11	12.2	13.3	14.4	15.6	16.7	17.9
$\hat{\sigma}_P$	2.7	2.3	2	1.9	1.9	2.2	2.8	4.6	7.3	10.2
CVaR	-5.6	-6.8	-7.6	-8.4	-8.9	-9.4	-9.7	-7.9	-4.7	-0.7
Weights										
ALSI	0	1.2	2.8	4	5.4	5.7	10	24.9	47.6	70.4
ALBI	0	14	27.4	41.9	57.8	73.7	71.7	50.1	27.4	4.6
MSCI	0	2.5	4.9	7.2	8.7	11.1	18.3	25	25	25
CASH	100	82.3	64.8	46.9	28.2	9.5	0	0	0	0

Figures 3.33 and 3.34 display the sources of individual contributors to (target) aggregate portfolio returns. Figures 3.35 and 3.36 display the sources of individual contributors to (target) aggregate portfolio volatility (covariance risk). In both mean-variance and mean-CVaR optimisation routines, the quantities of individual contributors to portfolio returns and portfolio risks may be constrained. For example, the contribution of volatility or CVaR risk to the portfolio from any one asset may be capped at 5%, a new efficient set of portfolios calibrated and the associated allocation weights obtained to construct the portfolio, or serve as a guide in portfolio adjustment decisions.

To improve insight into the characteristics of individual contributors to portfolio risks and returns, a layer of statistical traits at the conditioned asset level may be evaluated. The initial simulation generated 100,000 likely annual observations per variable (with the idiosyncratic and cross-asset characteristics maintained). The output that is conditioned on an anticipated input scenario requires all input ranges be simultaneously matched before retrieving the corresponding output (portfolio variable) values. After a complete matching set is extracted, each variable comprising the set will have a new and unique statistical profile, which can be evaluated.

Figure 3.33: Individual Contributors to Returns in the Optimal MVO Portfolios

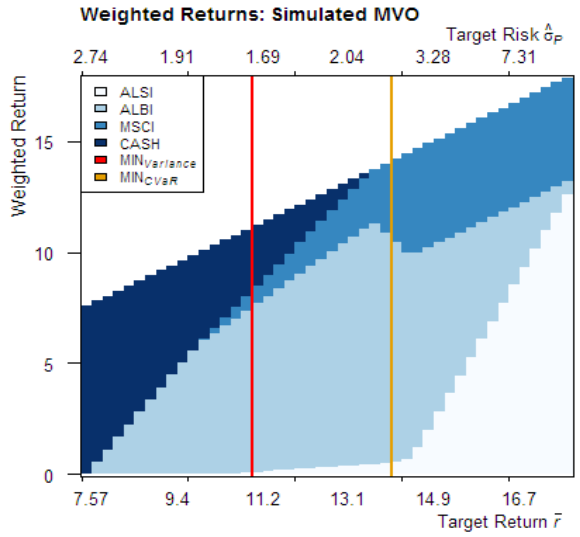


Figure 3.34: Individual Contributors to Returns in the Optimal Mean-CVaR Portfolios

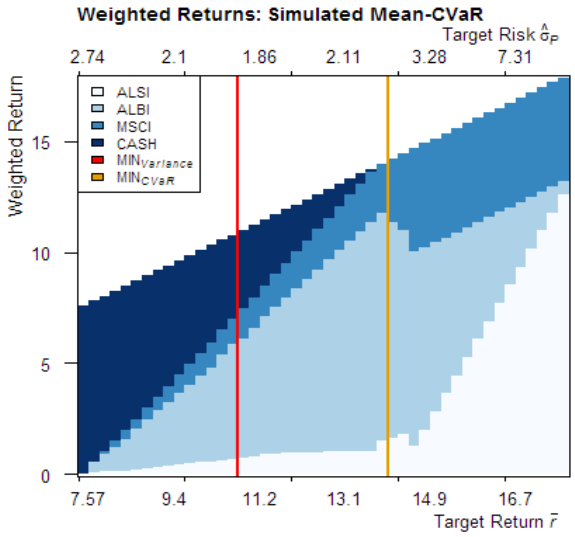


Figure 3.35: Individual Contributors to Covariance Risk in the Optimal MVO Portfolios

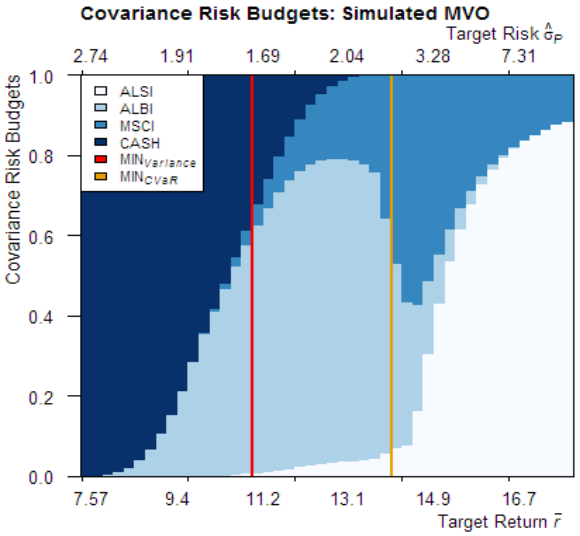
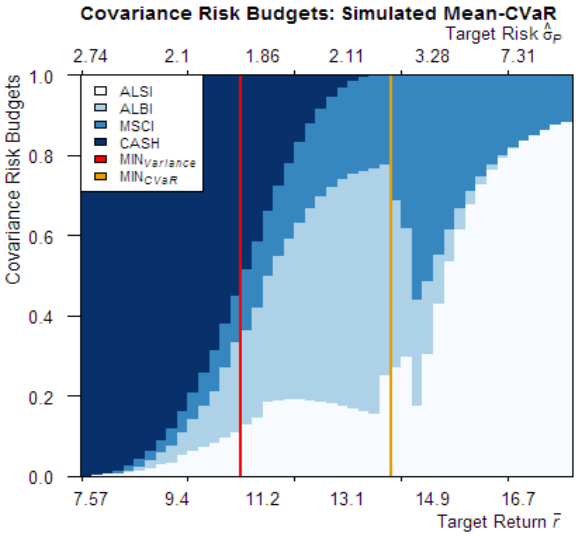


Figure 3.36: Individual Contributors to Covariance Risk in the Optimal Mean-CVaR Portfolios



The graphs in Figure 3.37 illustrate the basic statistical profiles of each of the 2014-to-2015 conditioned variables. Noticeable in each of the graphs is that the expected (and median) returns are all greater than the actual returns over the out-of-sample period. This may possibly be an artefact of the symmetric structure of the multivariate *t* distribution being unable to induce sufficient tail asymmetry into the simulations. A box-and-whisker plot (boxplot) is superimposed on each of the graphs, showing the median, interquartile range, maximum and minimum and outlier zones of the expected returns[103]. The histograms provide a visual display of the empirical or, in this case, expected range of returns and likelihoods of occurrence along the range of returns.

[103] The boxplot (see Appendix) provides a concise visual summary of essential data characteristics (viz., the median, interquartile range, skewness and presence or absence of outliers). The line splitting the rectangular box corresponds to the median return value of the range of returns.

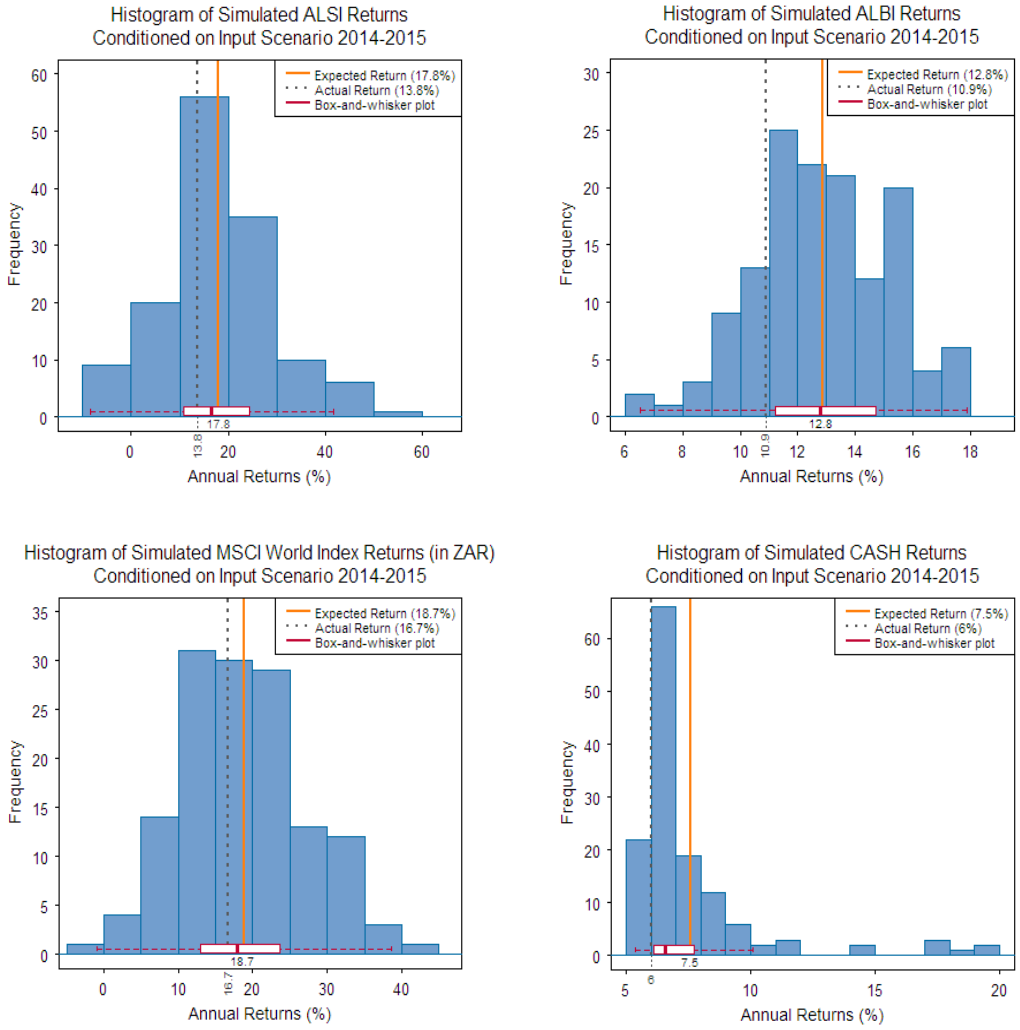


Figure 3.37: Comparison of Expected Returns versus Actual Returns in the Out-of-Sample Scenario

In the mean-variance and mean-CVaR methods discussed, a risk-averse investor is assumed to use either variance or CVaR, respectively, as the preferred risk measure. To conclude this chapter, actual results from the optimal variance-preference portfolio (#4 in Figure 3.31) and optimal CVaR-preference portfolio (#7 in Figure 3.32) are given in Table 3.14. Both methods overestimated the portfolio's expected annual returns. At the yearly scale, ex ante expected returns, volatility and absolute CVaR values are all greater in portfolio 7 than portfolio 4. These characteristics are preserved out-of-sample at the monthly scale.

Table 3.14: Comparison of Actual and Expected Portfolio Statistics

	Method	MVO	Mean-CVaR
	Portfolio #	4	7
Yearly	**Actual r**	**9.2%**	**12.3%**
	Expected $\bar{r}$	**11.0%**	**14.4%**
	Expected $\hat{\sigma}_P$	**1.7%**	**2.8%**
	Expected CVaR	**-8.1%**	**-9.7%**
Monthly	**Actual r**	**0.8%**	**1.0%**
	Actual σ_P	**1.4%**	**1.7%**
	Actual CVaR	**-1.5%**	**-1.7%**
	ALSI	0.4%	10%
	ALBI	56.6%	71.7%
	MSCI	3.6%	18.3%
	CASH	39.4%	0%

Chapter 4

Conclusion

This chapter summarises the entire dissertation and offers suggestions for future research.

4.1 Concluding Remarks

We acknowledge that global financial markets are trending towards unification, strengthening volatility linkages and quickening volatility transmissions through present-day financial networks. As it relates to South African fund managers, domestic portfolios are net recipients of returns and volatility shocks from major world markets. Financial crises continue to highlight the damaging effects of deep-tail, extreme events on portfolios. Therefore, this dissertation proposes a methodology to improve risk management systems in funds by building an asset allocation framework that offers practitioners an opportunity to explicitly model combinations of hypothesised global risks and the effects on their investments. The goal is to improve fund performances, particularly during periods of market stress.

Opportunity for improvement comes from pooling expert opinions on key near-term portfolio risk drivers. Aggregated opinions on key variables give rise to range forecasts used in a framework where any number of important variables can be combined to construct forward-looking scenarios, or anticipated states of the world. The scenario-building option offers practitioners and investors a pivotal mechanism through which many different effects on fund holdings can be scrutinised. The mechanism fits top-down, macro views onto a flexible, bottom-up, quantitative model. This dissertation brought together new technologies available to jointly model portfolio assets and their risk factor threats and elaborated on the requisite theoretical background. At the centre of the model is a dualistic simulation engine constructed to represent the multivariate and univariate data generating processes as accurately as possible.

Estimation of expected returns was not the focus of the research. It is widely accepted that accurate forecasts of returns are, at best, tentative (Campbell & Thompson, 2008; Goyal & Welch, 2008). Therefore, an ARMA structure was suggested for returns and a GARCH-family structure for the ARMA model errors. GARCH models have emerged to address the stylised facts characterising univariate financial data, such as leptokurtosis, asymmetry, autocorrelation and heteroskedasticity. To enhance capability in better representing the stylised facts of the data, GARCH models are paired with a range of conditional distributions designed with optionalities to absorb excess heavy-tailedness and skewness. Even so (and particularly in weekly or higher frequency data), there is typically a set of extreme observations in the tails of each variable that remains unrepresented by the model. Extreme Value Theory has been suggested to parametrically model the tail densities of

such extrema. This paper incorporated the theory to separately model upper and lower tails with generalised Pareto distributions, which, additionally, admit extrapolation beyond the limits set in the historical data. The high-density interiors were modelled with non-parametric Gaussian kernel smoothers. The tails and interior are spliced together to form a semi-parametric distribution, which not only admits asymmetry through separate modelling of the tails, but also represents (i) a cumulative distribution function used to create pseudo-copula data and (ii) a quantile function to transform copula-simulated data back to the domain of standardised residuals.

Modern Portfolio Theory relies on a Gaussian dependence structure based on Pearson's linear correlation. These structures provide a reliable framework for building diversified portfolios in normal markets, but they break down during stress periods. Volatile markets generate non-linear co-variation in multivariate financial data and copula theory is gaining momentum as the theory of choice for capturing these co-variation patterns, particularly those of observed tail dependencies. Although copulas are not a panacea for asset allocation problems, they admit a methodology that has been shown to materially improve the portfolio modelling framework when compared to the MPT frameworks currently employed by the majority of fund managers. They are most useful in modelling assets with less common or uncommon characteristics (such as those found in macro asset allocation problems), but their usefulness diminishes as variety among portfolio assets diminishes (such as in sector-concentrated funds comprising highly correlated assets).

The capstone of this study is portfolio optimisation. The effectiveness of portfolio optimisation is jointly influenced by the assumptions on asset return distributions, dependence structure and choice of risk measure. To obtain portfolios to optimise, we built a dualistic simulation engine. Univariate marginal semi-parametric distributions were modelled with a combination of heavy-tailed ARMA-GARCH structures and the EVT peaks-over-threshold approach. On the multivariate side, a meta Student *t* distribution was used to model joint portfolio asset and risk factor returns. The meta *t* distribution was constructed using the separation method admitted by Sklar's theorem. We used a static Student *t* copula to represent the multivariate (non-linear) dependence structure between variables and semi-parametric distributions to represent marginal variables. Multivariate simulations were generated from the meta *t* distribution thus characterised and the simulated output used as input to generate simulations from the ARMA-GARCH structures. The simulations resulted in financial data that was then filtered through a forward-looking scenario and what remained was portfolio data to be optimised.

Two asset allocation strategies were investigated: the classical Markowitz mean-variance strategy and a contemporary mean-CVaR strategy. The former uses a symmetric risk measure (standard deviation) to optimise portfolios; the latter, an asymmetric, downside risk measure (in this case, CVaR). Portfolios were constrained to "long-only" and a maximum of 25% invested offshore. The

forward-looking scenario was set to the year-ahead, out-of-sample period from 30 April 2014 to 30 April 2015. This scenario assumes perfect foresight in that the range for each risk factor variable was set to encapsulate the actual returns over the forecast horizon. Mean-CVaR optimisation was performed on the resulting conditioned asset class returns and compared to mean-variance optimisation performed on the entire historical data set. The optimisation on conditioned asset class returns showed a clear advantage to using the methodology proposed in this dissertation, however, subject to the caveat of being able to encapsulate ex ante risk factor returns in forward-looking range forecasts. The area of forecasting is where pooled expert opinion would be expected to contribute meaningfully to the model. Mean-variance and mean-CVaR optimisations were then performed on the conditioned returns. The risk-minimising portfolio from each strategy was selected as being optimal and the portfolios were compared out-of-sample. The optimal mean-CVaR portfolio returned more money than the optimal mean-variance portfolio, while enjoying lower monthly downside risk, but exhibited higher monthly volatility risk. Finally, a repeat of the 2008 financial crisis was input as a forward-looking scenario and expected versus actual returns compared. Scenarios were generated using actual returns on key variables from 31 December 2007 to 31 December 2008. As we considered this an in-sample evaluation, results were placed in the Appendix.

We concluded that the proposed asset allocation framework incorporating stylised facts of univariate and multivariate financial data, as well as contemporary portfolio optimisation techniques, leads to material risk-return gains for the risk-averse investor. In addition, the simulation component of the framework offers practitioners meaningful opportunity for improving resilience in portfolios against volatility threats.

4.2 Further Research

In this final section of the dissertation we list a number of recommendations for future research.

The model was fit and tested on low frequency data (monthly returns), leaving open an immediate extension to higher frequency data, such as daily data. Ng and Lam (2006) recommended sample sizes starting at 1000 observations for estimating the univariate conventional GARCH model, in order to avoid possible wrong optimal solutions. Since GARCH-processes are data hungry, satisfying the data requirements at the univariate level would likely feed through to improved parameter estimates at the multivariate level. Therefore, investigating multivariate parameter stability across larger and varying sample sizes (e.g., sizes $n =$ 2000, 3000, 5000, etc.) combined with higher and varying model dimensions (e.g., dimensions $d =$ 15, 20, 30, etc.) would be a logical next step for future research. As well, the research would benefit from evaluating the fitted copula and overall multivariate model with statistical robustness checks.

More interesting, however, would be to extend the study to admit both asymmetric and time-varying dependence in the copula structure. Correlations strengthen as asset classes synchronously absorb shocks to the financial system. Otherwise lower correlations that may ordinarily exist among assets in normal market environments trend quickly to high correlation levels during market downturns. This causes a dynamic and asymmetric dependence phenomenon, with adverse implications to portfolio diversification and hedging effectiveness. Furthermore, being quick to be able to recognise meaningful changes (as a function of time) in a dependence structure (whether symmetric or asymmetric) would be a valuable skill in portfolio risk management. To this end, extending the skewed *t* copula (Christoffersen, et al., 2012; Demarta & McNeil, 2005) and generalised hyperbolic skewed *t* copula (Allen & Satchell, 2014; Cerrato, Crosby, Kim & Zhao, 2015) to admit time-varying, or conditional, dependence are logical directions for future research. In terms of parsimony and manageable computational complexity, however, an alternative to the aforementioned would be to investigate the effects of varying the degrees of freedom parameter in the static meta *t* copula in the multivariate simulation stage. An exogenous metric (or metrics) may be determined and used to decide when and to what degree the tail dependence parameter should be varied (e.g., lowered to induce more tail dependence). Research in this direction would be similar to that of a regime-switching framework. More directly practical, however, and in terms of incorporating dynamic, conditional correlation into the (symmetric) static meta *t* copula framework, the DCC model of Engle (2002) and the Asymmetric Generalised DCC (AGDCC) model of Cappiello, Engle and Sheppard (2006) offer practitioners a promising opportunity to replace the static dispersion matrix P in equations (50) and (52) with a conditional, time-varying correlation matrix P_t (derived from the decomposition of the DCC or AGDCC model's conditional covariance matrix $H_t = D_t P_t D_t$). Recent and relevant time variation in asset and risk factor dynamics may, hereby, directly be incorporated into the multivariate framework proposed in this dissertation. Ghalanos (2012) and Le (2012) found significant advantages to constructing a model in this way. Although it may necessitate a trade-off in static, non-linear dependence for dynamic, conditional correlation, it would be interesting to investigate this step further.

In terms of programming, a natural extension for this and future, more complex, modelling frameworks would be to address the computational burden in model fitting. The resource burden is not only apparent in fitting and simulating in the dependence section, but is also apparent in the mean-CVaR portfolio optimisation section. The time it takes to calibrate high-dimensional models could be markedly reduced by rewriting the R code in C/C++, as well as by implementing parallel and grid computing techniques. Additionally, in the portfolio optimisation section, computational efficiencies can be gained by deploying smoothing algorithms (Alexander, Coleman & Li, 2006; Zhu, Coleman & Li, 2009) instead of the linear programming algorithm for the CVaR optimisation problem.

References

Aas, K. and Haff, I.H. (2006). The Generalized Hyperbolic Skew Student's *t*-distribution. *Journal of Financial Econometrics,* 4 (2):275-309.

Absa Capital (2013). South Africa Asset Allocation: The End of Modern Portfolio Theory. *South African Equity Strategy.* Johannesburg, South Africa: Absa Capital.

Ageeva, E. (2011). Bayesian Inference for Multivariate *t* Copulas Modeling Financial Market Risk. (*Unpublished manuscript)*, ETH Zürich, Switzerland.

Akgiray, V. and Booth, G.G. (1988). Mixed Diffusion-Jump Process Modeling of Exchange Rate Movements. *Review of Economics and Statistics,* 70 (4):631-637.

Alexander, C. (2008). *Market Risk Analysis IV: Value-at-Risk Models.* Chichester, UK: John Wiley & Sons.

Alexander, G.J. and Baptista, A.M. (2004). A Comparison of VaR and CVaR Constraints on Portfolio Selection with the Mean-Variance Model. *Management Science,* 50 (9):1261-1273.

Alexander, S., Coleman, T.F. and Li, Y. (2006). Minimizing CVaR and VaR for a Portfolio of Derivatives. *Journal of Banking and Finance,* 30 (2):583-605.

Allen, D. and Satchell, S. (2014). The Four Horsemen: Heavy-tails, Negative Skew, Volatility Clustering, Asymmetric Dependence. *Discussion Paper: 2014-004,* University of Sydney: Sydney, Australia.

Alles, L.A. and Kling, J.L. (1994). Regularities in the Variation of Skewness in Asset Returns. *Journal of Financial Research,* 17 (3):427-438.

Amenc, N., Goltz, F. and Lioui, A. (2011). Practitioner Portfolio Construction and Performance Measurement: Evidence from Europe. *Financial Analysts Journal,* 67 (3):39-50.

Amendola, A. and Francq, C. (2009). Concepts of and Tools for Nonlinear Time-Series Modelling. In D.A. Belsley and E.J. Kontoghiorghes (Eds.), *Handbook of Computational Econometrics* (pp. 377-427). Chichester, UK: John Wiley & Sons.

Ang, A., Chen, J. and Xing, Y. (2006). Downside Risk. *Review of Financial Studies,* 19 (4):1191-1239.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance,* 9 (3):203-228.

Bachelier, L. (1900). Théorie de la Spéculation. *Annales Scientifiques de l'École Normale Supérieure,* 17 (3):21-86.

Bai, X., Russell, J.R. and Tiao, G.C. (2003). Kurtosis of GARCH and Stochastic Volatility Models with Non-normal Innovations. *Journal of Econometrics,* 114 (2):349-360.

Baillie, R.T. and Bollerslev, T. (1989). The Message in Daily Exchange Rates: A Conditional-Variance Tale. *Journal of Business and Economic Statistics,* 7 (3):297-305.

Balkema, A.A. and de Haan, L. (1974) Residual Life Time at Great Age. *Annals of Probability,* 2 (5):792-804.

Ball, C.A. and Torous, W.N. (1983). A Simplified Jump Process for Common Stock Returns. *Journal of Financial and Quantitative Analysis,* 18 (1):53-65.

Bao, Y., Lee, T.-H. and Saltoğlu, B. (2004). A Test for Density Forecast Comparison with Applications to Risk Management. *Working Paper,* University of California: Riverside, CA.

Bao, Y., Lee, T.-H. and Saltoğlu, B. (2007). Comparing Density Forecast Models. *Journal of Forecasting,* 26 (3):203-225.

Bao, Y. (2015). Should we Demean the Data? *Annals of Economics and Finance,* 16 (1):163-175.

Baruník, J., Kočenda, E. and Vácha, L. (2014). Asymmetric Connectedness of Stocks: How does Bad and Good Volatility Spill Over the U.S. Stock Market? *Financial Distortions and Macroeconomic Performance, Working Paper No. 13*.

Basel Committee on Banking Supervision (1996). *Amendment to the Capital Accord to Incorporate Market Risks* [Online]. Retrieved 25 February 2015 from http://www.bis.org/publ/bcbs24.htm

Bibby, B.M. and Sørensen, M. (2003). Hyperbolic Processes in Finance. In S.T. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance* (pp. 329-384). Amsterdam, The Netherlands: Elsevier.

Blattberg, R.C. and Gonedes, N.J. (1974). A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *Journal of Business,* 47 (2):244-280.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics,* 31 (3):307-327.

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics,* 69 (3):542-547.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics,* 52 (1):5-59.

Boston Consulting Group (2014). *Global Asset Management 2014: Steering the Course to Growth* [Online]. Retrieved 10 June 2015 from https://www.bcgperspectives.com/content/articles/financial_institutions_global_asset_management_2014_steering_course_growth

Boubaker, H. and Sghaier, N. (2013). Portfolio Optimization in the Presence of Dependent Financial Returns with Long Memory: A Copula Based Approach. *Journal of Banking and Finance,* 37 (2):361-377.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G. and Roncalli, T. (2000). Copulas for Finance – A Reading Guide and Some Applications. *Groupe de Recherche Opérationnelle, Crédit Lyonnais, Working Paper*.

Box, G.E.P. and Tiao, G.C. (1962). A Further Look at Robustness Via Bayes's Theorem. *Biometrika,* 49 (3/4):419-432.

Box, G.E.P. and Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26 (2):211-252.

Brandimarte, P. (2014). *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics.* Hoboken, NJ: John Wiley & Sons.

Brochart, B., Taillardat, B. and Jourovski, A. (2012). Liquidity and Crowding within Low Volatility Equity Strategies. *Unigestion, Research Paper.*

Brock, W.A., Dechert, W.D., Scheinkman, J.A. and LeBaron, B. (1996). A Test for Independence Based on the Correlation Dimension. *Econometric Reviews,* 15 (3):197-235.

Brownlees, C., Engle, R.F. and Kelly, B. (2011). A Practical Guide to Volatility Forecasting through Calm and Storm. *Journal of Risk,* 14 (2):3-22.

Brunnermeier, M.K., Nagel, S. and Pedersen, L.H. (2009). Carry Trades and Currency Crashes. *National Bureau of Economic Research Macroeconomics Annual 2008,* 23:313-347.

Byström, H.N.E. (2004). Managing Extreme Risks in Tranquil and Volatile Markets using Conditional Extreme Value Theory. *International Review of Financial Analysis,* 13 (2):133-152.

Byström, H.N.E. (2005). Extreme Value Theory and Extremely Large Electricity Price Changes. *International Review of Economics and Finance,* 14 (1):41-55.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). *The Econometrics of Financial Markets.* Princeton, NJ: Princeton University Press.

Campbell, J.Y. and Thompson, S.B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies,* 21 (4): 1509-1531.

Cappiello, L. (2000). Do Fixed Income Securities also show Asymmetric Effects in Conditional Second Moments? *Financial Asset Management and Engineering, Working Paper No. 12*.

Cappiello, L., Engle, R.F. and Sheppard, K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, 4 (4):537-572.

Carmona, R. (2014). *Statistical Analysis of Financial Data in R* (2nd ed.). New York, NY: Springer.

Cerrato, M., Crosby, J., Kim, M. and Zhao, Y. (2015). Modeling Dependence Structure and Forecasting Market Risk with Dynamic Asymmetric Copula. *Working Paper,* Adam Smith Business School, University of Glasgow: Glasgow, UK.

Chakraborti, A., Toke, I. M., Patriarca, M. and Abergel, F. (2011). Econophysics Review: I. Empirical Facts. *Quantitative Finance,* 11 (7):991-1012.

Chavez-Demoulin, V. and Embrechts, P. (2004). Smooth Extremal Models in Finance and Insurance. *Journal of Risk and Insurance,* 71 (2):183-199.

Chavez-Demoulin, V., Davison, A.C. and McNeil, A.J. (2005). Estimating Value-at-Risk: A Point Process Approach. *Quantitative Finance,* 5 (2):227-234.

Chavez-Demoulin, V., Embrechts, P. and Sardy, S. (2014). Extreme-quantile Tracking for Financial Time Series. *Journal of Econometrics,* 181 (1):44-52.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance.* Chichester, UK: John Wiley & Sons.

Chicheportiche, R. and Bouchaud, J.-P. (2012). The Joint Distribution of Stock Returns is Not Elliptical. *International Journal of Theoretical and Applied Finance,* 15 (3):1-23.

Chinzara, Z. and Aziakpono, M.J. (2009). Dynamic Returns Linkages and Volatility Transmission between South African and World Major Stock Markets. *Studies in Economics and Econometrics,* 33 (3):69-94.

Christoffersen, P., Errunza, V., Jacobs, K. and Langlois, H. (2012). Is the Potential for International Diversification Disappearing? A Dynamic Copula Approach. *Review of Financial Studies,* 25 (12):3711-3751.

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance,* 1 (2):223-236.

Cotter, J. and Stevenson, S. (2007). Uncovering Volatility Dynamics in Daily REIT Returns. *Journal of Real Estate Portfolio Management,* 13 (2):119-128.

Cryer, J.D. and Chan, K.S. (2008). *Time Series Analysis with Applications in R* (2nd ed.). New York, NY: Springer.

Daníelsson, J. and de Vries, C.G. (1997). Beyond the Sample: Extreme Quantile and Probability Estimation. *Financial Markets Group, Discussion Paper No. 298,* London School of Economics: London, UK.

Daníelsson, J. (2002). The Emperor has No Clothes: Limits to Risk Modeling. *Journal of Banking and Finance,* 26 (7):1273-1296.

Daníelsson, J. (2011). *Financial Risk Forecasting.* Chichester, UK: John Wiley & Sons.

Daníelsson, J., Jorgensen, B.N., Samorodnitsky, G., Sarma, M. and de Vries, C.G. (2013). Fat Tails, VaR and Subadditivity. *Journal of Econometrics,* 172 (2):283-291.

Davis, R.A. and Mikosch, T. (2009). Extreme Value Theory for GARCH Processes. In T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (Eds.), *Handbook of Financial Time Series* (pp. 187-200). Berlin, Germany: Springer.

Davison, A.C. and Smith, R.L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52 (3):393-442.

Delatte, A.L. and Lopez, C. (2013). Commodity and Equity Markets: Some Stylized Facts from a Copula Approach. *Journal of Banking and Finance,* 37 (12): 5346-5356.

Demarta, S. and McNeil, A.J. (2005). The *t* Copula and Related Copulas. *International Statistical Review,* 7 (1):111-129.

Dempster, M.A.H., Pflug, G. and Mitra, G. (Eds.) (2009). *Quantitative Fund Management.* Boca Raton, FL: Chapman & Hall / CRC.

Diebold, F.X. (1986). Modeling the Persistence of Conditional Variances: A Comment. *Econometric Reviews,* 5 (1):51-56.

Diebold, F.X., Schuermann, T. and Stroughair, J.D. (2000). Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. *Journal of Risk Finance,* 1 (2):30-35.

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance,* 1 (1):83-106.

Duncan, A.S. and Kabundi, A. (2011). Volatility Spillovers across South African Asset Classes during Domestic and Foreign Financial Crises. *Economic Research Southern Africa, Working Paper No. 202*.

Duncan, A.S. and Kabundi, A. (2014). Global Financial Crises and Time-Varying Volatility Comovement in World Equity Markets. *South African Journal of Economics,* 82 (4):531-550.

Dupuis, D.J. (1999). Exceedances over High Thresholds: A Guide to Threshold Selection. *Extremes,* 1 (3):251-261.

Embrechts, P., McNeil, A.J. and Straumann, D. (1999). Correlation: Pitfalls and Alternatives. *Risk,* 12 (5):93-113.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (2001). *Modelling Extremal Events for Insurance and Finance* (3rd ed.). New York, NY: Springer.

Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling Dependence with Copulas and Applications to Risk Management. In S.T. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance* (pp. 329-384). Amsterdam, The Netherland: Elsevier.

Embrechts, P. and Hofert, M. (2014). Statistics and Quantitative Risk Management for Banking and Insurance. *Annual Review of Statistics and Its Application,* 1 (1):493-514.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica,* 50 (4):987-1007.

Engle, R.F. and Bollerslev, T. (1986). Modelling the Persistence of Conditional Variances. *Econometric Reviews,* 5 (1):1-50.

Engle, R.F., Lilien, D.M. and Robins, R.P. (1987). Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model. *Econometrica,* 55 (2):391-407.

Engle, R.F. and Ng, V.K. (1993). Measuring and Testing the Impact of News on Volatility. *Journal of Finance,* 48 (5):1749-1778.

Engle, R.F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, 20 (3):339-350.

Ergen, I. (2010). VaR Prediction for Emerging Stock Markets: GARCH Filtered Skewed t Distribution and GARCH Filtered EVT Method. *Working Paper.* Federal Reserve Bank of Richmond, Baltimore, MD.

European Central Bank (1999). *Monthly Bulletin, February* [Online]. Retrieved 12 March 2015 from https://www.ecb.europa.eu/pub/pubbydate/1999/html/index.en.html

Fabozzi, F.J., Focardi, S.M. and Jonas, C. (2007). Trends in Quantitative Equity Management: Survey Results. *Quantitative Finance,* 7 (2):115-122.

Fabozzi, F.J., Kolm, P.N., Pachamanova, D.A. and Focardi, S.M. (2007). *Robust Portfolio Optimization and Management.* Hoboken, NJ: John Wiley & Sons.

Fama, E.F. (1963). Mandelbrot and the Stable Paretian Hypothesis. *Journal of Business,* 36 (4):420-429.

Fama, E.F. (1965). The Behavior of Stock-Market Prices. *Journal of Business,* 38 (1):34-105.

Fang, H.B., Fang, K.T. and Kotz, S. (2002). The Meta-elliptical Distributions with Given Marginals. *Journal of Multivariate Analysis,* 82 (1):1-16.

Fenn, D.J., Porter, M.A., McDonald, M., Williams, S., Johnson, N.F. and Jones, N.S. (2011). Temporal Evolution of Financial-market Correlations. *Physical Review E,* 84 (2):026109.

Fernández, C. and Steel, M.F.J. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association,* 93 (441):359-371.

Fernandez, V. (2005). Risk Management under Extreme Events. *International Review of Financial Analysis,* 14 (2):113-148.

Ferreira, J.T.A.S. and Steel, M.F.J. (2006). A Constructive Representation of Univariate Skewed Distributions. *Journal of the American Statistical Association,* 101 (474):823-829.

Fischer, M., Köck, C., Schlüter, S. and Weigert, F. (2009). An Empirical Analysis of Multivariate Copula Models. *Quantitative Finance,* 9 (7):839-854.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Members of a Sample. *Proceedings of the Cambridge Philosophical Society,* 24:180-190.

Fisher, T.J. and Gallagher, C.M. (2012). New Weighted Portmanteau Statistics for Time Series Goodness of Fit Testing. *Journal of the American Statistical Association,* 107 (498):777-787.

Focardi, S.M. and Fabozzi, F.J. (2004). *The Mathematics of Financial Modeling and Investment Management.* Hoboken, NJ: John Wiley & Sons.

Forbes, K. (2012). The 'Big C': Identifying and Mitigating Contagion. *National Bureau of Economic Research, Working Paper No. 18465*.

Francq, C. and Zakoïan, J.M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications.* Chichester, UK: John Wiley & Sons.

Fredericks, G.A. and Nelsen, R.B. (2007). On the Relationship between Spearman's Rho and Kendall's Tau for Pairs of Continuous Random Variables. *Journal of Statistical Planning and Inference,* 137 (7):2143-2150.

Furió, D. and Climent, F.J. (2013). Extreme Value Theory versus Traditional GARCH Approaches Applied to Financial Data: A Comparative Evaluation. *Quantitative Finance,* 13 (1):45-63.

Genest, C., Nešlehová, J. and Ghorbal, N.B. (2011). Estimators Based on Kendall's Tau in Multivariate Copula Models. *Australian and New Zealand Journal of Statistics,* 53 (2):157-177.

Ghalanos, A. (2012). Higher Moment Models for Risk and Portfolio Management. (*Unpublished doctoral dissertation).* City University London: London, UK.

Ghalanos, A. (2014a). *Introduction to the rugarch package.* R package version 1.3-1.

Ghalanos, A. (2014b). *rugarch: Univariate GARCH Models.* R package version 1.3-4.

Ghalanos, A. (2014c). *spd: Semi-Parametric Distribution.* R package version 2.0-0.

Ghalanos, A., Rossi, E. and Urga, G. (2015). Independent Factor Autoregressive Conditional Density Model. *Econometric Reviews,* 34 (5):594-616.

Ghorbel, A. and Trabelsi, A. (2009). Measure of Financial Risk using Conditional Extreme Value Copulas with EVT Margins. *Journal of Risk,* 11 (4):51-85.

Glosten, L. R., Jagannathan, R. and Runkle, D.E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance,* 48 (5):1779-1801.

Goyal, A. and Welch, I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies,* 21 (4):1455-1508.

Graves, S. (2015). *FinTS: Companion to Tsay (2005) Analysis of Financial Time Series.* R package version 0.4-5.

Grégoire, V., Genest, C. and Gendron, M. (2008). Using Copulas to Model Price Dependence in Energy Markets. *Energy Risk,* 5 (5):58-64.

Grimshaw, S.D. (1993). Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics,* 35 (2):185-191.

Gujarati, D.N. and Porter, D.C. (2009). *Basic Econometrics* (5th ed.). New York, NY: McGraw-Hill.

Hansen, P.R. and Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *Journal of Applied Econometrics,* 20 (7):873-889.

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2014). *copula: Multivariate Dependence with Copulas.* R package version 0.999-13.

Huisman, R., Koedijk, K., Kool, C. and Palm, F. (2002). The Tail Fatness of FX Returns Reconsidered. *De Economist,* 150 (3):299-312.

Hutson, E. and Stevenson, S. (2008). Asymmetry in REIT Returns. *Journal of Real Estate Portfolio Management,* 14 (2):105-124.

Idzorek, T.M. (2006). Developing Robust Asset Allocations. *Ibbotson Associates, Working Paper*.

Inanoglu, H. and Ulman, S. (2009). *Revisiting Copula Dependency Modeling: A Case for Conservatism, Federal Deposit Insurance Corporation* [Online]. Retrieved 12 July 2014 from https://www.fdic.gov/bank/analytical/cfr/2009/Seminar2009/seminar_series_2009.html

Institute of International Finance (2014). *Financial Globalization: Maximizing Benefits, Containing Risks* [Online]. Retrieved 11 March 2015 from https://www.iif.com/file/7020/download?token=EiCOCQDk

International Monetary Fund (2011). *Global Financial Stability Report, September 2011* [Online]. Retrieved 10 February 2015 from http://www.imf.org/external/pubs/ft/gfsr/2011/02

International Monetary Fund (2014). *Global Financial Stability Report, October 2014* [Online]. Retrieved 10 February 2015 from http://www.imf.org/external/pubs/cat/longres.aspx?sk=41631

International Monetary Fund (2015). *Global Financial Stability Report, April 2015* [Online]. Retrieved 10 February 2015 from http://www.imf.org/external/pubs/cat/longres.aspx?sk=42422.0

Jalal, A. and Rockinger, M. (2008). Predicting Tail-related Risk Measures: The Consequences of using GARCH Filters for Non-GARCH Data. *Journal of Empirical Finance,* 15 (5):868-877.

Jansen, D.W. and de Vries, C.G. (1991). On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *Review of Economics and Statistics,* 73 (1):18-24.

Jenkinson, A.F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) Value of Meteorologica Elements. *Quarterly Journal of the Royal Meteorological Society,* 81 (348):158-171.

Joe, H. (2015). *Dependence Modeling with Copulas.* Boca Raton, FL: CRC Press.

Johnson, N.L. (1949). Systems of Frequency Curves Generated by Methods of Translation. *Biometrika,* 36:149-176.

Jondeau, E., Poon, S.-H. and Rockinger, M. (2007). *Financial Modeling Under Non-Gaussian Distributions.* London, UK: Springer.

J.P. Morgan (2011). Rise of Cross-asset Correlations. *Global Equity Derivatives & Delta One Strategy.* New York, NY: J.P. Morgan.

Kat, H.M. (2003). The Dangers of Using Correlation to Measure Dependence. *Journal of Alternative Investments,* 6 (2):54-58.

Kendall, M.G. (1938). A New Measure of Rank Correlation. *Biometrika,* 30 (1/2):81-93.

Kim, D. and Kon, S.J. (1994). Alternative Models for the Conditional Heteroscedasticity of Stock Returns. *Journal of Business,* 67 (4):563-598.

Klüppelberg, C., Kuhn, G. and Peng, L. (2007). Estimating the Tail Dependence Function of an Elliptical Distribution. *Bernoulli,* 13 (1):229-251.

Kole, E., Koedijk, K. and Verbeek, M. (2007). Selecting Copulas for Risk Management. *Journal of Banking and Finance,* 31 (8):2405-2423.

Kon, S.J. (1984). Models of Stock Returns—A Comparison. *Journal of Finance,* 39 (1):147-165.

Kowalski, C.J. (1972). On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 21 (1):1-12.

Krehbiel, T. and Adkins, L.C. (2005). Price Risk in the NYMEX Energy Complex: An Extreme Value Theory Approach. *Journal of Futures Markets,* 25 (4):309-337.

Krokhmal, P., Palmquist, J. and Uryasev, S. (2002). Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints. *Journal of Risk,* 4 (2):43-68.

Kuester, K., Mittnik, S. and Paolella, M.S. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics,* 4 (1):53-89.

Lakonishok, J., Shleifer, A. and Vishny, R.W. (1992). The Impact of Institutional Trading on Stock Prices. *Journal of Financial Economics,* 32 (1): 23-44.

Lane, P.R. and Milesi-Ferretti, G.M. (2008). The Drivers of Financial Globalization. *American Economic Review,* 98 (2):327-232.

Le, T.H. (2012). Essays on Multivariate Volatility Models: An Application to Emerging Financial Markets. (*Unpublished doctoral dissertation).* University of Birmingham: Birmingham, UK.

LeBaron, B. and Samanta, R. (2005). Extreme Value Theory and Fat Tails in Equity Markets. *Working Paper,* Brandeis University: Waltham, MA.

Lévy, M.P. (1940). Le Mouvement Brownien Plan. *American Journal of Mathematics,* 62 (1):487-550.

Li, B., Wang, T. and Tian, W. (2013). Risk Measures and Asset Pricing Models with New Versions of Wang Transform. In V.-N. Huynh, V. Kreinovich, S. Sriboonchitta and K. Suriya (Eds.), *Uncertainty Analysis in Econometrics with Applications* (pp. 155-167). New York, NY: Springer.

Lizieri, C., Satchell, S. and Qi, Z. (2007). The Underlying Return-generating Factors for REIT Returns: An Application of Independent Component Analysis. *Real Estate Economics,* 35 (4):569-598.

Ljung, G.M. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika,* 65 (2):297-303.

Longin, F.M. (1996). The Asymptotic Distribution of Extreme Stock Market Returns. *Journal of Business,* 69 (3):383-408.

Longin, F.M. (1999). Optimal Margin Level in Futures Markets: Extreme Price Movements. *Journal of Futures Markets,* 19 (2):127-152.

Longin, F.M. and Solnik, B. (2001). Extreme Correlation of International Equity Markets. *Journal of Finance,* 56 (2):649-676.

MacDonald, A., Scarrott, C.J., Lee, D., Darlow, B., Reale, M. and Russell, G. (2011). A Flexible Extreme Value Mixture Model. *Computational Statistics and Data Analysis,* 55 (6):2137-2157.

Malevergne, Y. and Sornette, D. (2006). *Extreme Financial Risks: From Dependence to Risk Management.* New York, NY: Springer.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business,* 36 (4):394-419.

Marimoutou, V., Raggad, B. and Trabelsi, A. (2009). Extreme Value Theory and Value at Risk: Application to Oil Market. *Energy Economics,* 31 (4):519-530.

Markowitz, H.M. (1952). Portfolio Selection. *Journal of Finance,* 7 (1):77-91.

Martellini, L. and Meyfredi, J.-C. (2007). A Copula Approach to Value-at-Risk Estimation for Fixed-Income Portfolios. *Journal of Fixed Income,* 17 (1):5-15.

Mashal, R. and Zeevi, A. (2002). Beyond Correlation: Extreme Co-movements Between Financial Assets. (*Unpublished manuscript*), Columbia University: New York, NY.

Mashal, R., Naldi, M. and Zeevi, A. (2003). On the Dependency of Equity and Asset Returns. *Risk,* 16 (10):83-87.

McFadden, D. (1978). Modeling the Choice of Residential Location. In A. Karlqvist, L. Lundqvist, F. Snickars and J.W. Wiebull (Eds.), *Spatial Interaction Theory and Planning Models* (pp. 75-96). Amsterdam, The Netherlands: North Holland.

McKinsey & Company (2013). Financial Globalization: Retreat or Reset? *McKinsey Global Institute.* New York, NY: McKinsey Global Institute.

McNeil, A. (1996). Estimating the Tails of Loss Severity Distributions using Extreme Value Theory. (*Technical report)*, ETH Zürich, Switzerland.

McNeil, A. and Saladin, T. (1997). The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions. *Proceedings of 28th International ASTIN Colloquium*.

McNeil, A. (1999). Extreme Value Theory for Risk Managers. (*Unpublished manuscript)*, ETH Zürich, Switzerland.

McNeil, A. and Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance,* 7 (3/4):271-300.

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

Merton, R.C. (1972). An Analytical Derivation of the Efficient Portfolio Frontier. *Journal of Financial and Quantitative Analysis,* 7 (4):1851-1872.

Meucci, A., Gan, Y., Lazanas, A. and Phelps, B. (2007). A Portfolio Manager's Guide to Lehman Brothers Tail Risk Model. *Lehman Brothers Publications*.

Meucci, A. (2011). A Short, Comprehensive, Practical Guide to Copulas. *GARP Risk Professional,* 22-27.

Mikosch, T. and Stărică, C. (2000). Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process. *Annals of Statistics,* 28 (5):1427-1451.

Miller, M.B. (2014). *Mathematics and Statistics for Financial Risk Management* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Mittnik, S. and Rachev, S.T. (1993). Modeling Asset Returns with Alternative Stable Distributions. *Econometric Reviews,* 12 (3):261-330.

Nelsen, R.B. (2006). *An Introduction to Copulas* (2nd ed.). New York, NY: Springer.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica,* 59 (2):347-370.

Newson, R. (2002). Parameters behind "Nonparametric" Statistics: Kendall's tau, Somers' D and Median Differences. *The Stata Journal,* 2 (1):45-64.

Ng, H.S. and Lam, K.P. (2006). How Does the Sample Size Affect GARCH Model? *Working Paper,* The Chinese University of Hong Kong: Hong Kong SAR, The People's Republic of China.

Nyblom, J. (1989). Testing for the Constancy of Parameters Over Time. *Journal of the American Statistical Association,* 84 (405):223-230.

Nyström, K. and Skoglund, J. (2002a). A Framework for Scenariobased Risk Management. *Swedbank, Group Financial Risk Control*.

Nyström, K. and Skoglund, J. (2002b). Univariate Extreme Value Theory, GARCH and Measures of Risk. *Swedbank, Group Financial Risk Control*.

O'Brian, J.M. and Szerszen, P.J. (2014). An Evaluation of Bank VaR Measures for Market Risk During and Before the Financial Crisis. *Finance and Economics Discussion Series Working Paper 2014-21.* Board of Governors of the Federal Reserve System, Washington, D.C.

Osborne, M. (1959). Brownian Motion in the Stock Market. *Operations Research,* 7 (2):145-173.

Owen, J. and Rabinovitch, R. (1983). On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice. *Journal of Finance,* 38 (3):745-752.

Pagan, A. (1996). The Econometrics of Financial Markets. *Journal of Empirical Finance,* 3 (1):15-102.

Palm, F.C. (1996). GARCH Models of Volatility. In G.S. Maddala and C.R. Rao (Eds.), *Handbook of Statistics, Volume 14* (pp. 209-240). Amsterdam, The Netherland: Elsevier.

Palmitesta, P. and Provasi, C. (2006). Maximum Likelihood Estimation of the APARCH Model with Skew Distributions for the Innovation Process. *Statistica Applicata,* 18 (3):499-520.

Paolella, M.S. (2007). *Intermediate Probability: A Computational Approach.* Hoboken, NJ: John Wiley & Sons.

Peiró, A. (1994). The Distribution of Stock Returns: International Evidence. *Applied Financial Economics,* 4 (6):431-439.

Pflug, G.C. (2000). Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk. In S. Uryasev (Ed.), *Probabilistic Constrained Optimization: Methodology and Applications* (pp. 272-281). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Philips, C.B. (2014). Global Equities: Balancing Home Bias and Diversification. *The Vanguard Group, Working Paper*.

Pickands, J. III (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics,* 3 (1):119-131.

Polakow, D.A. and Flint, E.J. (2014). Global Risk Factors and South African Equity Indices. *South African Journal of Economics.* DOI:10.1111/saje.12065.

Prause, K. (1999). The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures. (*Unpublished doctoral dissertation).* University of Freiburg: Freiburg, Germany.

PricewaterhouseCoopers (2015). *Africa Asset Management 2020* [Online]. Retrieved 2 May 2015 from http://www.amafrica2020.com/amafrica2020/index.html

Quessy, J.-F. and Bellerive, R. (2013). Statistical Procedures for the Selection of a Multidimensional Meta-elliptical Distribution. *Journal de la Société Française de Statistique,* 154 (1):78-101.

R Development Core Team (2014). R: A Language and Environment for Statistical Computing (Version 3.1.0) [Software]. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org

Riccetti, L. (2013). A Copula-GARCH Model for Macro Asset Allocation of a Portfolio with Commodities. *Empirical Economics,* 44 (3):1315-1336.

Rigby, R.A. and Stasinopoulos, D.M. (2005). Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 54 (3):507-554.

Rigby, R.A. and Stasinopoulos, D.M. (2010). *A Flexible Regression Approach Using GAMLSS in R. Notes for the Athens 2010 Short Course* [Online]. Retrieved 8 February 2014 from http://www.gamlss.org/?p=1080

Rocco, M. (2010). Extreme Value Theory for Finance: A Survey. *Department of Mathematics, Statistics and Informatics Working Paper No. 3,* University of Bergamo: Bergamo, Italy.

Rockafellar, R.T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *Journal of Risk,* 2 (3):21-41.

Rodgers, J.L. and Nicewander, W.A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician,* 42 (1):59-66.

Rootzén, H. and Tajvidi, N. (1997). Extreme Value Statistics and Wind Storm Losses: A Case Study. *Scandinavian Actuarial Journal,* 1:70-94.

Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. New York, NY: Springer.

Sarykalin, S., Serraino, G. and Uryasev, S. (2008). Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization. In Z.-L. Chen, S. Raghavan, P. Gray and H.J. Greenberg (Eds.), *Tutorials in Operations Research: State-of-the-Art Decision-Making Tools in the Information-Intensive Age* (pp. 270-294) [Online]. Retrieved 25 March 2015 from http://pubsonline.informs.org/doi/book/10.1287/educ.1080

Scaillet, O. and Fermanian, J.-D. (2002). Nonparametric Estimation of Copulas for Time Series. *Financial Asset Management and Engineering, Research Paper No. 57*.

Scherer, B. (2002). *Portfolio Construction and Risk Budgeting*. London, UK: Risk Books.

Schmidt, T. (2006). Coping with Copulas. In J. Rank (Ed.), *Copulas: From Theory to Application in Finance* (pp. 3-34). London, UK: Risk Books.

Schweizer, B. and Wolff, E.F. (1981). On Non-parametric Measures of Dependence for Random Variables. *Annals of Statistics,* 9 (4):879-885.

Sheikh, A.Z. and Qiao, H. (2010). Non-Normality of Market Returns: A Framework for Asset Allocation Decision Making. *Journal of Alternative Investments,* 12 (3):8-35.

Six, T. and Wiedemann, A. (2013). *Scenario-based Asset Allocation* [Online]. Retrieved 10 May 2015 from www.union-investment.com/data/docme/news/Risk-management-study-2013/E_1_12_Scenario_based_asset_allocation_pdf/document/E_1.12_Scenario_based_asset_allocation.pdf

Sklar, A. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *l'Institut de statistique de l'Université de Paris,* 8:229-231.

Smith, R.L. (1987). Estimating Tails of Probability Distributions. *Annals of Statistics,* 15 (3):1174-1207.

So, M.K.P and Yu, P.L.H. (2006). Empirical Analysis of GARCH Models in Value at Risk Estimation. *Journal of International Financial Markets, Institutions and Money,* 16 (2):180-197.

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *American Journal of Psychology,* 15 (1):72-101.

Stefanova, K. (2015). *How to Disrupt the Investing Business with Katina Stefanova (Ex-Bridgewater Management Committee) / Interviewer: D. Teten* [Online]. Retrieved 28 April 2015 from http://disruptinvesting.com/2015/04/how-to-disrupt-the-investing-business-with-katina-stefanova-ex-bridgewater-management-committee

Stoyanov, S.V., Rachev, S., Racheva-Iotova, B. and Fabozzi, F.J. (2010). Stochastic Models for Risk Estimation in Volatile Markets: A Survey. *Annals of Operations Research,* 176 (1):293-309.

Sun, W., Rachev, S., Fabozzi, F.J. and Kalev, P.S. (2009). A New Approach to Modeling Co-movement of International Equity Markets: Evidence of Unconditional Copula-based Simulation of Tail Dependence. *Empirical Economics,* 36 (1): 201-229.

Szegö, G. (2002). Measures of Risk. *Journal of Banking and Finance,* 26 (7):1253-1272.

Teräsvirta, T. (2009). An Intoduction to Univariate GARCH models. In T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (Eds.), *Handbook of Financial Time Series* (pp. 17-42). Berlin, Germany: Springer.

Tokpavi, S. and Vaucher, B. (2012). Conditional Value-at-Risk: An Alternative Measure for Low-Risk Strategies? *Research Paper, Unigestion*.

Tower Capital (Opalesque TV) (2013). *Getting Liquid Access to African Frontier Markets* [Online]. Retrieved 30 April 2014 from http://www.opalesque.tv/hedge-fund-videos/tower-capital/1

Tsay, R.S. (2010). *Analysis of Financial Time Series* (3rd ed.). New York, NY: John Wiley & Sons.

Tsay, R.S. (2012). *An Introduction to Analysis of Financial Data with R*. Hoboken, NJ: John Wiley & Sons.

Tsuchida, N., Giacometti, R., Fabozzi, F.J., Kim, Y.S. and Frey, R.J. (2014). Time Series and Copula Dependency Analysis for Eurozone Sovereign Bond Returns. *Journal of Fixed Income,* 24 (1):75-87.

Tucker, A.L. (1992). A Reexamination of Finite- and Infinite-Variance Distributions as Models of Daily Stock Returns. *Journal of Business and Economic Statistics,* 10 (1):73-81.

Uryasev, S. (2000). *Conditional Value-at-Risk: Optimization Algorithms and Applications. Financial Engineering News, 14:1-5* [Online]. Retrieved 20 August 2014 from http://www.ise.ufl.edu/uryasev/publications

Wang, I. and Zheng, P. (2010). Rebalancing, Conditional Value at Risk, and t-Copula in Asset Allocation. *Working Paper,* Duke University: Durham, NC.

Wang, P., Sullivan, R.N. and Ge, Y. (2012). Risk-Based Dynamic Asset Allocation with Extreme Tails and Correlations. *Journal of Portfolio Management,* 38 (4):26-42.

Wang, W. (2006). *Stochasticity, Nonlinearity and Forecasting of Streamflow Processes*. Amsterdam, The Netherlands: IOS Press.

Wei, W., Li, J., Cao, L., Sun, J., Liu, C. and Li, M. (2013). Optimal Allocation of High Dimensional Assets through Canonical Vines. In J. Pei, V.S. Tseng, L. Cao, H. Motoda and G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining, Part 1* (pp. 366-377). New York, NY: Springer.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica,* 50 (1):1-25.

Wolff, E.F. (1980). N-dimensional Measures of Dependence. *Stochastica,* 4 (3):175-188.

Wolters, M.A. (2012). Methods for Shape-Constrained Kernel Density Estimation. (*Unpublished doctoral dissertation*). University of Western Ontario: London, Canada.

Wuertz, D. and Chalabi Y. (2009). *fGarch: Rmetrics – Autoregressive Conditional Heteroskedastic Modelling*. R package version 3010.82.

Wuertz, D., Setz, T., Chalabi, Y. and Rmetrics Core Team (2014). *fPortfolio: Rmetrics – Portfolio Selection and Optimization*. R package version 3011.81.

Xiong, J.X. (2010). Nailing Downside Risk. *Ibbotson Associates, Morningstar Advisor*.

Xiong, J.X. and Idzorek, T.M. (2010). Mean-Variance Versus Mean-Conditional Value-at-Risk Optimization: The Impact of Incorporating Fat Tails and Skewness into the Asset Allocation Decision. *Ibbotson Associates, Morningstar Advisor*.

Xiong, J.X. and Idzorek, T.M. (2011). The Impact of Skewness and Fat Tails on the Asset Allocation Decision. *Financial Analysts Journal,* 67 (2):23-35.

Yan, J. (2007) Enjoy the Joy of Copulas: With a Package copula. *Journal of Statistical Software,* 21 (4):1-21.

Yu, J. (2001). Testing for Finite Variance in Stock Return Distributions. In J. Knight and S. Satchell (Eds.), *Return Distributions in Finance* (pp. 143-164). Oxford, England: Butterworth-Heinemann.

Zakoïan, J.M. (1994). Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control,* 18 (5):931-955.

Zhu, L., Coleman, T.F. and Li, Y. (2009). Min-max Robust and CVaR Robust Mean-variance Portfolios. *Journal of Risk,* 11 (3):1-31.

Zivot, E. and Wang, J. (2006). *Modeling Financial Time Series with S-PLUS*® (2nd ed.). New York, NY: Springer.

Appendices

Appendix A: Box-and-Whisker Plot

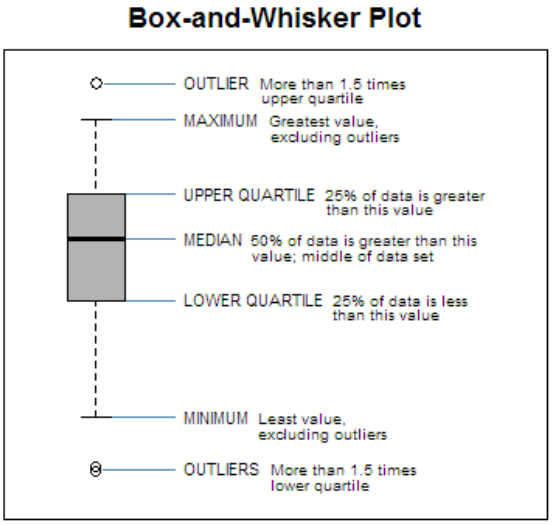


Figure A.1: Box-and-Whisker Plot

The box-and-whisker plot (boxplot) is a rectangular shaped box defined by the upper and lower quartiles of the data (i.e., 50% of the data falls inside the box) and is split by a median line (i.e., the 50^{th} percentile). The difference between the upper and lower quartile is called the interquartile range (IQR). There are two dashed lines, or whiskers. The whisker extending to the upper value ("maximum") marks the value of the largest observation that is less than or equal to the upper quartile plus 1.5 times the length of the IQR. Similarly, the whisker extending to the lower value ("minimum") marks the value of the smallest observation that is greater than or equal to the lower quartile less 1.5 times the length of the IQR. Outliers are observations that fall outside the lower or upper values. The boxplot gives a good indication of the shape of a distribution, including skewness and presence of outliers. For data from a normal distribution, outlier values occur fewer than once in every 100 observations.

Appendix B: In-Sample 2008 Financial Crisis Comparison

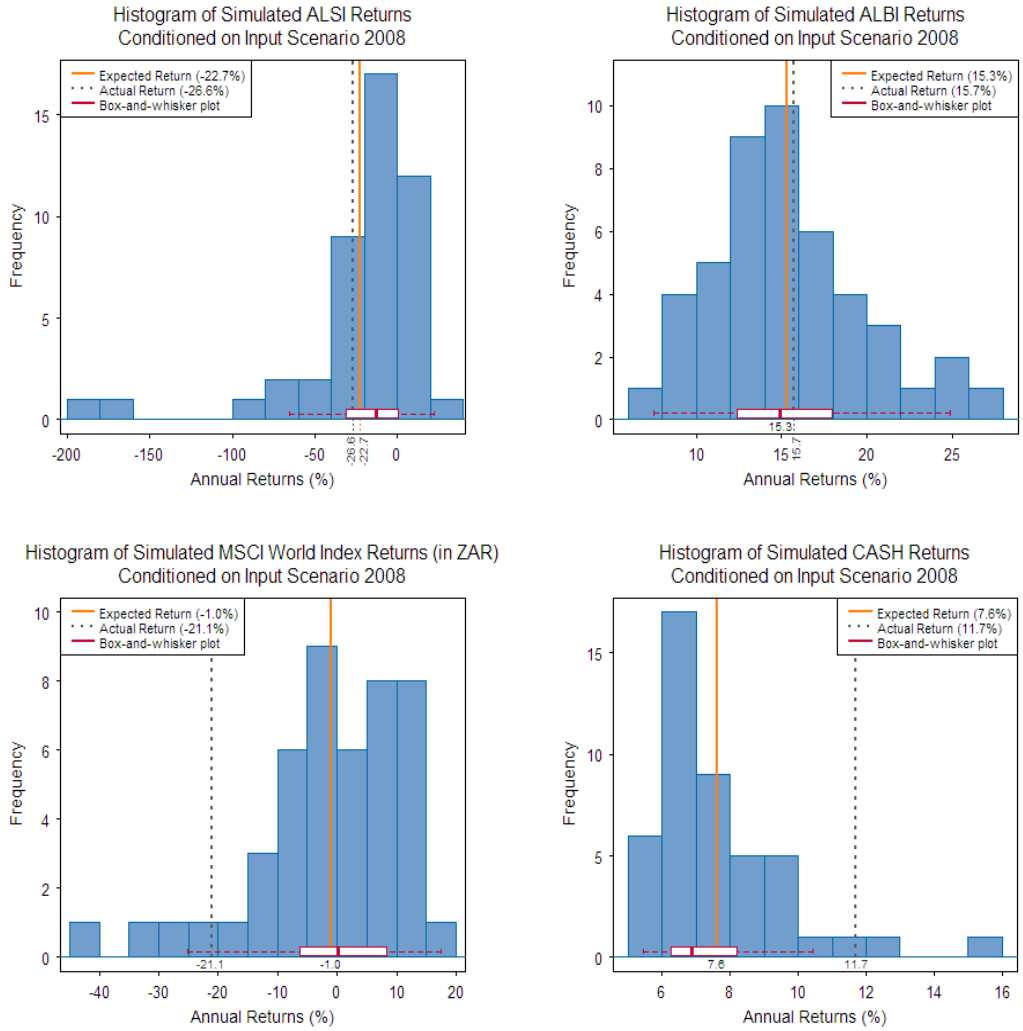


Figure B.1: Comparison of Expected Returns versus Actual Returns for the 2008 Financial Crisis Period

Table B.1: Risk Factor Range Forecasts for the 2008 Financial Crisis Period

Input Variables			Output Variables
Return Drivers	Range Forecasts	Actual	Asset Classes to be Optimised
USDZAR	+10% to +40%	+33.5%	FTSE/JSE All Share Index
EURUSD	-20% to +10%	-4.1%	FTSE/JSE All Bond Index
SA 10-Yr Bond Yield	0% to -30%	-13.1%	MSCI World Equity Index
Brent Crude	-20% to -99%	-93.3%	Domestic Cash Index
Commodities Index	-10% to -80%	-65.2%	
Gold Index	0% to +30%	+2.0%	
Platinum Index	-10% to -60%	-49.9%	
MSCI EM Index	-10% to -90%	-78.7%	
S&P 500 Index	-10% to -70%	-48.6%	

Range forecasts for key variables assume fairly accurate forecasts ahead of a repeat of the 2008 crash period. Actual returns for the period are included around which the range forecasts were built.

Appendix C: Ljung-Box and Lagrange Multiplier Test Results

Table C.1: *p*-Values from Ljung-Box Tests on Monthly Returns

Lag	ALSI	ALBI	MSCI.WRLD.ZAR	CASH	GLOUS	J253T	USDZAR	EURUSD	BRSPOT	GSCI	GLFX	PLAT	MSCI.EM$	FSPI	JPEMBI	USALCI	RBAS	JAYC10
1	0.568	0.606	0.510	0.000	0.089	0.209	0.713	0.754	0.652	0.453	0.084	0.013	0.011	0.146	0.833	0.155	0.000	0.620
2	0.846	0.852	0.607	0.000	0.109	0.361	0.719	0.692	0.885	0.484	0.130	0.034	0.017	0.331	0.554	0.028	0.000	0.878
3	0.512	0.895	0.439	0.000	0.203	0.563	0.864	0.252	0.796	0.571	0.165	0.048	0.022	0.178	0.599	0.026	0.000	0.967
4	0.549	0.368	0.350	0.000	0.258	0.465	0.942	0.179	0.799	0.657	0.124	0.065	0.042	0.197	0.740	0.045	0.000	0.389
5	0.226	0.412	0.486	0.000	0.115	0.403	0.799	0.277	0.887	0.617	0.202	0.058	0.062	0.264	0.846	0.016	0.000	0.426
6	0.303	0.136	0.551	0.000	0.171	0.294	0.632	0.344	0.943	0.736	0.266	0.098	0.029	0.284	0.879	0.005	0.000	0.106
7	0.331	0.184	0.338	0.000	0.026	0.396	0.675	0.261	0.221	0.293	0.319	0.106	0.043	0.352	0.933	0.010	0.000	0.118
8	0.407	0.225	0.327	0.000	0.031	0.362	0.413	0.352	0.174	0.131	0.418	0.009	0.061	0.341	0.963	0.019	0.000	0.159
9	0.313	0.304	0.304	0.000	0.019	0.402	0.444	0.426	0.229	0.186	0.507	0.010	0.093	0.423	0.976	0.029	0.000	0.212
10	0.352	0.230	0.259	0.000	0.024	0.450	0.522	0.435	0.292	0.210	0.589	0.014	0.130	0.514	0.959	0.015	0.000	0.205
11	0.437	0.271	0.160	0.000	0.006	0.533	0.481	0.524	0.015	0.067	0.086	0.022	0.137	0.581	0.912	0.019	0.000	0.222
12	0.520	0.281	0.203	0.000	0.005	0.510	0.564	0.513	0.014	0.032	0.119	0.012	0.182	0.501	0.906	0.028	0.000	0.235

At the 95% (or 90%) significance level, a *p*-value less than $\alpha = \mathbf{0.05}$ (or $\alpha = \mathbf{0.10}$) leads to the rejection of the null of no autocorrelation present in the time series (i.e. there is autocorrelation in the series). For a given lag length, m, a *p*-value near zero indicates that a time series has m jointly significant autocorrelations. The colour differentiation in the tables indicate where and at which significance level there is significant autocorrelation present.

Table C.2: *p*-Values from Ljung-Box Tests on Absolute Monthly Returns

Lag	ALSI	ALBI	MSCI.WRLD.ZAR	CASH	GLOUS	J253T	USDZAR	EURUSD	BRSPOT	GSCI	GLFX	PLAT	MSCI.EM$	FSPI	JPEMBI	USALCI	RBAS	JAYC10
1	0.368	0.000	0.008	0.000	0.022	0.039	0.000	0.725	0.000	0.149	0.003	0.000	0.058	0.004	0.001	0.070	0.000	0.000
2	0.001	0.000	0.003	0.000	0.074	0.028	0.000	0.119	0.000	0.015	0.005	0.000	0.010	0.000	0.000	0.138	0.000	0.000
3	0.002	0.000	0.005	0.000	0.146	0.001	0.001	0.205	0.000	0.032	0.007	0.000	0.011	0.000	0.001	0.166	0.000	0.000
4	0.002	0.000	0.007	0.000	0.160	0.000	0.001	0.212	0.000	0.064	0.000	0.000	0.009	0.000	0.001	0.167	0.000	0.000
5	0.001	0.000	0.000	0.000	0.251	0.000	0.000	0.027	0.000	0.012	0.000	0.000	0.017	0.000	0.003	0.262	0.000	0.000
6	0.001	0.000	0.000	0.000	0.196	0.000	0.000	0.036	0.000	0.013	0.000	0.000	0.005	0.000	0.006	0.216	0.000	0.001
7	0.000	0.000	0.001	0.000	0.255	0.000	0.000	0.056	0.000	0.014	0.000	0.000	0.001	0.000	0.006	0.287	0.000	0.002
8	0.000	0.000	0.001	0.000	0.261	0.000	0.000	0.089	0.000	0.025	0.000	0.000	0.001	0.000	0.002	0.299	0.000	0.003
9	0.000	0.000	0.002	0.000	0.128	0.000	0.000	0.105	0.000	0.041	0.000	0.000	0.001	0.000	0.001	0.310	0.000	0.005
10	0.000	0.000	0.003	0.000	0.177	0.000	0.000	0.117	0.000	0.053	0.000	0.000	0.000	0.000	0.000	0.374	0.000	0.006
11	0.000	0.000	0.005	0.000	0.057	0.000	0.000	0.158	0.000	0.079	0.000	0.000	0.001	0.000	0.000	0.423	0.000	0.004
12	0.000	0.000	0.004	0.000	0.056	0.000	0.000	0.197	0.000	0.028	0.000	0.000	0.001	0.000	0.000	0.057	0.000	0.001

At the 95% (or 90%) significance level, a *p*-value less than $\alpha = \mathbf{0.05}$ (or $\alpha = \mathbf{0.10}$) leads to the rejection of the null of no autocorrelation present in the time series (i.e. there is autocorrelation in the series). For a given lag length, m, a *p*-value near zero indicates that a time series has m jointly significant autocorrelations. The colour differentiation in the tables indicate where and at which significance level there is significant autocorrelation present.

Table C.3: *p*-Values from LM Tests for ARCH Effects in the Square of Demeaned Monthly Returns																		
Lag	ALSI	ALBI	MSCI.WRLD.ZAR	CASH	GLOUS	J253T	USDZAR	EURUSD	BRSPOT	GSCI	GLFX	PLAT	MSCI.EM$	FSPI	JPEMBI	USALCI	RBAS	JAYC10
1	0.877	0.000	0.039	0.000	0.060	0.012	0.033	0.569	0.007	0.092	0.008	0.000	0.057	0.000	0.450	0.041	0.009	0.000
2	0.023	0.000	0.039	0.000	0.164	0.001	0.103	0.000	0.001	0.003	0.018	0.000	0.079	0.001	0.714	0.033	0.012	0.001
3	0.052	0.000	0.043	0.000	0.308	0.001	0.133	0.000	0.003	0.009	0.044	0.000	0.125	0.000	0.876	0.078	0.017	0.003
4	0.104	0.000	0.072	0.000	0.442	0.001	0.242	0.000	0.004	0.018	0.006	0.000	0.222	0.000	0.951	0.151	0.035	0.006
5	0.168	0.000	0.000	0.000	0.570	0.002	0.239	0.000	0.006	0.017	0.011	0.000	0.324	0.000	0.982	0.249	0.065	0.012
6	0.255	0.000	0.001	0.000	0.568	0.002	0.025	0.000	0.012	0.033	0.023	0.000	0.319	0.000	0.992	0.347	0.094	0.021
7	0.343	0.000	0.001	0.000	0.517	0.004	0.047	0.000	0.019	0.020	0.043	0.000	0.191	0.000	0.997	0.330	0.147	0.031
8	0.297	0.000	0.001	0.000	0.629	0.007	0.078	0.001	0.028	0.035	0.043	0.000	0.265	0.001	0.999	0.418	0.214	0.052
9	0.380	0.000	0.002	0.000	0.599	0.010	0.122	0.001	0.048	0.034	0.070	0.000	0.347	0.002	0.999	0.514	0.292	0.079
10	0.337	0.000	0.004	0.000	0.694	0.014	0.093	0.002	0.074	0.055	0.109	0.000	0.235	0.003	0.942	0.619	0.380	0.118
11	0.384	0.000	0.008	0.000	0.031	0.023	0.120	0.003	0.086	0.085	0.158	0.000	0.262	0.005	0.966	0.500	0.459	0.147
12	0.413	0.001	0.008	0.000	0.022	0.033	0.090	0.003	0.115	0.098	0.184	0.000	0.292	0.008	0.979	0.265	0.544	0.168

At the 95% (or 90%) significance level, a *p*-value less than $\alpha = \mathbf{0.05}$ (or $\alpha = \mathbf{0.10}$) leads to the rejection of the null of no ARCH effect present in the time series (i.e. there is conditional heteroskedasticity in the series). For a given lag length, m, a *p*-value near zero indicates that a time series has m jointly significant autocorrelations in the square of the demeaned data. The colour differentiation in the tables indicate where and at which significance level there is significant ARCH effect present. Note that the `R` function `ArchTest` from the `FinTS` package (Graves, 2015) squares the demeaned data before performing the linear regression.