

Imperial College London
Department of Computing

Applications of Copulas in Optimisation

Iakovos Kakouris

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computing of Imperial College and
the Diploma of Imperial College, November 2013

To my grand parents Loukia, Eleni, George, Iakovos

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Declaration

This thesis presents my work in the Department of Computing at Imperial College London between October 2009 and November 2013.

I declare that the work presented in this thesis is my own, except where acknowledged.

Abstract

The methods for modelling uncertainty and assessing the risk of financial markets were placed under scrutiny after the 2008 crisis. To protect against the worst possible scenario, in a problem of asset allocation, robust optimisation is required. Still, within this framework, assumptions about the uncertainty set have to be made. In our work, we expand the possible options for describing uncertainty sets, through the use of *copulas*.

Copulas are a useful tool for describing uncertainty because of the modelling flexibility that they provide. They are able to easily describe *asymmetric dependence structures* and *tail risk*. Both are vital for emulating the financial markets behaviour, during periods of extreme shocks and comovements. Also, copulas are associated with robust measures of dependence.

We introduce copulas into the robust optimisation framework by following two different approaches. At first, we derive a *Worst Case Conditional Value at Risk* optimisation problem, in which the uncertainty set consists of a selection of copulas. We formulate the problem into a convex optimisation problem. The advantages of such a model are supported by numerical examples using real data. In the second approach, copulas are used as means for creating non-symmetric, convex uncertainty sets, in the form of domains. We present examples where these sets can be used in a robust optimisation problem.

Acknowledgements

I would like to thank everyone who helped and supported me throughout my Ph.D. and the writing of my thesis. I would particularly like to thank my supervisor Prof. Berç Rustem for his support and guidance, my co-supervisor Dr. Daniel Kuhn and Prof. Nicos Christofides.

A big thank you to my parents Andreas and Helen for their ethical and financial support throughout my studies, my flat-mates Nicodemos, Andreas, Kalamaras and my friends Angelos K., Lina for checking on me and for our long discussions. My friends from the Computational Optimization Group Michalis, Vladimir, Angelos G., Ryan, Wolfram, Kapsos, Panos, Christos, Clint, Phebe, Paula, Dimitra, Grani, Sei, Tran, Napat and my fellow PhD students Alexandra, Panagiotis, Prokopiou for the relaxing pub nights and the exchange of ideas and support. I should not forget the shisha fun club Maria, Athena, Katerina, Marianna, Elenitsa, Gavriella, Thalia and many more for the chill-out time spend together. Last but not least, a big thank you to my friends back home, my buddies Vasilis, Papas, Marios, Konstantinos, Eliana, for putting a big smile in my face whenever I visited.

To all those friends and relatives that I do not mention thank you as well.

Contents

1	Introduction	15
1.1	Notation	19
2	Optimisation	21
2.1	Convex optimisation	22
2.2	Duality	25
2.3	Examples	27
3	Copula Review	30
3.1	Copulas	31
3.2	Measures of Dependence	36
3.3	Methods for Constructing Copulas	40
3.3.1	Inversion	40
3.3.2	Shuffles	42
3.3.3	Archimedian Copulas	43
3.3.4	Non-Parametric Copulas	47
3.3.5	Mixture Copula	48

3.4	Copula Calibration Techniques	50
3.4.1	Maximum Likelihood based Methods	50
3.4.2	Calibration Methods Using Measures of Dependence	53
3.5	Examples	54
3.5.1	Modelling flexibility	54
3.5.2	Comparing quantile dependence	56
4	Robust Portfolio Optimisation with Mixture Copulas	60
4.1	Worst Case Conditional Value at Risk	63
4.1.1	Mixture Copula	66
4.2	Portfolio management under a worst case copula scenario	69
4.3	Numerical examples and empirical analysis	72
4.3.1	Fixed estimation window portfolio	73
4.3.2	Rolling estimation window portfolio	79
4.3.3	Performance analysis	81
4.3.4	Considering transaction cost	87
4.4	Summary, future work and conclusions	90
5	Non-symmetric convex uncertainty sets	97
5.1	Defining uncertainty sets through the use of copulas	99
5.1.1	Constructing a convex set	101
5.2	Formulating a convex optimisation problem	105
5.3	Summary, future work and conclusions	106

6 Summary and Conclusions	108
Bibliography	111
A	117
A.1 Definitions for sets	117
B	118
B.1 Theorem 3.3.6	118
C	120
C.1 Shares used from Mexican and Hong Kong stock exchanges	120
C.2 $\mathcal{P}_t^{tr}$ linear programming reformulation	121
D Notation and Acronyms	123
D.1 Notation	123
D.2 Acronyms	125

List of Tables

4.1	The mean, variance and covariance of the seven assets daily returns between November 1998 and June 2003	75
4.2	Comparison of the performance of Gaussian optimal, Worst Case optimal and Worst Case Markowitz optimal portfolios	77
4.3	Average daily return of Gaussian optimal, Worst Case optimal and Worst Case Markowitz optimal portfolios up to January 2008	77
4.4	Comparison of the performance of Gaussian optimal, the Worst Case optimal and Equally Weighted portfolios	82
4.5	Comparison of the performance of Gaussian optimal, the Worst Case optimal and Equally Weighted portfolios for Mexican stock exchange.	83
4.6	Comparison of the performance of Gaussian optimal, the Worst Case optimal and Equally Weighted portfolios for Hong Kong stock exchange.	85
4.7	Indices rolling estimation window portfolios: Worst periods of depreciation (Highest point/ Lowest point %).	85
4.8	Mexican stock exchange rolling estimation window portfolios: Worst periods of depreciation (%).	86
4.9	Hong Kong stock exchange rolling estimation window portfolios: Worst periods of depreciation (%).	86

4.10	Average turnover when the portfolios are rebalanced on daily basis	88
------	--	----

List of Figures

3.1	Gaussian with $\rho = 0.891$ and $\tau = 0.7$	41
3.2	Clayton with $\alpha = 4.67$ and $\tau = 0.7$	45
3.3	Gumbel with $\alpha = 3.33$ and $\tau = 0.7$	45
3.4	Frank with $\alpha = 11.41$ and $\tau = 0.7$	46
3.5	Gaussian probability density function with $X_1, X_2 \sim N(0, 2)$ and $r = 0.5$	54
3.6	Combining the Gaussian copula and chi-square margins with six degrees of freedom, $X_1, X_2 \sim \chi_6^2$	55
3.7	Combining the Clayton copula with a Gaussian $X_1 \sim N(0, 2)$ and a chi-square $X_2 \sim \chi_6^2$	56
3.8	The lower quantile dependence ν_L and upper quantile dependence ν_U of Gaussian, Clayton, Gumbel and Frank copulas for $\tau = 0.7$. The values at the boundaries give the lower tail dependence λ_L and upper tail dependence λ_U .	57
3.9	The diagonal of the empirical copula density function between FTSE and DAX 30 returns.	58
3.10	Empirical quantile dependence of FTSE and DAX 30 returns.	59
4.1	Seven indices from 1998 to 2011. All the indices are normalised to 1 on July 2003	74

4.2	Fixed estimation window portfolios: ‘Out of Sample’ portfolio performance for $\pi_r = 0$ (see equation (4.34))	76
4.3	Rolling estimation window portfolios performance with $\pi_r = \mathbf{w}_{eq}^T \hat{\boldsymbol{\mu}}$ (see equation (4.38))	81
4.4	Mexican stock exchange rolling estimation window portfolios performance with $\pi_r = \mathbf{w}_{eq}^T \hat{\boldsymbol{\mu}}$ (see equation (4.38))	83
4.5	Hong Kong stock exchange rolling estimation window portfolios performance with $\pi_r = \mathbf{w}_{eq}^T \hat{\boldsymbol{\mu}}$ (see equation (4.38))	84
4.6	Indices rolling estimation window portfolios: 30 trading days performance analysis (compounded return).	94
4.7	Mexican stock exchange rolling estimation window portfolios: 30 trading days performance analysis (compounded return).	95
4.8	Hong Kong rolling estimation window portfolios: 30 trading days performance analysis (compounded return).	96
5.1	The dark red area satisfies $\phi(x_1, x_2) \geq \gamma$. The grey area in $\mathbb{R}^2$ represents the corresponding uncertainty set, $S_x(\gamma) = \{\mathbf{x} \in \mathbb{R}^n : \phi(\mathbf{x}) \geq \gamma\}$	98
5.2	The dark red area satisfies $c_{CI}(\Phi_1(x_1), \Phi_2(x_2))/(f_1(x_1)f_2(x_2)) \geq \gamma$. The grey area in $\mathbb{R}^2$ represents the corresponding uncertainty set.	100
5.3	Contour plot of $c_{CI}(u_1, u_2)$. The different level curves represent different values of γ , (5.3).	101
5.4	Contour plot of $\log(f(\mathbf{x})) = \log(c_{CI}(\Phi_1(x_1), \Phi_2(x_2))/(f_1(x_1)f_2(x_2)))$. The different level curves represent different values of $\log(\gamma)$, (5.3).	102
5.5	The set $\mathbb{P}$ in blue, bounding $S_x(\gamma)$ in dark red.	104
C.1	Mexican stock exchange shares from 2000 to 2011. All the indices are normalized at the beginning of the period	120

C.2 Hong Kong stock exchange shares from 2000 to 2011. All the indices are normalized at the beginning of the period	121
---	-----

Chapter 1

Introduction

From very early times, people started to realise the need to model uncertainty and reduce risk within their daily operations. The need for monitoring risk was particularly evident in financial markets. People tried to give a mathematical and economical reasoning why the market movements are happening. Modigliani and Miller [28] introduced a theory for assessing the value of a firm and as an extension its share price, a theory that provided the basis for modern thinking on firms capital structure. Itô's work [21, 22] formed the basis for Itô calculus which extends the theory of calculus to stochastic processes. Itô calculus was subsequently used by Black and Scholes in their groundbreaking work on option pricing [5] and by Hull and White in their widely used interest rate model [20].

While the above dealt with modelling the movements of financial assets, Markowitz [25] was one of the pioneers in proposing a risk minimisation model. Markowitz [25] suggested a mean variance framework for minimising the risk within a portfolio of financial assets, with the risk of the assets described by the covariance matrix. Since Markowitz, a lot of schools of thoughts for describing uncertainty within an optimisation framework have been developed. The most prevailing ones are stochastic programming, worst case optimisation and robust optimisation, with the latter two being very closely related.

In stochastic programming, the random variable is assumed to be described by some probability distribution, which is either given or estimated. This distribution is usually used to create a scenario tree based analysis, with either a random number of simulations from the assumed distribution, e.g., using Monte-Carlo techniques, or structured scenario trees that satisfy the moments of the assumed distribution, such as binomial trees. This way an approximation of the uncertainty space via possible realisations of the random variable is created. The optimisation problem in this case, has to be evaluated using these realisations [31]. The stochastic programming problem is prone to both systematic and estimation errors, which stem from the modelling distribution of the random variable and the estimation of the distribution's free parameters. Both can significantly affect the outcome of the optimisation problem, which makes stochastic programming a sensitive methodology.

This issue can be overcome by employing worst-case and robust optimisation techniques in which assumptions about the distribution of the random variable are relaxed. Using robust optimisation we obtain the best possible solution by optimising against all possible scenarios defined by a prescribed set. The prescribed set is selected in way that is representative of the random variable's uncertainty space. For example, the set may contain 95% of the instances that the random variable can obtain. Hence, by giving a set of possible realisations, we avoid making assumptions about the distribution of the random variable. In worst case optimisation, some characteristics of the random variable are used to describe a set of possible distributions. The worst case distribution from this set governs the scenarios over which the optimal solution is computed. Consequently, such a solution will also be optimal over the set of all possible scenarios hence, establishing the equivalence with robust optimisation. Thus, in this work we will consider robust and worst case scenario optimisation as equivalent problems. An example in which both methods coincide, explained in detail in the following chapter, is the case where the first and the second moments of the random variable are given. Worst case optimisation will optimise against the worst possible distribution that satisfies these moments, resulting in an ellipsoidal uncertainty set, while in robust optimisation, given the ellipsoidal uncertainty set, it optimises against all possible scenarios. The outcome

is an equivalent problem formulation for both methods.

While the worst case and robust optimisation offer theoretically robust approaches, in practice this is difficult to be achieved. Due to computational tractability issues, the number of uncertainty sets that can be considered is limited. The decision maker usually attempts to construct a convex optimisation problem. As we will discuss in more detail later on, convex optimisation problems inherit a lot of properties that makes them attractive for practical applications. The most common uncertainty sets which result in a convex optimisation problem formulation are box uncertainty, ellipsoidal uncertainty and a set of pre-selected distributions [4, 15, 48, 49, 50]. The first two are very structured, symmetric uncertainty sets. Similarly, the most commonly used multivariate distributions are the Gaussian and the Student-t, which are characterised by linear correlation and symmetric dependence structures. Studies suggest that at least in the context of financial markets, assets exhibit stronger comovements during a crisis as opposed to prosperity [1, 18, 19]. Hence, when selecting the uncertainty set representative of a portfolio of financial assets, it should be able to capture the asymmetric properties of the financial markets. Thus, we need uncertainty sets that diverge from the classical ones. *In this study we will provide alternative uncertainty sets, that are able to provide more realistic, non-symmetric structures. We will construct these uncertainty sets through the use of copulas.*

Copulas are multivariate distribution functions, the one-dimensional margins of which are uniformly distributed on the closed interval $[0,1]$. Copulas were introduced by Sklar in 1959 [42] and they belong to a special family of probabilistic metric spaces that has the same properties as probability distributions [39]. Thus, copulas provide an alternative way of looking into probability distributions. As the name suggests, copulas *couple* marginal distribution functions. They are a probabilistic measure of distance or dependence between marginal distributions, the univariate distributions of the individual random variables, as opposed to probability distributions that describe directly the dependence between random variable.

In this study we define uncertainty sets through the use of copulas by following two different approaches. The first approach involved the use of mixture copulas, which are constructed through a convex combination of copulas. The mixture copula is introduced within a worst case optimisation framework, in which the copulas comprising the mixture copula, form the uncertainty set. This method provides alternative ways of pre-selecting the distributions which form the uncertainty set in the optimisation problem. As demonstrated throughout this thesis, copulas provide a vast array of possible multivariate distributions, which are easy to construct and calibrate. These copulas are characterised by asymmetric dependence structures that prove to be very useful in modelling multi-asset dependencies.

The second approach attempts to generate uncertainty sets beyond box and ellipsoidal uncertainty. For this analysis we utilise the idea of robust optimisation, that we are interested in the set that covers a high percentage of all possible realisations of the random variables, but not in the actual distribution that characterises them. To do so, we use copulas as a starting point for defining uncertainty sets. These copula-constructed uncertainty sets have interesting characteristics, that diverge from the usual uncertainty sets mentioned above.

With our work we provide alternative solutions and open new possibilities for worst case and robust optimisation frameworks. The above two ideas and the motivation behind them is introduced in more detail in the corresponding chapters, together with the related background literature.

The thesis is organised as follows. In Chapter 2 we formally define what an optimisation problem is, with a particular focus on convex optimisation. Copulas together with their background literature are discussed in Chapter 3. Chapter 4 is the first of two original contribution chapters. Here, we formally show how Value at Risk (VaR) and Conditional VaR (CVaR) risk measures can be formulated with the use of copulas. The CVaR formulation is then used to formulate a Worst Case CVaR optimisation problem through the use of mixture copulas. Detailed numerical analysis is performed to investigate the advantages and disadvantages of such a model. Chapter 5 is the second original contribution chapter,

in which we discuss the use of copulas in defining uncertainty sets. Chapters 4 and 5 also include discussion on possible future direction.

Part of the work in Chapter 4 can be found in

1. I. Kakouris and B. Rustem. *Robust portfolio optimization with copulas*. Under Review for European Journal of Operational Research.

Also, parts of the work in Chapter 4 were presented in EURO conference, Vilnius 2012, and EURO INFORMS conference, Rome 2013. The work in Chapter 5 can be found in

2. I. Kakouris and B. Rustem. *Non-symmetric uncertainty spaces, a copula approach*. Working Paper.

1.1 Notation

Throughout this thesis we will denote column vectors by bold, lower case letters, e.g., $\mathbf{x}$ and matrices by bold upper case letters, e.g., $\mathbf{M}$. By $\leq$ and $<$ we denote the component wise inequality and strict inequality between vectors. For the set of real numbers we use $\mathbb{R}$ and for the set of numbers between and including 0 and 1 by $\mathbb{I}$. $\mathbb{N}$ denotes the set of natural numbers. General sets will be denoted by S . For other, more specific sets, e.g., the set vector $\mathbf{w}$ belongs to, we use upper case letters with the same font as for real numbers, i.e. W . $S_1 \subseteq S_2$ and $S_1 \subset S_2$ denotes that S_1 is a subset and a strict subset of S_2 respectively.

μ is used to denote the mean, σ the standard deviation and Σ the covariance matrix. $C(\cdot)$ is used for copulas, $F(\cdot)$ for cumulative distribution functions, $\phi(\cdot)$ and $\Phi(\cdot)$ for the probability and cumulative density functions of Gaussian distribution. By $F^{-1}(\cdot)$ we denote the inverse of a function. **dom** is the domain and **ran** is the range.

By $\lim_{u \rightarrow a}$ we denote the limit as u tends to a . By $\lim_{u \rightarrow a^+}$ we denote the limit as u tends to a from above, i.e., $u \geq a$, and $\lim_{u \rightarrow a^-}$ we denote the limit as u tends to a from below, i.e., $u \leq a$,

Any additional notation is given and defined in each chapter. A full notation, abbreviations and acronyms list can be found in Appendix D.

Chapter 2

Optimisation

Most optimisation problems arise from everyday challenges. From the delivery route planning of a logistics company, to the management of electricity production according to the demand forecasts, to risk minimisation of a portfolio of financial assets. In order to solve these problems we have to make some model assumptions and formulate optimisation problems.

In general, an optimisation problem has the form

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, m \\ & && f_i(\mathbf{x}) = 0 \quad i = m + 1, \dots, k. \end{aligned} \tag{P2.1}$$

In this problem $\mathbf{x}$ is called the *decision vector* or *optimisation variable*. Function $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ is the *objective* function and $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, k$, are the *constraint* functions. The solution of the problem is a decision vector $\mathbf{x}^*$ which satisfies the following conditions:

C1. $f_i(\mathbf{x}^*) \leq 0$, for $i = 1, \dots, m$, and $f_i(\mathbf{x}^*) = 0$, for $i = m + 1, \dots, k$,

C2. for all $\mathbf{z} \in \mathbb{R}^n$ that satisfy $f_i(\mathbf{z}) \leq 0$, for $i = 1, \dots, m$, and $f_i(\mathbf{z}) = 0$, for $i = m + 1, \dots, k$, we have $f_0(\mathbf{z}) \geq f_0(\mathbf{x}^*)$

Optimisation problems can be grouped into families of problems. This categorisation depends on the special characteristics of the objective function and the constraint functions. For our study we are interested in convex optimisation problems.

The aim of this chapter is to introduce the main definitions and theory of convex optimisation. Section 2.1 motivates the study of this topic. Section 2.2 introduces the duality theory, which plays a very important role in the theory of optimisation. Finally we conclude this chapter with Section 2.3 where we give some examples from the literature. Most of the theory discussed in this chapter can be found in [6] and references therein.

2.1 Convex optimisation

Before we define what a convex optimisation problem is, we give some definitions important to the subject. We first start with the definitions of affine and convex sets.

Definition 2.1.1. *A set $S_{af} \in \mathbb{R}^n$ is affine if the line through any two distinct points in S_{af} lies in S_{af} , i.e. for all $\mathbf{x}_1, \mathbf{x}_2 \in S_{af}$ and $\theta \in \mathbb{R}$, we have*

$$\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2 \in S_{af}. \quad (2.1)$$

We can easily extend (2.1) to an affine combination for $\mathbf{x}_i$, $i = 1, \dots, m$. Since $\mathbf{x}_1 \in S_{af}$, we can define $\mathbf{x}_{1,1}, \mathbf{x}_{1,2} \in S_{af}$ and $\theta_2 \in \mathbb{R}$ such that $\mathbf{x}_1 = \theta_2 \mathbf{x}_{1,1} + (1 - \theta_2) \mathbf{x}_{1,2}$. Then we have $\theta(\theta_2 \mathbf{x}_{1,1} + (1 - \theta_2) \mathbf{x}_{1,2}) + (1 - \theta) \mathbf{x}_2 = \theta \theta_2 \mathbf{x}_{1,1} + \theta(1 - \theta_2) \mathbf{x}_{1,2} + (1 - \theta) \mathbf{x}_2 \in S_{af}$ with $\theta \theta_2 + \theta(1 - \theta_2) + (1 - \theta) = 1$. Hence, by induction, Definition 2.1.1 can be extended to any affine combination of vectors.

Definition 2.1.2. *A set S_{con} is convex if the line segment between any two points in S_{con} lies in S_{con} , i.e. for all $\mathbf{x}_1, \mathbf{x}_2 \in S_{con}$ and any $\theta \in \mathbb{I}$, we have*

$$\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2 \in S_{con}. \quad (2.2)$$

Again, we can extend (2.2) to a convex combination of $\mathbf{x}_i, i = 1, \dots, m$. This can be deduced in a similar manner as to before. From the above two definitions, it can also be deduced that every affine set is convex since it contains the whole line between any two points. However, the converse is not true.

Definition 2.1.3. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if $\text{dom } f$ is a convex (concave) set and if for all $\mathbf{x}, \mathbf{y} \in \text{dom } f$, and $\theta \in \mathbb{I}$, we have

$$f(\theta\mathbf{x} + (1 - \theta)\mathbf{y}) \leq (\geq) \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}). \quad (2.3)$$

Another way to identify a convex function is if the function satisfies first and second order conditions. For the *first order* condition we have to assume that the function f is differentiable. Then a function f is convex if and only if $\text{dom } f$ is convex and

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad (2.4)$$

holds for all $\mathbf{x}, \mathbf{y} \in \text{dom } f$. For the *second order* condition we have to assume that the function f is twice differentiable. Then a function f is convex if and only if $\text{dom } f$ is convex and

$$\nabla^2 f(\mathbf{x}) \succeq 0, \quad (2.5)$$

i.e. its Hessian is positive semidefinite for all $\mathbf{x} \in \text{dom } f$. For a concave function we require $\leq$ and $\preceq$ for (2.4) and (2.5) respectively. In general, we can define a concave function as $-f$ where f is a convex function. Adjusting (2.3)-(2.5) for $-f$ we obtain the required form for concave functions.

Now that we have given the necessary definitions for a convex set (Definition 2.1.2) and a

convex function (Definition 2.1.3), we can continue with the definition of a convex optimisation problem. A *convex optimisation problem* is an optimisation problem consisting only of convex functions, both in the objective and in the constraints, i.e. f_i for $i = 0, \dots, m$ are all convex. Also, all equality constraints are linear and hence also convex. The problem description in standard form is

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, m \\ & && \mathbf{A}\mathbf{x} + \mathbf{a} = 0 \end{aligned} \tag{P2.2}$$

where, $\mathbf{A} \in \mathbb{R}^{l \times n}$ and $\mathbf{a} \in \mathbb{R}^l$. In the same manner as in (P2.1) the optimal solution of the problem $\mathbf{x}^*$ has to satisfy the two conditions C1 and C2.

Convexity is the main reason why a problem like (P2.2) is attractive. The convexity of the objective function guaranties a single global optimal point. Especially for the case of an unconstrained problem, i.e. $m = 0$ and $l = 0$, we can easily find a numerical method that will provide $\mathbf{x}^*$. The solution of the problem becomes more complicated with the addition of constraints. Because the constraints in (P2.2) are convex there are efficient algorithms that still allow us to solve the problem. These is mainly due to the convexity of the **dom** f_i , $i = 1, \dots, m$, which preserves the convexity of the problem's domain

$$\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom} f_i.$$

When the convexity of (P2.2) is combined with duality theory it can provide strong results that the algorithms can take advantage of. We will discuss more about duality theory in the next section. On the other hand, if convexity in the constraints did not hold we would have to resolve to global optimisation techniques, where finding $\mathbf{x}^*$ is not guaranteed.

2.2 Duality

Duality is very important in the theory of optimisation. The most common application of duality uses the Lagrangian function. The *Lagrangian* function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{k-m} \rightarrow \mathbb{R}$ associated with (P2.1), called the *primal problem*, is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=m+1}^k \nu_i f_i(\mathbf{x}). \quad (2.6)$$

The domain of L is $\mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^{k-m}$. Vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are called the *Lagrangian multipliers*. From (2.6) we can define the *Lagrangian dual function* (dual function) as

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=m+1}^k \nu_i f_i(\mathbf{x}) \right). \quad (2.7)$$

An important characteristic of $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is that it is linear with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$. This is true even if the original problem (P2.1) is not convex. Hence, it is possible, through the use of the Lagrangian function, to reformulate a non convex problem into a convex one.

The dual function is always an underestimator of the original problem, i.e. for any $\boldsymbol{\lambda} \geq 0$ and $\boldsymbol{\nu} \in \mathbb{R}^{k-m}$ we have $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\mathbf{x}^*)$. Thus, we can formulate the *dual problem* as follows

$$\begin{aligned} & \underset{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^{k-m}}{\text{maximize}} && g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ & \text{subject to} && \boldsymbol{\lambda} \geq 0. \end{aligned} \quad (\text{P2.3})$$

The solution of the dual problem are the decision vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ where $\boldsymbol{\lambda}^* \geq 0$ and for all $\boldsymbol{\lambda} \geq 0$ and $\boldsymbol{\nu} \in \mathbb{R}^{k-m}$ we have $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$. These two conditions are the equivalent to C1 and C2 for (P2.1) respectively.

The difference between $f_0(\mathbf{x}^*)$ and $g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is called the *duality gap* and is the minimum distance between the solutions of (P2.1) and (P2.3). If the duality gap is strictly greater than zero we say that we have *weak duality*. If, on the other hand, we have strict equality,

then we achieve *strong duality* and we say that strong duality holds. Hence, in the case of strong duality, by solving (P2.3), we obtain the solution of the original problem (P2.1), i.e. $f_0(\mathbf{x}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$.

In general strong duality is not achievable, but there are methods that can be used to identify problems where strong duality holds. Here we will look into two methods.

For most convex problems (P2.2) strong duality holds, but this is not always the case. In order to verify whether strong duality is achievable we can use Slater's constraint qualification. Here we give the more general version. *Slater's constraint qualification* states that if for a convex problem of the form

$$\begin{aligned}
 & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} && f_0(\mathbf{x}) \\
 & \text{subject to} && f_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, m \\
 & && \mathbf{Ax} + \mathbf{a} \leq 0 \\
 & && \mathbf{Bx} + \mathbf{b} = 0
 \end{aligned} \tag{P2.4}$$

where $\mathbf{A} \in \mathbb{R}^{l \times n}$, $\mathbf{a} \in \mathbb{R}^l$, $\mathbf{B} \in \mathbb{R}^{p \times n}$ and $\mathbf{b} \in \mathbb{R}^p$, there exist $\mathbf{x} \in \mathbf{relint} \mathcal{D}$ (see Definition A.1.2) of (P2.4), which satisfies

$$\begin{aligned}
 & f_i(\mathbf{x}) < 0, \quad i = 1, \dots, m, \\
 & \mathbf{Ax} + \mathbf{a} \leq 0, \\
 & \mathbf{Bx} + \mathbf{b} = 0,
 \end{aligned} \tag{2.8}$$

then strong duality holds.

Another way to see whether strong duality for a convex problem holds is through the use of the *Karush-Kuhn-Tucker* (KKT) conditions. If for (P2.2) we have $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\nu}}$ which satisfy the KKT conditions

$$\begin{aligned}
f_i(\tilde{\mathbf{x}}) &\leq 0, & i = 1, \dots, m \\
\mathbf{A}\tilde{\mathbf{x}} + \mathbf{a} &= 0, & i = 1, \dots, l, \\
\tilde{\lambda} &\geq 0, \\
\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) &= 0, & i = 1, \dots, m, \\
\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^m \lambda_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=m+1}^k v_i \nabla (\mathbf{A}_i \tilde{\mathbf{x}} + \mathbf{a}) &= 0,
\end{aligned} \tag{2.9}$$

then $\tilde{\mathbf{x}}$ and $(\tilde{\lambda}, \tilde{\nu})$ are the solutions of (P2.2) and (P2.3), with zero duality gap. Here $\mathbf{A}_i$ denotes the i^{th} row of matrix $\mathbf{A}$.

2.3 Examples

In this section we will give some examples of optimisation problems that are convex or can be reformulated into a convex optimisation problem. Some of the ideas and concepts introduced in these examples will be used in our work later on.

Let's consider a very common example from financial portfolio optimisation. In this problem we seek to minimise the risk with respect to VaR for an asset portfolio [49]. We can write the problem as

$$\begin{aligned}
&\underset{\gamma \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \gamma \\
&\text{subject to} && P(\gamma \geq -\mathbf{w}^T \mathbf{r}) \geq 1 - \epsilon.
\end{aligned} \tag{E2.3.1}$$

Here, the decision vector $\mathbf{w}$ represents the amount invested in each asset, i.e. w_i is the amount invested in the i^{th} asset. Also, $\mathbf{r}$ denotes the random vector of asset returns, with r_i denoting the return of the i^{th} asset, and level ϵ , which is a given constant, gives $(1 - \epsilon)$ percentile of the distribution¹. Because of the probability $P(\cdot)$ in the constraints and $\mathbf{r}$ being a random vector, (E2.3.1) constitutes a *chance-constrained stochastic program*.

Chance-constrained stochastic programs are in general difficult to solve. They tend to have non-convex and sometimes disconnected feasible sets [49]. Also, they require precise

¹Common values for ϵ are 0.1%, 1% and 5%.

knowledge of the distribution of random vector $\mathbf{r}$. This, in practice is unrealistic. Finally, the convexity of VaR is highly dependent on the underlying distribution of $\mathbf{r}$ [10].

The most common way of solving (E2.3.1) is by making assumptions about the underlying distribution of $\mathbf{r}$. Taking into account the complexities that (E2.3.1) demonstrates, practitioners usually assume that asset prices are log-normally distributed. Thus, the returns of the assets are normally distributed. This gives an analytic evaluation of VaR,

$$\text{VaR}_\epsilon(\mathbf{w}) = -\boldsymbol{\mu}^T \mathbf{w} - \Phi^{-1}(\epsilon) \sqrt{\mathbf{w}^T \boldsymbol{\Sigma}^{-1} \mathbf{w}}, \quad (2.10)$$

which is a convex function with respect to $\mathbf{w}$. Here, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and the covariance matrix of $\mathbf{r}$. The problem with this approach is that assuming that the returns are normally distributed is a weak assumption [18, 19, 45]. At the same time if $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are assumed or estimated, hence the possibility of a systematic or estimation error is very high. A lot of research trying to address these problems. We will particularly focus in the cases where assumptions about the distribution of the random vector $\mathbf{r}$ are omitted and attention is on the uncertainty set related to $\mathbf{r}$, i.e. $\text{dom } \mathbf{r}$.

The *robust* approach is to consider an uncertainty set for $\mathbf{r}$ and then optimise against all possible scenarios with respect to the uncertainty set. In this example the optimisation problem is written as

$$\begin{aligned} & \underset{\gamma \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \gamma \\ & \text{subject to} && \gamma \geq -\mathbf{w}^T \mathbf{r} \quad \forall \mathbf{r} \in S_r. \end{aligned} \quad (\text{E2.3.2})$$

In order to evaluate (E2.3.2) we have to make assumptions about the uncertainty set $S_r \subseteq \mathbb{R}^n$. The selection of S_r has to be representative of the possible realisations of $\mathbf{r}$. One common uncertainty set used in the literature, which utilises nicely the available information from $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, is the *ellipsoidal* uncertainty set,

$$S_r(\delta) = \left\{ \mathbf{r} \in \mathbb{R}^n : (\mathbf{r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \leq \delta^2 \right\}. \quad (2.11)$$

Then (E2.3.2) can be written as

$$\begin{aligned}
 & \underset{\gamma \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \gamma \\
 & \text{subject to} && \sup_{\mathbf{r} \in \mathbb{R}^n} \{ \gamma \geq -\mathbf{w}^T \mathbf{r} \} \\
 & && (\mathbf{r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \leq \delta^2.
 \end{aligned} \tag{E2.3.3}$$

In general problems with sup in the constraints are very difficult to solve, but in this example we can use the nice properties of ellipsoidal uncertainty. It can be shown that by applying duality theory on sup we can rewrite (E2.3.3) as a convex optimisation problem,

$$\begin{aligned}
 & \underset{\gamma \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \gamma \\
 & \text{subject to} && \gamma \geq -\boldsymbol{\mu}^T \mathbf{w} + \delta \|\boldsymbol{\Sigma}^{1/2} \mathbf{w}\|_2.
 \end{aligned} \tag{E2.3.4}$$

where $\|\cdot\|_2$ is the Euclidean norm [3, 4]. In practice we can replace $\boldsymbol{\mu}$ with any vector and $\boldsymbol{\Sigma}$ with any positive semidefinite matrix.

Another way to approach (E2.3.1) is by taking into account a *worst case scenario*. One example is the work by El Ghaoui et al.[15]. In this case study full knowledge of the first and second moments is assumed, i.e. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Furthermore, $\boldsymbol{\Sigma} \succ 0$. Then, the constraints in (E2.3.3) are replaced with

$$\sup_{P \in \mathcal{P}_r} P(\gamma \geq -\mathbf{w}^T \mathbf{r}) \geq 1 - \epsilon, \tag{2.12}$$

where $\mathcal{P}_r$ is the set of all probability distribution with first and second moments equal to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. With the use of duality theory, El Ghaoui et al. [15] show that (2.12) can be written as

$$\gamma \geq -\boldsymbol{\mu}^T \mathbf{w} + \kappa(\epsilon) \|\boldsymbol{\Sigma}^{1/2} \mathbf{w}\|_2, \tag{2.13}$$

where $\kappa(\epsilon) = \sqrt{(1 - \epsilon)/\epsilon}$. Comparing (2.13) with the constraints in (E2.3.4) we can see that if $\delta = \kappa(\epsilon)$, then the two problems become equivalent.

Chapter 3

Copula Review

Copulas were introduced by Sklar in 1959 [42]. They belong to a special family of probabilistic metric spaces that have the same properties as probability distributions [39]. Thus, copulas provide an alternative way of looking into probability distributions. As the name suggests, copulas, *couple* marginal distribution functions. They are a probabilistic measure of distance or dependence between marginal distributions, the univariate distributions of the individual random variables, as opposed to probability distributions that directly describe the dependence between random variable.

Copulas, in view of their theoretical foundation as measures of distance between marginal distributions, inherit many interesting properties. For example, they are scale-free measures of dependence, which are invariant under monotonic transformations [29, 43]. In addition, the fact that they can measure distance between univariate margins, makes them very attractive in modelling and find their way to a variety of applications (see for example [13, 18, 19, 35, 45]). Finally, in [38, 41, 43], through a series of arguments, it is shown that there are operations on distribution functions that do not correspond directly to operations on random variables. However, these operations can be achieved through copulas. This instates copulas as an interesting topic on their own right.

We are interested in copulas from a modelling perspective, because we would like to take advantage of the modelling flexibility that they provide. In particular, we will rigorously examine a special class of copulas called Archimedian. In Section 3.1 we introduce the relevant theory and formally define copulas. In Section 3.2 we present some measures of dependence associated with copulas, which are used in subsequent chapters. Then, in Section 3.3 we describe some of the methods for creating copulas, and in Section 3.4 we go through main estimation techniques. Finally, we conclude this chapter with Section 3.5 by demonstrating some of copulas' properties and applications.

3.1 Copulas

Before we give a rigorous definition of copulas, we firstly supply some of the background theory. This theory provides an in depth understanding of copulas in the context of probabilistic metric spaces. In the sequel the copulas are defined and the theorems that associate them with probability distributions are stated. The background literature in this section can be found in [29, 38, 39, 43, 44] and the references therein. Here, we omit proofs of the theorems but we indicate the references where the proofs can be found.

The first few definitions focus on the monotonic properties of a function. We begin with two definitions regarding an n -dimensional subset of $\mathbb{R}^n$, and its associated vertices.

Definition 3.1.1. *Let n be a positive integer. An n -interval is the Cartesian product of n real intervals, and an n -box the Cartesian product of n closed intervals. If $\mathbf{J}$ is the n -interval $J_1 \times \dots \times J_m \times \dots \times J_n$, where for $m = 1, \dots, n$, and the interval J_m has end points a_m and b_m then the vertices of $\mathbf{J}$ are the points $\mathbf{z} = (z_1, \dots, z_n)$ such that each z_m is equal to either a_m or b_m .*

Definition 3.1.2. *Let $\mathbf{B}$ be the n -box $[\mathbf{a}, \mathbf{b}]$, where $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$. Let $\mathbf{z} = (z_1, \dots, z_n)$ be a vertex of $\mathbf{B}$. If the vertices of $\mathbf{B}$ are all distinct (which is equivalent to saying $\mathbf{a} < \mathbf{b}$),*

then

$$\text{sgn}_{\mathbf{B}}(\mathbf{z}) = \begin{cases} 1, & \text{if } z_m = a_m \text{ for an even number of } m\text{'s} \\ -1, & \text{if } z_m = a_m \text{ for an odd number of } m\text{'s} \end{cases} \quad (3.1)$$

If the vertices are not distinct, then $\text{sgn}_n(\mathbf{Z}) = 0$.

Using the above two definitions we can define the volume of an n -box under a function F and the properties that F has to satisfy in order to be considered a distribution function.

Definition 3.1.3. Let S_n be nonempty subset of $\mathbb{R}^n$, and let $F : S_n \rightarrow \mathbb{R}$ (F is also called an n -place function). If $\mathbf{B}$ is an n -box whose vertices are all in $\mathbf{dom} F$, then the F -volume of $\mathbf{B}$ is the sum

$$V_F(\mathbf{B}) = \sum_{\mathbf{z} \in \mathcal{Z}_{\mathbf{B}}} \text{sgn}_{\mathbf{B}}(\mathbf{z})F(\mathbf{z}) \quad (3.2)$$

where $\mathcal{Z}_{\mathbf{B}}$ is the set of all vertices of $\mathbf{B}$.

For example, consider the following special case. Let $\mathbf{B}$ be the 2-box $[a_1, b_1] \times [a_2, b_2]$. Then, the set of vertices is $\mathcal{Z}_{\mathbf{B}} = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$. Hence, the F -volume is

$$\begin{aligned} V_F(\mathbf{B}) &= \sum_{\mathbf{z} \in \mathcal{Z}_{\mathbf{B}}} \text{sgn}_{\mathbf{B}}(\mathbf{z})F(\mathbf{z}) \\ &= F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1) + F(a_2, b_2) \end{aligned}$$

The two definitions and the lemma that follows are vital in establishing fundamental properties of a distribution function.

Definition 3.1.4. An n -place real function F is n -increasing if $V_F(\mathbf{B}) \geq 0$ for all n -boxes $\mathbf{B}$ the vertices of which lie in $\mathbf{dom} F$.

Definition 3.1.5. An n -place function F is grounded if $F(x_1, \dots, x_m, \dots, x_n) = 0$ where $x_m = a_m$, the endpoint of $\mathbf{dom} F$, for $m = 1, 2, \dots, n$.

Lemma 3.1.6. Let an n -place function F be n -increasing and grounded with $\mathbf{dom} F = S_n \subseteq \mathbb{R}^n$. Then F is nondecreasing in each argument.

The proof for the bivariate case of this lemma can be found in Nelsen [29] and for the multivariate in Schweizer and Sklar [39]. This lemma extends the monotonicity property of the multivariate function to the monotonicity of its margins.

Now we are in a position to rigorously define distribution functions.

Definition 3.1.7. *A distribution function is a function F with domain $\mathbb{R}$ such that:*

1. F is nondecreasing
2. $F(-\infty) = 0$ and $F(\infty) = 1$
3. F is right-contained with left limits

Definition 3.1.8. *A joint n -place distribution function is a function F with domain $\mathbb{R}^n$ such that*

1. F is n -increasing
2. F is grounded
3. $F(\infty) = 1$
4. $F(\infty, \dots, \infty, x_m, \infty, \dots, \infty) = F_m(x_m)$,

where $m = 1, 2, \dots, n$.

We now continue with copulas. For the sake of completeness, since copulas are a special case of a family of multivariate functions called *subcopulas*, we will start by giving the definition of these functions.

Definition 3.1.9. *An n -dimensional subcopula (also called n -subcopula) is a function C' with the following properties:*

1. $\text{dom } C' = S_n \subset \mathbb{I}^n$

2. C' is grounded and n -increasing
3. $C'(1, \dots, 1, u_m, 1, \dots, 1) = u_m$ for all $u_m \in \mathbb{I}$ where $m = 1, 2, \dots, n$.

Copulas are a special case of sub-copulas where $S_n \equiv \mathbb{I}^n$, i.e., their domain is the unit hypercube with n dimensions.

Definition 3.1.10. *An n -dimensional copula (n -copula) is an n -subcopula C whose domain is $\mathbb{I}^n$. Equivalently, a copula is a function $C : \mathbb{I}^n \rightarrow \mathbb{I}$ with the following properties*

1. C is grounded
2. $C(1, \dots, 1, u_m < 1, \dots, 1) = u_m$ for all $u_m \in \mathbb{I}_m$ where $m = 1, 2, \dots, n$
3. C is n -increasing

The following lemma gives a rigorous connection between the copula and the sub-copula functions. The proofs for the bivariate and the multivariate cases can be found in Nelsen [29] and Sklar [44] respectively.

Lemma 3.1.11. *Let C' be an n -subcopula. Then there exists a n -copula C such that $C'(u) = C(u)$ for all $u \in \text{dom } C'$.*

Finally, consider the following definition that will be of use in the sequel.

Definition 3.1.12. *The support of a copula is the complement of the union of all open subsets of $\mathbb{I}^n$ with V_C zero (see Definition 3.1.3).*

We continue with what is probably the most important theorem. *Sklar's* theorem is probably the most fundamental connection between copulas and statistics. This theorem, together with the corollary that follows, give the structure which enables us to easily construct new multivariate models. Sklar's theorem was introduced by Sklar in 1959 [42] where also the proof of its bivariate case can be found. For the multivariate case see Schweizer and Sklar [39] (the proofs of the corollary are contained in references therein).

Theorem 3.1.13 (Sklar's Theorem). *Let F be an n -place distribution function with margins $F_1, \dots, F_n$. Then there exists an n -copula C such that, for all $x \in \mathbb{R}^n$,*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (3.3)$$

Furthermore, if $F_1, \dots, F_n$ are continuous, then C is unique; otherwise C is unique on $\text{ran } F_1 \times \dots \times \text{ran } F_n$.

Corollary 3.1.14. *Let F be an n -place distribution function with margins $F_1, \dots, F_n$, and let C be an n -copula. Then, for any $\mathbf{u} \in \mathbb{I}^n$,*

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \quad (3.4)$$

where $F_1^{-1}, \dots, F_n^{-1}$ are the quasi-inverses of the marginals.

Knowing the relation that exists between the joint distribution function F and a copula C we can also define the joint density function of copulas.

Definition 3.1.15. *The copula density of an n -copula C is the function $c : \mathbb{I}^n \rightarrow [0, \infty)$ such that*

$$c(u_1, \dots, u_n) \equiv \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n} = \frac{f(x_1, \dots, x_n)}{\prod_{i=1}^n f_i(x_i)}. \quad (3.5)$$

Another interesting connection between copulas and random variables is the case where random variables are independent.

Theorem 3.1.16. *Let $\mathbf{x}$ be a random vector in $\mathbb{R}^n$ with n -copula C . Then $\mathbf{x}$ has independent entries if and only if $C = u_1 \cdot \dots \cdot u_n$, also called the product copula and is denoted by Π .*

Finally, global bounds for copulas are given. In the case where we do not know or cannot construct an n -copula or a multivariate distribution function to represent our random variables, we can use these bounds to construct the best and the worst-case scenarios between

the random variables. This description applies mostly when we want to model the dependence between the different assets and we want to know at least the worst and best possible outcomes.

Theorem 3.1.17. *Let C' be an n -subcopula. Then for every $\mathbf{u} = (u_1, u_2, \dots, u_n)$ in $\text{Dom } C'$*

$$W(\mathbf{u}) \leq C' \leq M(\mathbf{u}) \quad (3.6)$$

where $W(\mathbf{u}) = \max(u_1 + u_2 + \dots + u_n - n + 1, 0)$ and $M(\mathbf{u}) = \min(u_1, u_2, \dots, u_n)$.

W and M are also called *Fréchet-Hoeffding* lower and upper bounds. Theorem 3.1.17 applies to n -copulas as well since by Definition 3.1.10 an n -copula is a special case of n -subcopula. W and M are copulas themselves. W is copula for $n = 2$ and M for $n \geq 2$. The proof of Theorem 3.1.17 can be found in Nelsen [29] for the bivariate case and in Schweizer and Sklar [39] for the multivariate case.

3.2 Measures of Dependence

Classical measures of dependence make use of joint distribution functions in their structure. Using Sklar's Theorem 3.1.13, we can replace the joint distribution functions with copulas. As we will see in this section, by substituting the joint distribution functions with copulas, it is possible to fully remove the information from the random variable. This will reinstate the measure as non-parametric [40]. Also, the measure will inherit some of copulas' properties that we mentioned above. Probably, the most important property is the invariance under monotonic transformations. This makes the measure a rank statistic, a measure of comovements between the random variables, which can be considered as a more *robust* measure of dependence [8].

In this section we introduce some of the measures that are associated with copulas. For comparison purposes and for the sake of completeness, we provide both the copula and

distribution formulation of each measure. All the measures of dependence are defined for random variables $X_1, X_2 \in \mathbb{R}$ and $U_1, U_2 \in \mathbb{I}$.

The most widely used measure of dependence is the *linear correlation*, ρ , defined as,

$$\begin{aligned}\rho(X_1, X_2) &= \frac{1}{\sigma_{X_1}\sigma_{X_2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x_1, x_2) - F_1(x_1)F_2(x_2)] dx_1 dx_2 \\ &= \frac{1}{\sigma_{X_1}\sigma_{X_2}} \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2] dF_1^{(-1)}(u_1) dF_2^{(-1)}(u_2),\end{aligned}\tag{3.7}$$

where σ_{X_i} is the standard deviation of variable X_i . ρ , is always dependent on X_1 and X_2 because of σ_{X_1} and σ_{X_2} , and its dependence cannot be removed by the copula substitution. This makes ρ a strictly parametric measure of dependence, which, as the name suggests, can only measure linear dependencies. Thus, it is sensitive to inconsistencies that may occur in sample data.

As opposed to ρ , two classical measures of dependence that do not suffer from the same deficiencies are *Spearman's* ρ_s and *Kendal's* τ . They are defined as ,

$$\begin{aligned}\rho_s(X_1, X_2) &= 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x_1, x_2) - F_1(x_1)F_2(x_2)] dF(x_1) dF(x_2) \\ &= 12 \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2] du_1 du_2,\end{aligned}\tag{3.8}$$

$$\begin{aligned}\tau(X_1, X_2) &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x_1, x_2) dF(x_1, x_2) - 1 \\ &= 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.\end{aligned}\tag{3.9}$$

Their copula version does not have any dependence on X_1 and X_2 , which renders them rank statistics.

We now move on to some measures of dependence that were particularly created with copulas in mind. Schweizer and Wolff [40] argue that since random variables X_1, X_2 are independent if and only if $C(u_1, u_2) = u_1 u_2$ (Theorem 3.1.16), any suitable normalised measure of distance between the surface $z = C(u_1, u_2)$ and $z = u_1 u_2$ should yield a symmetric non-

parametric measure of dependence. Hence, they take the L_p norm and demonstrate the cases where $p = 1, 2$ and ∞ . These measures are

$$\begin{aligned}\omega(U_1, U_2) &= 12 \int_0^1 \int_0^1 |C(u_1, u_2) - u_1 u_2| du_1 du_2 \\ &= 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F(x_1, x_2) - F_1(x_1)F_2(x_2)| dF_1(x_1)dF_2(x_2),\end{aligned}\quad (3.10)$$

$$\begin{aligned}\gamma(U_1, U_2) &= \left(90 \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2]^2 du_1 du_2 \right)^{\frac{1}{2}} \\ &= \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x_1, x_2) - F_1(x_1)F_2(x_2)]^2 dF_1(x_1)dF_2(x_2) \right)^{\frac{1}{2}}, \\ \kappa(U_1, U_2) &= 4 \sup_{u_1, u_2 \in \mathbb{I}} |C(u_1, u_2) - u_1 u_2| \\ &= 4 \sup_{x_1, x_2 \in \mathbb{R}} |F(x_1, x_2) - F_1(x_1)F_2(x_2)|.\end{aligned}\quad (3.11)$$

Other interesting measures of dependence are the ones that calculate *tail dependence*. The tail dependence of two random variables explains how the two variables behave compared to each other at the extreme ends of their distributions. These measures are very useful in finance, because they quantify how two assets may behave, when compared to each other, in periods of large gains (prosperity) or losses (crisis) [13, 45]. The first pair of these measures is the *tail dependence coefficients*. The *upper* tail dependence coefficient is defined as

$$\begin{aligned}\lambda_U(X_1, X_2) &= \lim_{u \rightarrow 1^-} P[X_2 > F_2^{(-1)}(u) \mid X_1 > F_1^{(-1)}(u)] \\ &= \lim_{u \rightarrow 1^-} P[X_1 > F_1^{(-1)}(u) \mid X_2 > F_2^{(-1)}(u)]\end{aligned}$$

and the *lower* tail dependence coefficient as

$$\begin{aligned}\lambda_L(X_1, X_2) &= \lim_{u \rightarrow 0^+} P[X_2 \leq F_2^{(-1)}(u) \mid X_1 \leq F_1^{(-1)}(u)] \\ &= \lim_{u \rightarrow 0^+} P[X_1 \leq F_1^{(-1)}(u) \mid X_2 \leq F_2^{(-1)}(u)].\end{aligned}$$

Upper and lower tail dependence can be expressed strictly in copula terms by

$$\lambda_U(C) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u} \quad \text{and} \quad \lambda_L(C) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}. \quad (3.12)$$

Lower and upper *quantile dependence* is another way of measuring tail dependence. Nelsen [29] refers to them as *left tail decreasing* and *right tail increasing* respectively. He also discusses a series of theorems that describe their relation with copulas (the theorems and other measures of dependence are discussed in [29]). The *upper* quantile dependence is defined as

$$\begin{aligned} \nu_U(u; X_1, X_2) &= P[X_2 > F_2^{(-1)}(u) \mid X_1 > F_1^{(-1)}(u)] \\ &= P[X_1 > F_1^{(-1)}(u) \mid X_2 > F_2^{(-1)}(u)], \end{aligned}$$

and the *lower* quantile dependence as

$$\begin{aligned} \nu_L(u; X_1, X_2) &= \text{Prob}[X_2 \leq F_2^{(-1)}(u) \mid X_1 \leq F_1^{(-1)}(u)] \\ &= \text{Prob}[X_1 \leq F_1^{(-1)}(u) \mid X_2 \leq F_2^{(-1)}(u)]. \end{aligned}$$

Their copula only version is

$$\nu_U(u) = \frac{1 - 2u + C(u, u)}{1 - u} \quad \text{and} \quad \nu_L(u) = \frac{C(u, u)}{u}. \quad (3.13)$$

Renyi [32] gives a series of axioms that define a coherent measure of risk between two random variables. Some of the above measures satisfy some of this conditions [2, 40, 46]. The linear correlation ρ is not a coherent measure of risk [2]. This is one more reason why a normal distribution is not appropriate to describe multivariate dependencies, because the dependence in the distribution is solely defined by the linear correlation.

Wolff [47] extends some of this measures to n-dimensions. He constructs the multivariate versions of Spearman's ρ_S (3.8), σ (3.10) and κ (3.11). He also takes Renyi's axioms [2, 8, 32] and constructs their multivariate versions. Thus, he associates them with the axioms that

they satisfy. Trying to describe the dependence between more than two variables with a single parameter can be inappropriate. By trying to provide an all in one description a lot of information related to individual pairs is neglected. Thus, we omit the multivariate versions of ρ_S , σ and κ , but the interested reader can refer to Wolff [47].

The formulas of the above measures for sample data can be found in Smillie [45].

3.3 Methods for Constructing Copulas

In this section we review some of the methods for constructing copulas. Some, make use of Theorem 3.1.13, hence we directly obtain their joint distribution equivalent. Others, come from copula theory and from empirical formulations. For an extended discussion about methods for constructing copulas see Nelsen [29].

3.3.1 Inversion

Nelsen [29] discusses various methods of constructing copulas. The simplest is the inversion method. For a given n -place distribution function F , we use Theorem 3.1.13 in conjunction with Definition 3.1.8 and Definition 3.1.10 to get

$$F(\infty, \dots, \infty, x_m, \infty, \dots, \infty) = F_m(x_m) = u_m = C(1, \dots, 1, u_m, 1, \dots, 1)$$

for $m = 1, 2, \dots, n$. Using the above, we can define x_m in terms of u_m by $x_m = F^{(-1)}(u_m)$, if the inverse exist, and replace it in F to determine C .

Many families of copulas are a result of this inversion. The most common, typically used in finance, is the *Gaussian Copula* and the *t-copula* [8, 9, 13, 35, 45].

Definition 3.3.1. Given a n -place standard multivariate normal distribution function Φ_n parameterised by a correlation matrix $\mathbf{R} \in [-1, 1]^{n \times n}$, the Gaussian copula is the function $C_{Ga} : \mathbb{I}^n \rightarrow \mathbb{I}$

such that

$$C_{Ga}(\mathbf{u}) = \Phi(\Phi_1^{-1}(u_1), \dots, \Phi_n^{-1}(u_n)).$$

It is important to note that in Definition 3.3.1, for C_{Ga} to be called Gaussian, all margins have to be *standard normal distributions*, i.e., it is important to have Φ as the multivariate distribution and Φ_i for margins. The relation between the margins and the random variables is $\Phi((x_i - \mu_i)/\sigma_i) = \Phi_i(z_i) = u_i$, where $Z_i \sim N(0, 1)$, for $i = 1, \dots, n$. Using Definition 3.1.15 we can also define the Gaussian copula density function [8, 45]

$$c_{Ga}(\mathbf{u}) = \frac{\phi(\Phi_1^{-1}(u_1), \dots, \Phi_n^{-1}(u_n))}{\phi(\Phi_1^{-1}(u_1)) \cdot \dots \cdot \phi(\Phi_n^{-1}(u_n))} = \frac{1}{|\mathbf{R}|^{1/2}} \exp \left[-\frac{1}{2} \zeta^T (\mathbf{R}^{-1} - \mathbf{I}_n) \zeta \right],$$

where $\zeta = (\Phi_1^{-1}(u_1), \dots, \Phi_n^{-1}(u_n))^T$ and $\mathbf{I}$ is the $n \times n$ identity matrix.

Figure 3.1: Gaussian with $\rho = 0.891$ and $\tau = 0.7$

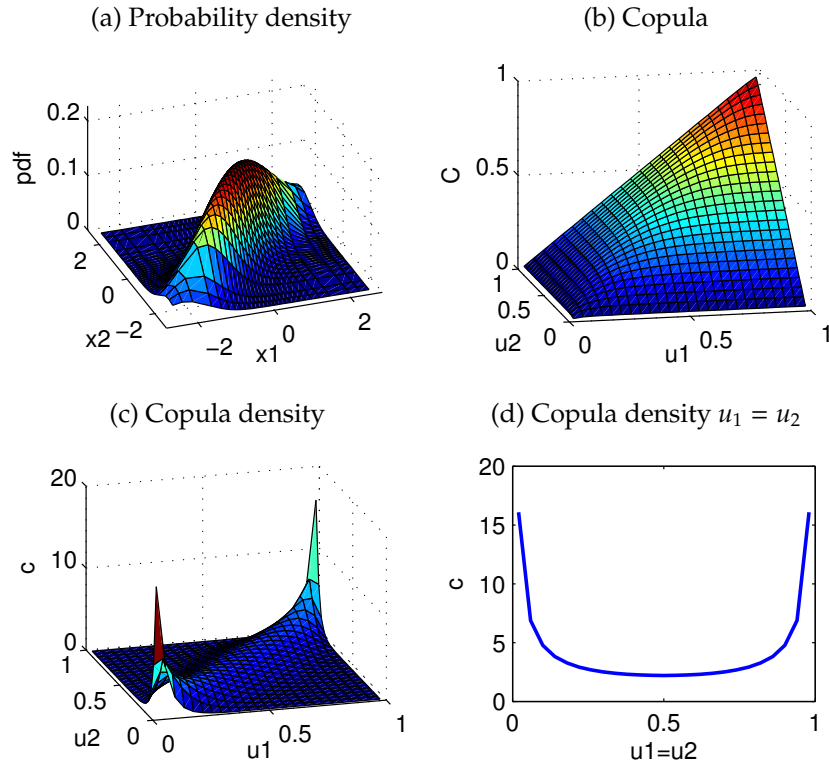


Figure 3.1a shows the probability density function of a Gaussian bivariate distribution where

$X_1, X_2 \sim N(0, 2)$ and $\rho(X_1, X_2) = 0.891$. Figure 3.1b to Figure 3.1d, show the copula related functions. By comparing Figure 3.1a with Figure 3.1c, we can see the difference between $c_{Ga}(\mathbf{u})$ and $\phi(\mathbf{x})$ plot. As in the case of high correlation, i.e. strong dependence, the mass of $\phi(\mathbf{x})$ is concentrated along the diagonal. In the case of $c_{Ga}(\mathbf{u})$ the strong dependence is reflected on the high picks of the tails of the diagonal. Also, Figure 3.1d illustrates nicely the symmetric dependence structure that the Gaussian copula implies. An example from finance, demonstrating this symmetry, is the probability of positive returns which is the same as the one for negative asset returns.

Definition 3.3.2. Given a n -place multivariate t -distribution function $T_{n,\nu}$, parameterised by a correlation matrix $\mathbf{R} \in [-1, 1]^{n \times n}$ and degrees of freedom $\nu \in \mathbb{R}_+$ and $\nu > 2$ ¹, the Student's t copula is the function $C_T : \mathbb{I}^n \rightarrow \mathbb{I}$ such that

$$C_T(\mathbf{u}) = T_\nu(T_{\nu,1}^{-1}(u_1), \dots, T_{\nu,n}^{-1}(u_n)),$$

where $T_\nu^{-1}(\cdot)$ denotes the inverse of the univariate t distribution function having ν degrees of freedom.

The relation between the margins and the random variables is given by $T_{\nu,i}((x_i - \mu_i)/\sigma_i) = u_i$ for $i = 1, \dots, n$. For the t -copula density function c_T , see [8, 45]. For calibration and simulation techniques for C_{Ga} and C_T , see [8, 45].

3.3.2 Shuffles

Shuffles is a general method of constructing copulas of a more theoretical nature, and belongs to a broader class of methods called *geometric* [29]. One example is the shuffles of Fréchet-Hoeffding upper bound M (see Definition 3.6), which essentially partitions the support of M

¹This assumption is required in order to guaranty the existence of the correlation matrix. If $\nu \leq 2$ the correlation matrix does not exist.

(see Definition 3.1.12) and shuffles it within $\text{dom } C$. What makes this method interesting is the following theorem for a 2-copula.

Theorem 3.3.3. *For any $\epsilon > 0$, there exists a shuffle of M , which we denote C_ϵ , such that*

$$\sup_{\mathbf{u} \in \mathbb{I}^2} |C_\epsilon(\mathbf{u}) - C(\mathbf{u})| < \epsilon. \quad (3.14)$$

Theorem 3.3.3 implies that it is possible to approximate any 2-copula by shuffles of M . This may be of interest, especially from an algorithmic or optimisation perspective, because by e.g., any given set of empirical data, C_ϵ can be used to approximate the empirical copula (see Subsection 3.3.4). Hence, we will have an analytic formula for the approximation. The proof of the theorem can be found in Nelsen [29] and references therein.

3.3.3 Archimedean Copulas

The family of *Archimedean* copulas were first introduced by Ling [24]. They are a unique in that their construction stems from probabilistic metric spaces theory. Rather than using Theorem 3.1.13, they are constructed directly from a function φ , known as the generator, which enables us to write the expression for the copula in closed form.

Definition 3.3.4. *Given a function $\varphi : \mathbb{I} \rightarrow [0, \infty)$ such that $\varphi(1) = 0$ and $\varphi(0) = \infty$ and having inverse $\varphi^{[-1]}$ completely monotone, an n -place **Archimedean** copula is a function $C_\varphi : \mathbb{I}^n \rightarrow \mathbb{I}$ such that*

$$C_\varphi(\mathbf{u}) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_n)).$$

Nelsen [29] gives a more rigorous definition regarding the characteristics of the generator φ , and its pseudo-inverse $\varphi^{[-1]}$. For the sake of completeness we also give this definition.

Definition 3.3.5. Let φ be a continuous, strictly decreasing function from $\mathbb{I}$ to $[0, \infty]$ such that $\varphi(1) = 0$. The pseudo-inverse of φ is the function $\varphi^{[-1]}$ with **dom** $\varphi^{[-1]} = [0, \infty]$ and **ran** $\varphi^{[-1]} = \mathbb{I}$ given by

$$\varphi^{[-1]} = \begin{cases} \varphi^{-1} & \text{for } 0 \leq t \leq \varphi(0) \\ 0 & \text{for } \varphi(0) < t < \infty \end{cases}.$$

Note that $\varphi^{[-1]}$ is continuous and non-increasing on $0, \infty$, and strictly decreasing on $0, \varphi(0)$. Furthermore $\varphi^{[-1]}(\varphi(u)) = u$ on $\mathbb{I}$ and

$$\varphi(\varphi^{[-1]}(t)) = \begin{cases} t & \text{for } 0 \leq t \leq \varphi(0) \\ \varphi(0) & \text{for } \varphi(0) \leq t \leq \infty \end{cases} = \min(t, \varphi(0)).$$

Finally, if $\varphi(0) = \infty$, then $\varphi^{[-1]} = \varphi^{-1}$.

For simplicity, we will only use the notation of inverse, φ^{-1} . We proceed by giving a theorem which describes some of the properties that apply to all Archimedean copulas. We give the proof for the multivariate case of Theorem 3.3.6 in Appendix B.1.

Theorem 3.3.6. Let C be an Archimedean n -copula with generator φ , and let $\mathbf{u} \in \mathbb{I}^n$. Then

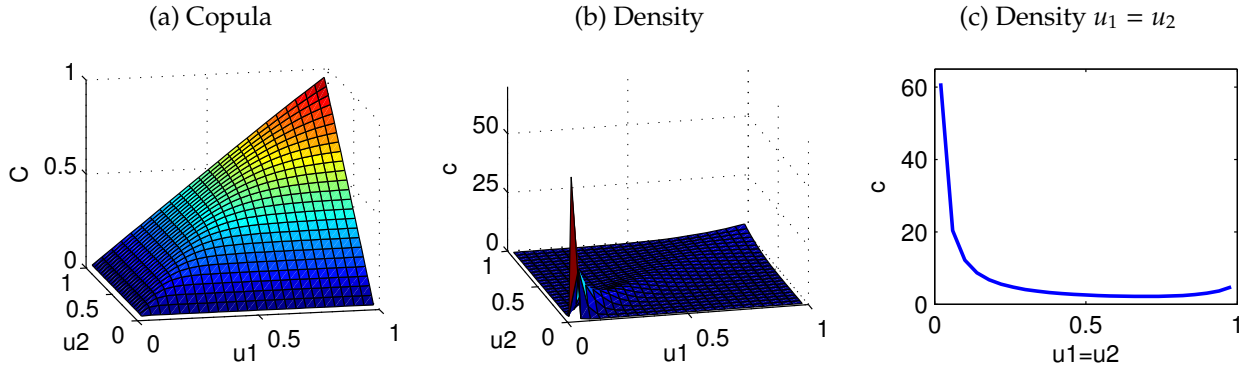
1. C is symmetric; i.e., $C(u_1, \dots, u_i, \dots, u_j, \dots, u_n) = C(u_1, \dots, u_j, \dots, u_i, \dots, u_n)$ for all $i, j = 1, 2, \dots, n$
2. C is associative, i.e. $C(C(\mathbf{u}_{-i}, u_i), \mathbf{w}_{-i}) = C(C(\mathbf{w}_{-j}, u_j), \mathbf{u}_{-j})$ for all $i, j = 1, \dots, n$, where $\mathbf{u}_{-i} = (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n) \in \mathbb{R}^{n-1}$ and similarly for $\mathbf{w}_{-i}$.
3. If $\lambda > 0$ is any constant, the $\lambda\varphi$ is also a generator for C .

Constructing Archimedean copulas is relatively easy. The only thing needed is a generator φ that satisfies the conditions in Definition 3.3.4. In this study we will particularly focus in three Archimedean copulas, namely *Clayton*, *Gumbel* and *Frank* [8, 19, 18, 29, 45] (for a more general review on Archimedean copulas see [29, Chapter 4]).

Definition 3.3.7. Given a generator of the form $\varphi(u) = u^{-\alpha} - 1$ with $\alpha \in (0, \infty)$ then, the Clayton n -copula is given by

$$C_{Cl}(\mathbf{u}) = \max[(u_1^{-\alpha} + \dots + u_n^{-\alpha} - n + 1)^{-1/\alpha}, 0].$$

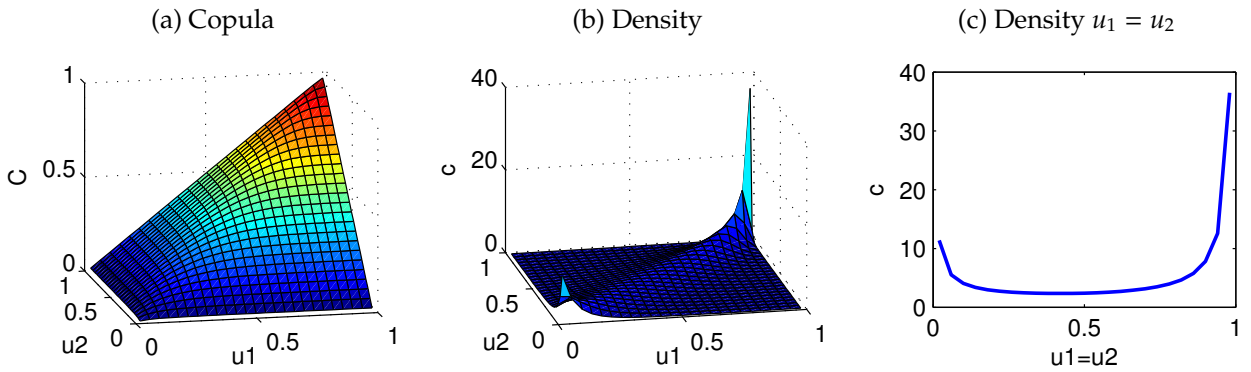
Figure 3.2: Clayton with $\alpha = 4.67$ and $\tau = 0.7$



Definition 3.3.8. Given a generator of the form $\varphi(u) = (-\ln(u))^\alpha$ with $\alpha \in (1, \infty)$ then, the Gumbel n -copula is given by

$$C_{Gu}(\mathbf{u}) = \exp \left\{ - [(-\ln u_1)^\alpha + \dots + (-\ln u_n)^\alpha]^{1/\alpha} \right\}.$$

Figure 3.3: Gumbel with $\alpha = 3.33$ and $\tau = 0.7$

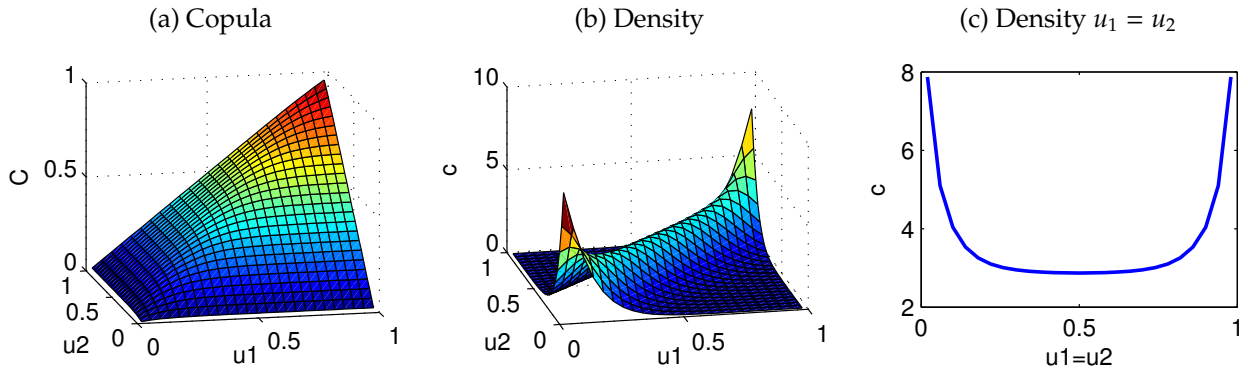


Definition 3.3.9. Given a generator of the form $\varphi(u) = \ln \left(\frac{\exp(-\alpha u) - 1}{\exp(-\alpha) - 1} \right)$ with $\alpha \in (0, \infty)$ then, the

Frank n -copula is given by

$$C_{Fr}(\mathbf{u}) = -\frac{1}{\alpha} \ln \left\{ 1 + \frac{(e^{-\alpha u_1} - 1) \cdot \dots \cdot (e^{-\alpha u_n} - 1)}{(e^{-\alpha} - 1)^{n-1}} \right\}.$$

Figure 3.4: Frank with $\alpha = 11.41$ and $\tau = 0.7$



For the bivariate case the ranges of α are slightly different from the ones given in Definitions 3.3.7 and 3.3.9. For the bivariate case the ranges of α become $[-1, 0) \cup (0, \infty)$ and $(-\infty, 0) \cup (0, \infty)$ respectively.

There is a number of reasons why we are interested in these three copulas. Firstly, they are easy to calibrate, something that will become evident in Section 3.4. Secondly, there are easy methods for simulating data, as demonstrated by Melchiori [27]. Furthermore, these copulas have some interesting properties that are best illustrated in Figures 3.2-3.4. Figures 3.2-3.4 show plots of the three copulas, together with their density functions and the diagonal of each density function ($u_1 = u_2$). Comparing Figures 3.2-3.4 to Figure 3.1, we can see that, with exception of Frank, Clayton and Gumbel have an asymmetric dependence structure. Clayton implies a strong negative dependence between two variables, whereas Gumbel a strong positive one. These asymmetries prove to be very useful in modelling. For the selection of margins, we can assign $u_i = F_i(x_i)$, where F_i can be any univariate distribution of our choice, and does not have to be the same F_j . In fact, each margin can be a different univariate distribution.

Before we end this subsection we would like to make a remark regarding Archimedian copulas. Although Archimedian copulas are very easy to construct, they have one main disadvantage. Most of them, like the ones given above, have only one free parameter, α , in their construction (some examples of Archimedian copulas with two free parameters can be found in [8, 29]). This restricts their flexibility in describing the dependence between more than two distributions. Ways for addressing this issue are discussed later on in this thesis.

3.3.4 Non-Parametric Copulas

Non-parametric copulas are mainly used to describe sets of data and do not make any assumptions about their copula. This prevents any systematic error that arises from modelling assumptions. The simplest form of a non-parametric copula is the empirical one.

Definition 3.3.10. Suppose we have an n -dimensional empirical sample of length T , denoted $\{\bar{x}_1^{(t)}, \bar{x}_2^{(t)}, \dots, \bar{x}_n^{(t)}\}_{t=1}^T$, and having rank statistics $\{\bar{r}_1^{(t)}, \bar{r}_2^{(t)}, \dots, \bar{r}_n^{(t)}\}_{t=1}^T$. The function $\hat{C}$ that is defined on the lattice

$$\mathbf{L}_n^T = \left\{ \left(\frac{i_1}{T}, \frac{i_2}{T}, \dots, \frac{i_n}{T} \right) : i_j = 0, 1, \dots, T; j = 1, 2, \dots, n \right\}$$

as

$$\hat{C}\left(\frac{i_1}{T}, \frac{i_2}{T}, \dots, \frac{i_n}{T}\right) = \frac{1}{T} \sum_{t=1}^T \prod_{j=1}^n \mathbf{1}_{[\bar{r}_j^{(t)} \leq i_j]}$$

is called an empirical copula.

Smillie [45] argues that this, in reality, is an empirical subcopula because its not defined outside of $\mathbf{L}_n^T$. Thus, he gives an extension to Definition 3.3.10 which applies to the whole of $\mathbb{I}^n$,

$$\hat{C}(\mathbf{u}) = \frac{1}{T} \sum_{t=1}^T \prod_{j=1}^n \mathbf{1}_{[\bar{r}_j^{(t)} \leq u_j T]}.$$

In addition, Smillie [45] mentions a further two non-parametric methods for approximating copulas. The first one is the *Partitions of Unity* and the second are the *Kernel Estimators*.

We give only the definition of the former, because as Smillie [45] demonstrates, the latter is not very efficient, especially when the dimensions of the system are increased. For further reading refer to [8, 29, 37, 45] and references therein.

Definition 3.3.11. *A collection of functions $h_1, \dots, h_n \in L^1[(0, 1)]$ is called a Partition of Unity if*

1. $h_i(u) \geq 0$ for $i = 1, 2, \dots, n$
2. $\int_0^1 h_i(t)dt = \frac{1}{n}$ for $i = 1, 2, \dots, n$
3. $\sum_{i=1}^n h_i(u) = 1$ for all $u \in [0, 1]$.

Theorem 3.3.12. *Given a partition of Unity $h_1, \dots, h_d$ and an n -place subcopula C' defined on the lattice $\mathbf{L}_n^d$ (see Definition 3.3.10), the function $C_h : \mathbb{I}^n \rightarrow \mathbb{I}$ defined by*

$$C_h(\mathbf{u}; C') = d^n \sum_{i_1=1}^d \sum_{i_2=1}^d \dots \sum_{i_n=1}^d \Delta_{i_1, i_2, \dots, i_n}(C') \prod_{j=1}^n \int_0^{u_j} h_{i_j}(t)dt$$

is a copula.

For the proof see e.g., [45].

3.3.5 Mixture Copula

Mixture copula provide the flexibility of using known copulas to create new ones through a convex combination. All copula properties given in Definition 3.1.10 are preserved. Hence, the end result is a new copula.

Definition 3.3.13. *Given a set of n -copulas C_i and weights $p_i \in (0, 1)$, for $i = 1, 2, \dots, d$ such that $\sum_{i=1}^d p_i = 1$ the Mixture Copula is defined as*

$$C_M(\mathbf{u}) = p_1 C_1(\mathbf{u}) + \dots + p_d C_d(\mathbf{u}).$$

C_i , for $i = 1, 2, \dots, d$, can be any copula independent of others in the convex combination. Actually, if they are the same, i.e. $C^{(i)} \equiv C$ for all $i = 1, 2, \dots, d$, the mixture copula will reduce to its component, $C_M \equiv C$.

Smillie [45], in his thesis proposes a specific mixture copula which only uses multivariate normal distributions and he names it *Gaussian Mixture* copula.

Definition 3.3.14. *Given a d -multivariate normal distribution functions, each parametrised by mean vectors $\mu^{(k)} \in \mathbb{R}^n$, positive definite correlation matrices $P^{(k)} \in [-1, 1]^{n \times n}$, variance vectors $\sigma^{(k)} \in [0, \infty)^n$ and probabilities $p^{(k)}$ such that $\sum_{k=1}^d p^{(k)} = 1$, we define the Gaussian Mixture distribution*

$$\Phi_{GM}(\mathbf{x}) \equiv \sum_{k=1}^d p^{(k)} \Phi(\mathbf{x}; \mu^{(k)}, \sigma^{(k)}, P^{(k)}).$$

Definition 3.3.15. *Given an n -place Gaussian mixture distribution function Φ_{GM} , the Gaussian Mixture copula is the function $C_{GM} : \mathbb{I}^n \rightarrow \mathbb{I}$ such that*

$$C_{GM}(\mathbf{u}) = \Phi_{GM}(\Phi_{GM_1}^{-1}(u_1), \Phi_{GM_2}^{-1}(u_2), \dots, \Phi_{GM_n}^{-1}(u_n))$$

where Φ_{GM_i} denotes the i -th marginal distribution, i.e.

$$\Phi_{GM_i}(\mathbf{x}) \equiv \sum_{k=1}^d p^{(k)} \Phi(x_i; \mu_i^{(k)}, \sigma_i^{(k)}).$$

Using the above framework, Smillie [45] constructs a model that is easy to build since its components are the very well known multivariate normal distributions. Also, the model can easily handle many dimensions because of the virtues of multivariate normal distribution. At the same time, it can exhibit behaviours that the multivariate normal distribution on its own cannot.

Similarly, Hu [18, 19] makes use of the Archimedian copulas and a single multivariate normal

distribution. In [18] he uses Clayton, Gumbel and Frank copulas (see Definitions 3.3.7-3.3.9), which exhibit different dependence behaviours, to describe the link between stock indices. The same approach is followed in [19], but instead of the Frank copula the author uses a multivariate normal distribution.

Hu's approach overcomes the parameter restriction of the Archimedian copulas. By using a mixture copula, Hu, creates an "extended" Archimedian copula with more free parameters, demonstrating the flexibility that mixture copulas provide.

3.4 Copula Calibration Techniques

In this section we go through the most common copula calibration techniques. We consider some general methodologies and some that are directly linked to specific copulas. Most of this section reviews the work done in [8, Chapter 5]. We follow a similar structure and add references where appropriate.

In [8], it is mentioned that the most of the statistical inference theory is not applicable to copulas. The only methodology that is applicable is the *Maximum Likelihood Estimation* (MLE). Also, in most cases, an analytic solution of the objective function will not exist. Thus, a numerical approximation is required.

Throughout this section we consider time-series data used for the calibration of copulas. We denote the time-series of observed data by $\bar{x}_{i,t}$ for variables $i = 1, 2, \dots, n$ and time-steps $t = 1, 2, \dots, T$.

3.4.1 Maximum Likelihood based Methods

Consider the simple MLE method. There are three ways to approach the MLE method for copulas. The first one, also referred the "*One step MLE*" [45], follows the classical

approach where the log-likelihood function is maximised. Here, we consider an n -copula C with marginals $F_1, \dots, F_n$. Also, we denote by θ_C the parameters of the copula, by $\theta_1, \dots, \theta_n$ the parameters of the corresponding marginals (θ_i can be a set of parameters), and by $\theta = \{\theta_C, \theta_1, \dots, \theta_n\}$ the entire parameter set. Then, the log likelihood function is defined as

$$l(\theta) = \sum_{t=1}^T \ln c(F_1(\bar{x}_{1,t}), \dots, F_n(\bar{x}_{n,t})) + \sum_{t=1}^T \sum_{i=1}^n \ln f_i(\bar{x}_{i,t}), \quad (3.15)$$

where c is the copula density function from Definition 3.1.15 and f_i is the distribution density function of marginal F_i . Using (3.15) we can find the empirical estimate of θ by

$$\hat{\theta} = \max_{\theta} l(\theta).$$

In [8] it is shown that under regularity conditions the MLE exists, is consistent and is asymptotically efficient and normally distributed ²

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta_0) \xrightarrow{d} N(0, \mathfrak{F}^{-1}(\theta_0)),$$

where $\mathfrak{F}(\theta_0)$ is Fisher's information matrix and θ_0 is the true value (for more information see [8] and the references therein).

Although one step MLE is quite straight forward, the fact that the whole set θ is maximised simultaneously makes the method slow and inefficient, especially when n is large. To overcome this, the "Two Step MLE" was proposed. Also known as *Inference Function for Marginals* method (IFM), the two step MLE takes advantage of the structure of copulas [8, 45]. Since copulas couple together the distribution functions of different variables and do not make direct use of the random variables themselves, it is possible to first calibrate each marginal individually before calibrating the copula parameters θ_C . Thus, we define the

²The symbol $\xrightarrow{d}$ defines convergence in distribution.

likelihood function of the univariate marginals

$$l_i(\theta_i) = \sum_{t=1}^T \ln f_i(\bar{x}_{i,t})$$

and find the empirical estimate

$$\hat{\theta}_i = \max_{\theta_i} l_i(\theta_i)$$

for $i = 1, 2, \dots, n$. After estimating the θ_i of the marginals, we construct the likelihood of the copula

$$l_C(\theta_C | \theta_1, \dots, \theta_n) = \sum_{t=1}^T \ln c(F_1(\bar{x}_{1,t} | \hat{\theta}_1), \dots, F_n(\bar{x}_{n,t} | \hat{\theta}_n))$$

and we calculate the empirical estimate

$$\hat{\theta}_C = \max_{\theta_C} l_C(\theta_C | \theta_1, \dots, \theta_n).$$

IFM is not equivalent to MLE [8] but it has similar limiting behaviours

$$\sqrt{T}(\hat{\theta}_{IFM} - \theta_0) \xrightarrow{d} N(0, \mathbb{G}^{-1}(\theta_0))$$

with $\mathbb{G}(\theta_0)$ be the Godambe information matrix [8].

The final method that involves MLE is called *Canonical Maximum Likelihood* (CML). Instead of making assumptions about the distribution of the marginals F_i and calibrating their parameters θ_i , the marginals are calculated empirically, defined as $\hat{F}_i$ (see Subsection 3.3.4; the same methods can be applied to a univariate distribution) and then used to calibrate the parameters θ_C of the copula. Thus, in this case the log-likelihood function becomes

$$l_C(\theta_C) = \sum_{t=1}^T \ln c(\hat{F}_1(\bar{x}_{1,t}), \dots, \hat{F}_n(\bar{x}_{n,t}))$$

and $\hat{\theta}_C$ is estimated in a similar manner as previously.

3.4.2 Calibration Methods Using Measures of Dependence

For some families of copulas, it is possible to use measures of dependence in order to estimate their parameters (for an extensive discussion see [8] and the references therein). This applies to the Archimedean Copulas defined in Subsection 3.3.3. Let us denote the free parameters of each of the Archimedean copulas defined in Subsection 3.3.3 by α_{Cl} for Clayton (Definition 3.3.7), α_{Gu} for Gumbel (Definition 3.3.8), and α_{Fr} for Frank (Definition 3.3.9). For these three cases, we have closed form relations with one of Spearman's ρ_S (3.8) and Kendall's τ (3.9) or both. For Clayton we have that

$$\tau = 1 - \alpha_{Cl}^{-1}, \quad (3.16)$$

for Gumbel

$$\tau = \frac{\alpha_{Gu}}{\alpha_{Gu} + 2}, \quad (3.17)$$

and for Frank we have

$$\tau = 1 + \frac{4[D_1(\alpha_{Fr})]}{\alpha_{Fr}} \quad (3.18)$$

$$\rho_S = 1 - \frac{12[D_2(-\alpha_{Fr}) - D_1(-\alpha_{Fr})]}{\alpha_{Fr}}, \quad (3.19)$$

where

$$D_k(\alpha) = \frac{k}{\alpha^k} \int_0^\alpha \frac{x}{\exp(x) - 1} dx \text{ for } k = 1, 2.$$

The measures of dependence can also be used in the calibration of the Gaussian copula (see Definition 3.3.1). Under the assumption that the data are normally distributed the following two relations hold [45]

$$\rho = \sin\left(\frac{\pi\tau}{2}\right), \quad (3.20)$$

$$\rho = 2 \sin\left(\frac{\pi\rho_S}{6}\right), \quad (3.21)$$

where ρ is as defined in (3.7).

3.5 Examples

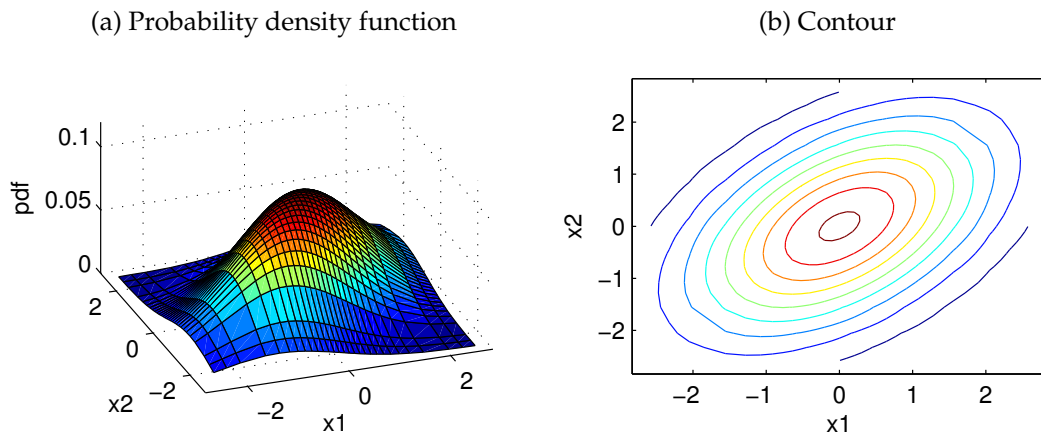
Before this chapter is brought to a close, we would like to go through some examples. These examples will serve as means of illustrating some of the properties and characteristics of the copulas and their theory that have been covered in this chapter.

3.5.1 Modelling flexibility

From Definition 3.1.10, we have that copulas are a function $C : \mathbb{I}^n \rightarrow \mathbb{I}$. In general, they are not directly associated to a known probability distribution $F(\mathbf{x})$, although Theorem 3.1.13 allows us to construct one from the other. Thus, with the exception of copulas constructed by the inversion method (Subsection 3.3.1), random variable X_i can follow any univariate probability distribution F_i , and then one can assign $u_i = F_i(x_i)$. This principle is also reflected in the two step MLE method discussed in Section 3.4.

For comparison purposes let us begin with the simple case where the random variables $X_1, X_2 \sim N(0, 2)$ and their dependence is described by Gaussian copula with $r = 0.5$ (Definition 3.3.1). The probability density function and its contours are shown in Figure 3.5.

Figure 3.5: Gaussian probability density function with $X_1, X_2 \sim N(0, 2)$ and $r = 0.5$



For our first example, suppose that we would like to describe the dependence between X_1, X_2 with linear correlation, but we do not believe that X_1 and X_2 are normally distributed. Assume that in this instance $X_1, X_2 \sim \chi_6^2$, the chi-square distribution with six degrees of freedom. Then the copula will become

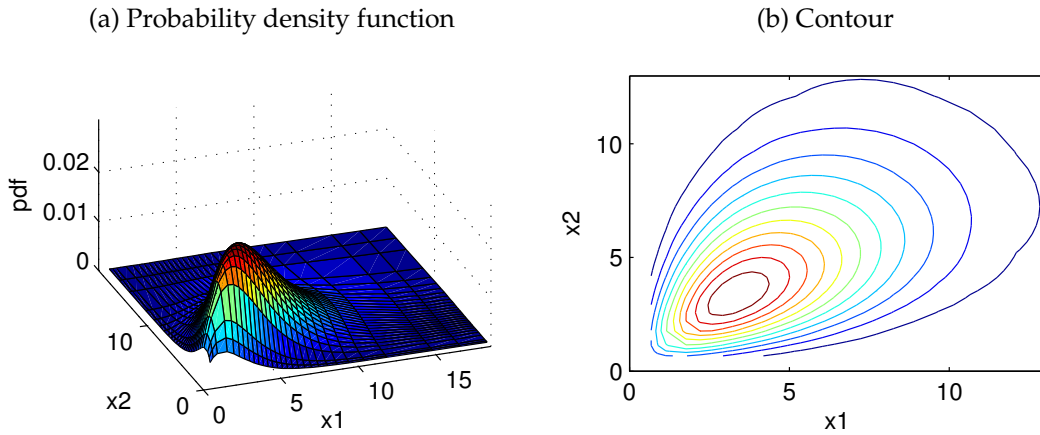
$$C_{Ga}(F_{\chi_6^2,1}(x_1), F_{\chi_6^2,2}(x_2)) = \Phi(\Phi^{-1}(F_{\chi_6^2,1}(x_1)), \Phi^{-1}(F_{\chi_6^2,2}(x_2))) \quad (3.22)$$

and from (3.5) we have that the probability density function is

$$f(x_1, x_2) = c_{Ga}(F_{\chi_6^2,1}(x_1), F_{\chi_6^2,2}(x_2)) f_{\chi_6^2,1}(x_1) f_{\chi_6^2,2}(x_2). \quad (3.23)$$

In Figure 3.6, we illustrate what this newly created $f(x_1, x_2)$ looks like. It can be seen that the original structure of Gaussian probability density function, seen in Figure 3.5, has been totally altered. Now we have an asymmetric function with all its mass concentrated on the positive quantile.

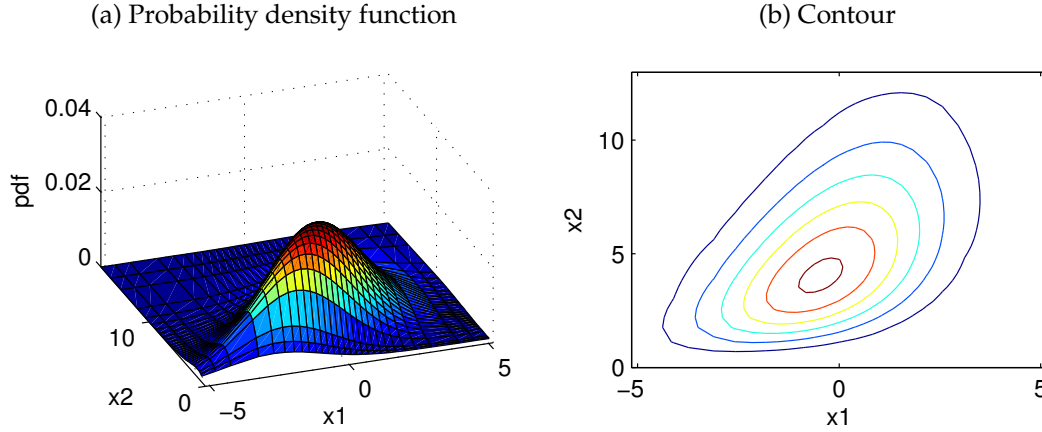
Figure 3.6: Combining the Gaussian copula and chi-square margins with six degrees of freedom, $X_1, X_2 \sim \chi_6^2$



We conclude this section with an example using one of the Archimedian copulas (Section 3.3.3). In particular, we use Clayton (Definition 3.3.7). For the margins we assume that $X_1 \sim N(0, 2)$ and $X_2 \sim \chi_6^2$, i.e. $u_1 = \Phi(x_1)$ and $u_2 = F_{\chi_6^2,2}(x_2)$. The probability density function is given in Figure 3.7. The strong negative dependence of Clayton shown in Figure 3.2, is

presented here in the form of the long tail towards the $(-5,0)$ mark.

Figure 3.7: Combining the Clayton copula with a Gaussian $X_1 \sim N(0,2)$ and a chi-square $X_2 \sim \chi^2_6$



The above examples in Figures 3.5-3.7, demonstrate how flexible copulas can be, particularly when the idea of mixture copulas is added, as discussed in Section 3.3.5.

3.5.2 Comparing quantile dependence

Reconsider the copulas shown in Figures 3.1-3.4, i.e., the Gaussian and the three Archimedean copulas, Clayton, Gumbel and Frank. All four copulas were shown for the case where $\tau = 0.7$ (see (3.9)). Consider the case that τ was calculated from a dataset. From Section 3.4.2 we know that we can use τ to calibrate all four copulas. Figures 3.1-3.4 though, suggest that using different copulas should give different results. Indeed, if we were to inspect them independently we would have never considered the fact that they represent the same dataset. This demonstrates the importance of carefully selecting a copula to represent a dataset. Hence, we cannot depend on a single measure of dependence, to understand the underlying characteristics of a given dataset. Thus, in this example we will make use of the measure of quantile dependence ν (3.13) and its limiting case the tail dependence λ (3.12), to help us investigate further the four copulas.

Figure 3.8: The lower quantile dependence ν_L and upper quantile dependence ν_U of Gaussian, Clayton, Gumbel and Frank copulas for $\tau = 0.7$. The values at the boundaries give the lower tail dependence λ_L and upper tail dependence λ_U .

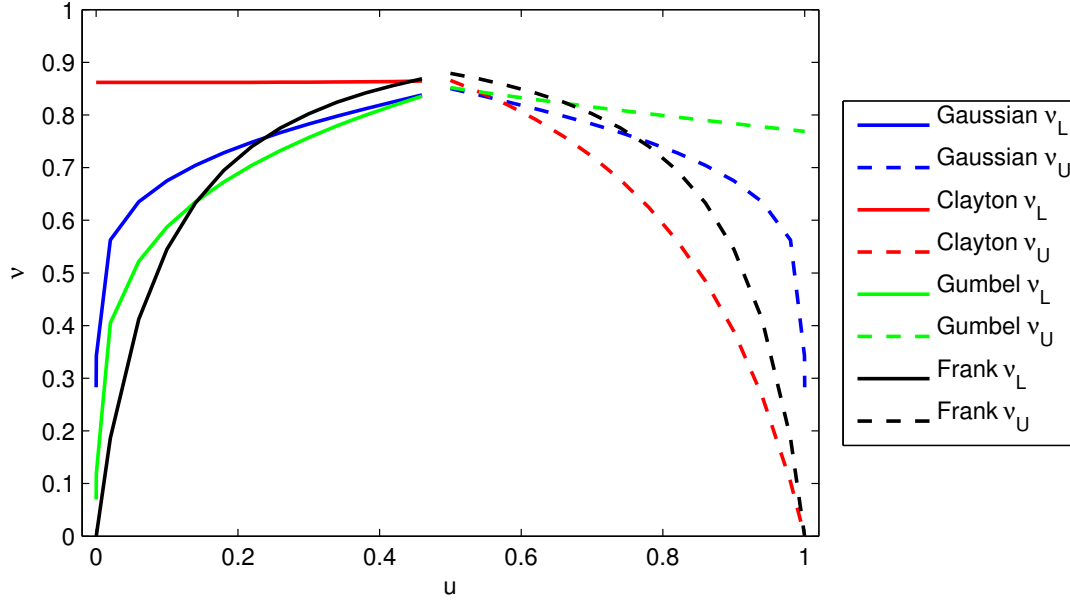


Figure 3.8 shows the lower and upper quantile dependence of the four copulas in question. Gaussian and Frank copulas are symmetric with respect to quantile dependence, i.e., $\nu_L(u) = \nu_U(1 - u)$, with Gaussian copula having higher ν than Frank, as u approaches the boundaries. For both copulas, $\lambda_L = \lambda_U = 0$, i.e., they both have zero tail dependence. This is not the case for Clayton and Gumbel copulas, which are asymmetric in their quantile dependence. Clayton copula has high ν_L and $\lambda_L > 0$. However it has the lowest ν_U in the group, with $\lambda_U = 0$. On the other side of the spectrum, Gumbel copula has high ν_U and $\lambda_U > 0$, while it has low ν_L and $\lambda_L = 0$.

Consider a real example using FTSE and DAX30 stock indexes. Kendall's τ for the historical data is 0.52. From this value alone we can create all the above copulas, with the Gaussian copula being the most likely candidate, but by quick inspection of Figure 3.9, we can see that Gaussian copula does not match the empirical copula density function created from the dataset. The empirical function is asymmetric with heavier lower tail. This can be confirmed with the calculation of the quantile dependence between FTSE and DAX30. Figure 3.10

illustrates that v_L is significantly higher as $u \rightarrow 0$, than v_U is as $u \rightarrow 1$. This behaviour bears closer resemblance to Clayton copula than the Gaussian copula. It follows that copulas are extremely useful in the modelling of financial assets.

Figure 3.9: The diagonal of the empirical copula density function between FTSE and DAX 30 returns.

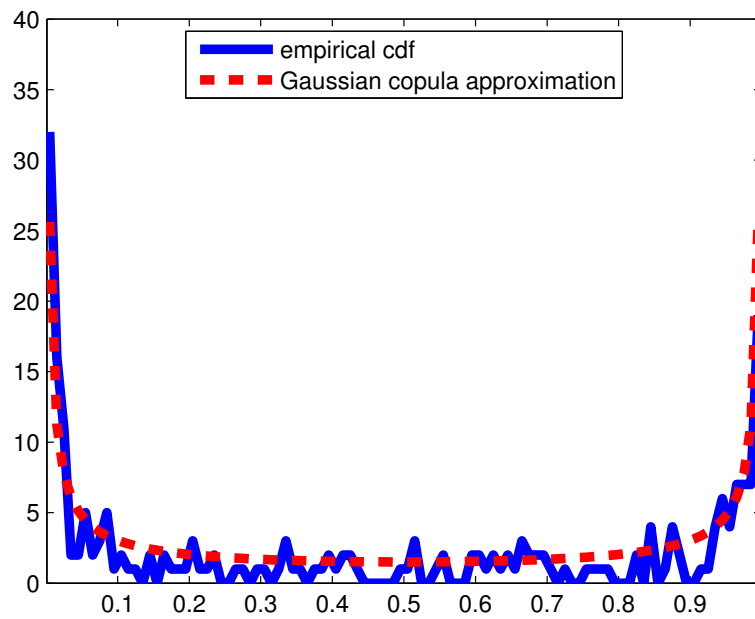
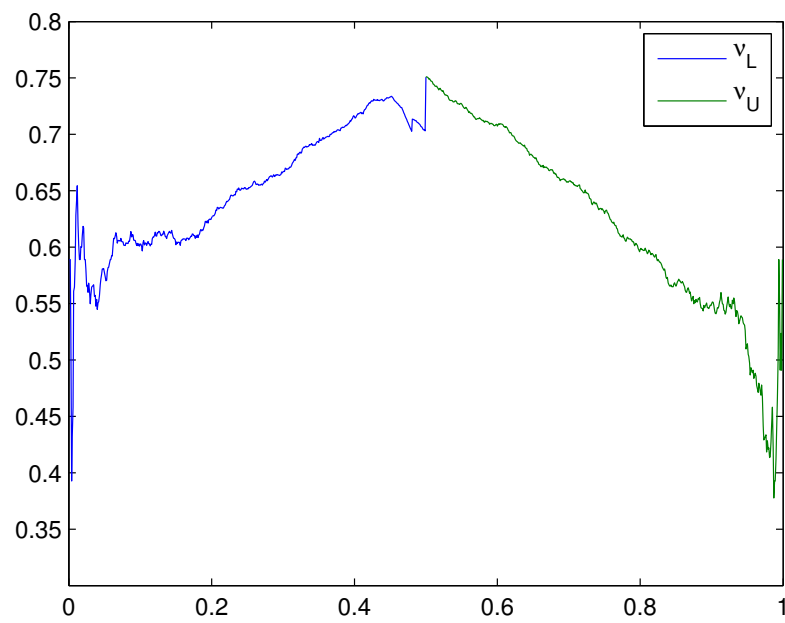


Figure 3.10: Empirical quantile dependence of FTSE and DAX 30 returns.



Chapter 4

Robust Portfolio Optimisation with Mixture Copulas

In this chapter we look into the problem of portfolio optimisation where the assets of the portfolio are described by random variables. In this situation the selection of the optimal portfolio depends on the underlying assumptions on the behaviour of the assets and the choice on the measure of risk. Usually the objective is to find the optimal risk-return trade-off.

One of the pioneers in portfolio optimisation was Markowitz [25] who proposed the mean-variance framework for risk return analysis. Although the most common measure for the estimation of the return of the portfolio remains the expected return, many other ways of calculating the risk have been developed. A widely used measure of risk is *Value at Risk* (VaR). VaR is the measure of risk that is recommended as a standard by the Basel Committee. However, VaR has been criticised in recent years mainly for two reasons. Firstly, VaR does not satisfy sub-additivity and hence it is not a coherent measure of risk in the way that is defined by Artzner et al.[2]. Also, as discussed in Section 2.3, the convexity of VaR is highly dependent on the assumptions of the underlying distribution. Secondly, it gives a percentile of the loss distribution that does not provide an adequate picture of the possible losses in the

tail of the distribution. Szego [46] uses this argument to state that “VaR does not measure risk”. He then goes on to recommend alternative measures of risk with one of them being *Conditional Value at Risk* (CVaR).

CVaR is the expectation of the distribution above VaR. Thus, the value of CVaR is affected by the thickness of the tail of the distribution. Hence, CVaR provides a better description of the loss on the tail of the distribution. Rockafellar and Uryasev [33, 34] propose a minimisation formulation that usually results in a convex or linear problem. These are desirable properties of the CVaR and have paved the way for its use in risk management and portfolio optimisation. A literature review on CVaR can be found in [48] and references therein.

Following the formulation in [33, 34], in order to calculate CVaR, one has to make some assumptions on the underlying distribution of the assets. This can also be in the form of an uncertainty domain, like a hypercube or an ellipsoidal set in which all feasible uncertainty values lie [48] (see (E2.3.4)). Alternatively, one can assume that the assets are described by some multivariate distribution [48]. In this chapter we focus on the selection of multivariate distributions.

The Gaussian is the most commonly used multivariate distribution. Not only is it easy to calibrate, but there are also very efficient algorithms to simulate Gaussian data. This also applies to some extent to the elliptical family of distributions with student- τ distribution being widely used in Credit Risk [7, 9, 35]. One disadvantage of using the Gaussian distribution is its symmetry (see Figures 3.1 and 3.8). This implies that the probability of losses is the same as the probability of gains. Studies suggest that at least in the context of financial markets, assets exhibit stronger comovements during a crisis as opposed to prosperity [1, 18, 19]. The second disadvantage, is that it uses linear correlation as a measure of dependence. As the name suggests, linear correlation is characterised by linear dependencies. Since the observation of asymmetric comovements, mentioned above, suggests non-linear dependencies, linear correlation may not be an adequate measure of dependence [2, 46].

One way of addressing the limitations of the symmetry underlying elliptical distributions is

to consider mixture distributions. A convex combination of a set of distributions is used to fit the given sample by optimising the combination of weights. Hasselblad [16, 17] was one of the first to look into mixture distributions and how their parameters can be estimated. Zhu and Fukushima [48] avoid the assumptions needed on the set of distributions and their parameters and also avoid the estimation of the weights. Subsets of historical returns are used to represent data arising from different distributions and a worst-case scenario approach is applied to avoid the calibration of the weights. In [18, 19, 45] they use mixture copulas to fit their data samples for the bivariate case. The work of Hu [18, 19] and Zhu and Fukushima [48], motivates us to introduce copulas within a worst case scenario framework.

The reasoning behind the use of copulas is to take advantage of their properties discussed in Chapter 3. We are interested in the separation of the univariate distribution selection from the copula, the association of copulas with monotonic measures of dependence and the asymmetric properties of the three Archimedian copulas discussed in Section 3.3.3. The combination of copulas within the worst case scenario framework provides us a robust and flexible model for assessing the dependence between financial assets.

The chapter is structured as follows. In Section 4.1, we derive the CVaR for copulas. We extend CVaR to *Worst Case Conditional Value at Risk* (WCVaR) through the use of mixture copulas. We conclude the section by stating the generalised optimisation problem for the WCVaR. In Section 4.2, we construct a portfolio optimisation model based on the theory of the previous section. We also introduce the competing models that ours will be compared against. Then, in Section 4.3, we discuss the technical details of our numerical analysis and we provide numerical examples that assess the performance of our model. Finally, we end with summary, future work and conclusions.

4.1 Worst Case Conditional Value at Risk

In Section 3.1 we introduced the theorems linking copulas to probability distributions. In this section we will derive the copula formulation of the WCVaR. Zhu and Fukushima [48] derive the WCVaR for distributions. We follow a similar approach to derive WCVaR for copulas. At every step involving the use of distributions we present the equivalent copula formulation. For the derivation of the copula formulation we use equations (3.3)-(3.5).

In order to define the WCVaR we first have to define VaR and CVaR. We first discuss VaR.

Proposition 4.1.1. *Let $\mathbf{w} \in \mathbb{W} \subseteq \mathbb{R}^m$ be a decision vector, $\mathbf{u} \in \mathbb{I}^n$ a random vector, $\tilde{g}(\mathbf{w}, \mathbf{u})$ the cost function and $\mathbf{F}(\mathbf{x}) = (F_1(x_1), \dots, F_n(x_n))^T$ a set of marginal distributions where $\mathbf{u} = \mathbf{F}(\mathbf{x})$. Also, assume that $\mathbf{u}$ follows a continuous distribution with copula function $C(\cdot)$. Then for a confidence level β , VaR_β is defined as*

$$\text{VaR}_\beta(\mathbf{w}) = \min\{\alpha \in \mathbb{R} : C(\mathbf{u} \mid \tilde{g}(\mathbf{w}, \mathbf{u}) \leq \alpha) \geq \beta\}. \quad (4.1)$$

Proof. Given a decision $\mathbf{w} \in \mathbb{W}$ and a random vector $\mathbf{x} \in \mathbb{R}^n$ which follows a continuous distribution with density function $f(\cdot)$, the probability of $g(\mathbf{w}, \mathbf{x})$ not exceeding a threshold α is represented by

$$\begin{aligned} \Psi(\mathbf{w}, \alpha) &= \int_{g(\mathbf{w}, \mathbf{x}) \leq \alpha} f(\mathbf{x}) d\mathbf{x} \\ &= \int_{g(\mathbf{w}, \mathbf{x}) \leq \alpha} c(\mathbf{F}(\mathbf{x})) \prod_{i=1}^n f_i(x_i) d\mathbf{x} \\ &= \int_{\tilde{g}(\mathbf{w}, \mathbf{u}) \leq \alpha} c(\mathbf{u}) d\mathbf{u} \\ &= C(\mathbf{u} \mid \tilde{g}(\mathbf{w}, \mathbf{u}) \leq \alpha), \end{aligned}$$

where $f_i(x_i) = \frac{\partial F_i(x_i)}{\partial x_i}$ is the univariate probability distribution of the individual elements of the random vector $\mathbf{x}$ and $c(\cdot)$ is the copula density function of $C(\cdot)$ (see Definition 3.1.15).

$\tilde{g}(\mathbf{w}, \mathbf{u}) = g(\mathbf{w}, \mathbf{F}^{-1}(\mathbf{u}))$ where $\mathbf{F}^{-1}(\mathbf{u}) = (F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))^T$ which transforms the domain of $g(\cdot, \cdot)$ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n \times \mathbb{I}^n$ from $\mathbb{R}^n$ to $\mathbb{I}^n$, as implied by the transformation $u_i = F_i(x_i)$. For the derivation of the copula version of $\Psi(\mathbf{w}, \alpha)$ we use equation (3.5). Having defined $\Psi(\mathbf{w}, \alpha)$, we consider the VaR. Given a fixed $\mathbf{w} \in \mathbb{W}$ and a confidence level β , VaR_β is defined as

$$\begin{aligned}\text{VaR}_\beta(\mathbf{w}) &= \min\{\alpha \in \mathbb{R} : \Psi(\mathbf{w}, \alpha) \geq \beta\} \\ &= \min\{\alpha \in \mathbb{R} : C(\mathbf{u} \mid \tilde{g}(\mathbf{w}, \mathbf{u}) \leq \alpha) \geq \beta\}.\end{aligned}$$

□

We proceed with the definition of the CVaR with respect to VaR.

Proposition 4.1.2. *Given $\mathbf{w}$, $\mathbf{u}$, $\mathbf{F}(\mathbf{x})$ and $\tilde{g}(\mathbf{w}, \mathbf{u})$ as in Proposition 4.1.1 we define CVaR_β for a confidence level β as*

$$\text{CVaR}_\beta(\mathbf{w}) = \frac{1}{1 - \beta} \int_{\tilde{g}(\mathbf{w}, \mathbf{u}) \geq \text{VaR}_\beta(\mathbf{w})} \tilde{g}(\mathbf{w}, \mathbf{u}) c(\mathbf{u}) d\mathbf{u}. \quad (4.2)$$

Proof. We start from the equation of CVaR that arises from the probability density function $f(\cdot)$ and we derive the copula form.

$$\begin{aligned}\text{CVaR}_\beta(\mathbf{w}) &= \frac{1}{1 - \beta} \int_{g(\mathbf{w}, \mathbf{x}) \geq \text{VaR}_\beta(\mathbf{w})} g(\mathbf{w}, \mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{1 - \beta} \int_{g(\mathbf{w}, \mathbf{x}) \geq \text{VaR}_\beta(\mathbf{w})} g(\mathbf{w}, \mathbf{x}) c(\mathbf{F}(\mathbf{x})) \prod_{i=1}^n f_i(x_i) d\mathbf{x} \\ &= \frac{1}{1 - \beta} \int_{\tilde{g}(\mathbf{w}, \mathbf{u}) \geq \text{VaR}_\beta(\mathbf{w})} \tilde{g}(\mathbf{w}, \mathbf{u}) c(\mathbf{u}) d\mathbf{u}.\end{aligned} \quad (4.3)$$

□

Following [33] we re-write equation (4.3) as

$$\begin{aligned} G_\beta(\mathbf{w}, \alpha) &= \alpha + \frac{1}{1-\beta} \int_{\mathbf{x} \in \mathbb{R}^n} [g(\mathbf{w}, \mathbf{x}) - \alpha]^+ f(\mathbf{x}) d\mathbf{x} \\ &= \alpha + \frac{1}{1-\beta} \int_{\mathbf{u} \in \mathbb{I}^n} [\tilde{g}(\mathbf{w}, \mathbf{u}) - \alpha]^+ c(\mathbf{u}) d\mathbf{u}. \end{aligned} \quad (4.4)$$

where $[x]^+$ is defined as $\max\{0, x\}$, and which gives us the following minimisation problem

$$\text{CVaR}_\beta(\mathbf{x}) = \min_{\alpha \in \mathbb{R}} G_\beta(\mathbf{w}, \alpha). \quad (4.5)$$

By solving the minimisation problem in equation (4.5), we directly obtain both the values of CVaR and VaR. From Proposition 4.1.1 we have that the value of VaR is the value of α .

In order for the above definitions to be computed, exact knowledge of the distribution $f(\mathbf{x})$ or copula density $c(\mathbf{u})$ and the margins $\mathbf{F}(\mathbf{x})$ is needed. As the aim of this thesis is to represent distributions with copulas, we shall omit using $f(\mathbf{x})$ and use $c(\mathbf{u})$ instead. The equivalence of the two has been discussed in Section 3.1. Knowledge of copula $C(\mathbf{u})$ and its margins $\{u_i = F_i(x_i)\}_{i=1}^n$ implies knowledge of $f(\mathbf{x})$ and $c(\mathbf{u})$. A copula representation of the distribution of $\mathbf{x}$ cannot be expected to be exact. Thus, we assume that our copula representation belongs to a set of copulas $c(\cdot) \in \mathcal{C}$. The notion of robustness with respect to $\mathcal{C}$ involves the worst performing copula (or copulas, since the worst case may not be unique), i.e. the copula for which we obtain the greatest CVaR. Hence, we define WCVaR.

Definition 4.1.3. *The WCVaR for fixed $\mathbf{w} \in \mathbb{W}$ and confidence level β , with respect to $\mathcal{C}$ is defined as*

$$\text{WCVaR}_\beta(\mathbf{w}) = \sup_{c(\cdot) \in \mathcal{C}} \text{CVaR}_\beta(\mathbf{w}). \quad (4.6)$$

It is known that CVaR is a coherent measure of risk [2, 46, 48]. A measure of risk ρ is coherent when it satisfies the following properties:

- (i) Subadditivity: for all random vectors X and Y , $\rho(X + Y) \leq \rho(X) + \rho(Y)$;

- (ii) Positive homogeneity : for positive constant λ , $\rho(\lambda X) = \lambda \rho(X)$;
- (iii) Monotonicity: if $X \leq Y$ for each outcome, then $\rho(X) \leq \rho(Y)$;
- (iv) Translation invariance: for constant m , $\rho(X + m) = \rho(X) + m$.

Zhu and Fukushima [48] prove that WCVaR preserves coherence. They also give the following lemma from Fan [14], which allows us to formulate the problem into a tractable one.

Lemma 4.1.4. *Suppose that $\mathbb{W}$ and $\mathbb{X}$ are nonempty convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and the function $z(\mathbf{w}, \mathbf{x})$ is convex in $\mathbf{w}$ for any $\mathbf{x}$, and concave in $\mathbf{x}$ for any $\mathbf{w}$. Then we have*

$$\min_{\mathbf{w} \in \mathbb{W}} \max_{\mathbf{x} \in \mathbb{X}} z(\mathbf{w}, \mathbf{x}) = \max_{\mathbf{x} \in \mathbb{X}} \min_{\mathbf{w} \in \mathbb{W}} z(\mathbf{w}, \mathbf{x}) \quad (4.7)$$

We also use Lemma 4.1.4 to extend the proof from [48] to copulas and eventually formulate our problem as a *minmax* problem.

4.1.1 Mixture Copula

In this example the distribution of the vector of returns $\mathbf{x}$ is described by a mixture copula

$$C(\mathbf{F}(\mathbf{x})) = \boldsymbol{\lambda}^T \vec{C}, \quad (4.8)$$

where $\boldsymbol{\lambda} \in \Lambda = \{\boldsymbol{\lambda} : \mathbf{e}^T \boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \in \mathbb{R}^l\}$ and $\vec{C} = (C_1(\mathbf{F}(\mathbf{x})), \dots, C_l(\mathbf{F}(\mathbf{x})))^T$ is the vector with copulas and $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), \dots, F_n(\mathbf{x}))^T$ is the vector of the cumulative univariate distributions. We can apply equation (3.5) to equation (4.8) to obtain the density of the mixture copula. Then, we can use this density in (4.1.2) to obtain the CVaR of the mixture copula. The CVaR

of the mixture copula, given in (4.4) representation, is

$$G_\beta(\mathbf{w}, \lambda, \alpha) = \alpha + \frac{1}{1-\beta} \int_{\mathbf{u} \in \mathbb{I}^n} [\tilde{g}(\mathbf{w}, \mathbf{u}) - \alpha]^+ \sum_{i=1}^l \lambda_i c_i(\mathbf{u}) d\mathbf{u} = \sum_{i=1}^l \lambda_i G_\beta^i(\mathbf{w}, \alpha), \quad (4.9)$$

where

$$G_\beta^i(\mathbf{w}, \alpha) = \alpha + \frac{1}{1-\beta} \int_{\mathbf{u} \in \mathbb{I}^n} [\tilde{g}(\mathbf{w}, \mathbf{u}) - \alpha]^+ c_i(\mathbf{u}) d\mathbf{u} \quad \text{for } i = 1, 2, \dots, l, \quad (4.10)$$

is the CVaR_β of the individual copulas, used within the mixture. The optimisation problem that we need to solve is stated by the following theorem and corollary from [48]:

Theorem 4.1.5. *For each $\mathbf{w}$, $\text{WCVaR}_\beta(\mathbf{w})$ with respect to $\mathcal{C}$ is given by*

$$\text{WCVaR}_\beta(\mathbf{w}) = \min_{\alpha \in \mathbb{R}} \max_{\lambda \in \Lambda} G_\beta(\mathbf{w}, \lambda, \alpha), \quad (4.11)$$

where $\Lambda = \{\lambda : \mathbf{e}^T \lambda = 1, \lambda \geq \mathbf{0}, \lambda \in \mathbb{R}^l\}$.

Corollary 4.1.6. *Minimising $\text{WCVaR}_\beta(\mathbf{w})$ over $\mathbb{W}$ can be achieved by the following minimization*

$$\min_{\mathbf{w} \in \mathbb{W}} \text{WCVaR}_\beta(\mathbf{w}) = \min_{\mathbf{w} \in \mathbb{W}} \min_{\alpha \in \mathbb{R}} \max_{\lambda \in \Lambda} G_\beta(\mathbf{w}, \lambda, \alpha). \quad (4.12)$$

More specifically, if $(\mathbf{w}^*, \alpha^*, \lambda^*)$ attains the right hand side minimum, then $\mathbf{w}^*$ attains the left hand side minimum.

Zhu and Fukushima [48] provide the proof for the case of mixture distributions. The theorems in Section 3.1, together with Proposition 4.1.1 and Proposition 4.1.2, show that Theorem 4.1.5 and Corollary 4.1.6 can be applied to copulas. For the sake of completeness we give the motivation behind the proof and we continue with the formulation of the optimisation problem.

In order to optimise the portfolio we need to solve

$$\min_{\mathbf{w} \in \mathbb{W}} \text{WCVaR}_\beta(\mathbf{w}) \equiv \min_{\mathbf{w} \in \mathbb{W}} \max_{\lambda \in \Lambda} \min_{\alpha \in \mathbb{R}} G_\beta(\mathbf{w}, \lambda, \alpha) \quad (4.13)$$

Since the mixture copula (4.8) is linear in λ , Zhu and Fukushima [48] use Lemma 4.1.4 to show that (4.13) can be written as

$$\min_{\mathbf{w} \in \mathbb{W}} \min_{\alpha \in \mathbb{R}} \max_{\lambda \in \Lambda} G_{\beta}(\mathbf{w}, \lambda, \alpha). \quad (4.14)$$

An epigraph formulation can be used to reduce problem (4.14) into a minimisation problem as follows

$$\min_{(w, \alpha, \theta) \in \mathbb{W} \times \mathbb{R} \times \mathbb{R}} \left\{ \theta : \sum_{i=1}^l \lambda_i G_{\beta}^i(\mathbf{w}, \alpha) \leq \theta, \forall \lambda \in \Lambda \right\} \quad (4.15)$$

and θ must satisfy

$$G_{\beta}^i(\mathbf{w}, \alpha) \leq \theta, \quad \text{for } i = 1, 2, \dots, l. \quad (4.16)$$

Problem (4.15) can thus be reduced to

$$\min_{(w, \alpha, \theta) \in \mathbb{W} \times \mathbb{R} \times \mathbb{R}} \left\{ \theta : G_{\beta}^i(\mathbf{w}, \alpha) \leq \theta, i = 1, 2, \dots, l \right\}. \quad (4.17)$$

A straightforward approach for evaluating problem (4.17) is by Monte Carlo simulation. Rockafellar and Uryasev [33] give an approximation of $G_{\beta}(\mathbf{w}, \alpha)$, where Monte Carlo simulation can be used. They write $G_{\beta}(\mathbf{w}, \alpha)$ as

$$\hat{G}_{\beta}(\mathbf{w}, \alpha) = \alpha + \frac{1}{S(1 - \beta)} \sum_{k=1}^S [\tilde{g}(\mathbf{w}, \mathbf{u}_{[k]}) - \alpha]^+, \quad (4.18)$$

where $\mathbf{u}_{[k]}$ is the k^{th} sample vector (again here we give the copula version where $\mathbf{u}_{[k]} = \mathbf{F}(\mathbf{x}_{[k]})$). Thus, using equation (4.18) we can express problem (4.17) for evaluation using Monte Carlo simulations

$$\min_{(w, \alpha, \theta) \in \mathbb{W} \times \mathbb{R} \times \mathbb{R}} \left\{ \theta : \alpha + \frac{1}{S^i(1 - \beta)} \sum_{k=1}^{S^i} [\tilde{g}(\mathbf{w}, \mathbf{u}_{[k]}^i) - \alpha]^+ \leq \theta, i = 1, 2, \dots, l \right\}, \quad (4.19)$$

where $\mathbf{u}_{[k]}^i$ is the k^{th} sample arising from copula $C_i(\cdot)$ of the mixture copula (equation (4.8)). S^i is the size of the sample that arises from $C_i(\cdot)$.

Following Zhu and Fukushima [48], we write the minimisation problem as

$$\text{minimize} \quad \theta \quad (4.20)$$

$$\text{subject to} \quad \mathbf{w} \in \mathbb{W}, \mathbf{v} \in \mathbb{R}^m, \alpha \in \mathbb{R}, \theta \in \mathbb{R} \quad (4.21)$$

$$\alpha + \frac{1}{S^i(1-\beta)}(\mathbf{1}^i)^T \mathbf{v}^i \leq \theta, \quad i = 1, \dots, l \quad (4.22)$$

$$v_k^i \geq \tilde{g}(\mathbf{w}, \mathbf{u}_{[k]}^i) - \alpha, \quad k = 1, \dots, S^i, \quad i = 1, \dots, l \quad (4.23)$$

$$v_k^i \geq 0, \quad k = 1, \dots, S^i, \quad i = 1, \dots, l, \quad (4.24)$$

where $\mathbf{v} = (\mathbf{v}^1; \dots; \mathbf{v}^l) \in \mathbb{R}^m$ with $m = \sum_{i=1}^l S^i$ and $\mathbf{1}^i = (1; \dots; 1) \in \mathbb{R}^{S^i}$.

4.2 Portfolio management under a worst case copula scenario

In this section we demonstrate how the theory in Section 4.1 can be used for the optimisation of a portfolio of financial assets. Financial assets can be described by distributions and their risk can be measured using CVaR.

We consider a portfolio of n financial assets $A_1, \dots, A_n$. We assume that the returns of the assets are log-normally distributed and they are consistent with the Black and Scholes [5] representation, given by

$$x_i = \frac{dA_i(t)}{A_i(t)} = \mu_i dt + \sigma_i dB_i(t), \quad (4.25)$$

where μ_i and σ_i is the mean and the standard deviation of the random variable x_i and $dB_i(t)$ denotes a Wiener process. We have the return vector $\mathbf{x}^T = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\mathbf{u}^T = (u_1, \dots, u_n) = (\Phi_1(x_1), \dots, \Phi_n(x_n))$. We define the decision vector $\mathbf{w}^T = (w_1, \dots, w_n) \in \mathbb{W} \subseteq \mathbb{R}^n$, which denotes the percentage, out of the portfolio's total value, invested in each financial asset. We also define the loss function

$$\tilde{g}(\mathbf{w}, \mathbf{u}) = -\mathbf{w}^T \boldsymbol{\Phi}^{-1}(\mathbf{u}), \quad (4.26)$$

where $\Phi^{-1}(\mathbf{u}) = (\Phi_1^{-1}(u_1), \dots, \Phi_n^{-1}(u_n))^T$. Hence, the loss function is the negative of the portfolio return $\mathbf{w}^T \Phi^{-1}(\mathbf{u})$.

Above, We have selected the univariate distribution that describes the asset returns. We now consider the selection of the copula that describes the dependency between these returns. We first solve a simple optimisation problem using the Gaussian copula. Consider the problem

$$\underset{\mathbf{w} \in \mathbb{W}}{\text{minimize}} \text{ CVaR}(\mathbf{w}), \quad (4.27)$$

where $\mathbb{W}$ defines the domain of $\mathbf{w}$ as described by its constraints and $\mathbf{u} \sim C_{Ga}$ (see equation (4.3)). This is equivalent to equations (4.20)-(4.24) when $l = 1$. As we discussed in the introduction of this chapter, although the Gaussian copula has many nice characteristics, it suffers from modelling deficiencies. Regardless of that, since it is the most widely used copula in portfolio optimisation, it is suitable benchmark portfolio.

We aim to compensate for some of these disadvantages by using a mixture copula. In problem (4.28), the mixture set $\mathcal{C}$ contains the three Archimedian copulas from Section 3.3.3 and the Gaussian copula (see 3.3.1). The aim is to cover all types of dependencies and thereby use a robust measure of dependence (Kendall's τ (3.9)). This robustness is further augmented by the worst-case approach. Thus, the second problem we solve is

$$\underset{\mathbf{w} \in \mathbb{W}}{\text{minimize}} \text{ WCVaR}(\mathbf{w}), \quad (4.28)$$

where $\mathbb{W}$ is defined as above and $\mathbf{u} \sim c$ and $c \in \mathcal{C}$ is a set of copulas (see equation (4.6))

For further comparison we introduce a third and final model. The third model is an extension of Markowitz's original model [25], shown below,

$$\min_{\theta \in \mathbb{R}, \mathbf{w} \in \mathbb{W}} \{\theta : \mathbf{w}^T \Sigma \mathbf{w} \leq \theta\}, \quad (4.29)$$

where Σ , the covariance matrix, is symmetric positive semidefinite. Problem (4.29) con-

stitutes a quadratic convex problem. The extended model takes into consideration the uncertainty that governs Σ . The solution is provided within a worst case framework where Σ is unknown, but is assumed to belong in a set $\mathbb{S}_+$, which comprises symmetric positive semidefinite matrices ¹ [23, 36]. Worst Case Markowitz (WCM) has the form

$$\min_{\theta \in \mathbb{R}, \mathbf{w} \in \mathbb{W}} \sup_{\Sigma \in \mathbb{S}_+} \{\theta : \mathbf{w}^T \Sigma \mathbf{w} \leq \theta\}. \quad (4.30)$$

Given a discrete set $\mathbb{S}_+$, problem (4.30) can be reformulated into the following convex optimisation problem

$$\min_{\theta \in \mathbb{R}, \mathbf{w} \in \mathbb{W}} \{\theta : \mathbf{w}^T \Sigma_i \mathbf{w} \leq \theta, \forall \Sigma_i \in \mathbb{S}_+\}. \quad (4.31)$$

WCM provides us with a nice comparison model since it is an extended, robust version of the well-known classic Markowitz model, and uses Σ to describe the dependence structure, in a similar fashion to problem (4.27).

Problems (4.27) and (4.28) are solved using equations (4.20)-(4.24) and WCM using (4.31). Also, for all three problems we assume $\mathbb{W}$ to be convex and, without loss of generality (and for simplicity), we define it by the following constraints:

$$\mathbf{1}^T \mathbf{w} = 1, \quad (4.32)$$

where $\mathbf{1}$ is the vector with all-unity elements. Furthermore, to assure portfolio diversification, the additional constrains

$$\underline{\mathbf{w}} \leq \mathbf{w} \leq \overline{\mathbf{w}} \quad (4.33)$$

can be imposed, for example in the case of a portfolio comprising only long positions. At this instance we may require $\underline{\mathbf{w}} = 0.1 \cdot \mathbf{1}^n$ and $\overline{\mathbf{w}} = \mathbf{1}^n$, where $\underline{\mathbf{w}}, \overline{\mathbf{w}}$ are lower and upper bounds respectively. This way we prevent the portfolio from having a single asset position. Of course, such constraints are not compulsory and vary according to the investors' preference.

¹The estimation of the covariance matrix from empirical data will usually yield a positive definite matrix. One case where the covariance matrix is positive semidefinite is when is estimated for a large number of assets with sparse returns. Another case is when redundant assets are taken into account.

Finally, since we optimise an asset portfolio we are interested in its performance. Hence, it is often desirable to impose an additional performance restriction in terms of the minimum expected return π_r

$$E(\mathbf{w}^T \mathbf{F}^{-1}(\mathbf{u})) \geq \pi_r. \quad (4.34)$$

4.3 Numerical examples and empirical analysis

In Section 4.1 we derived a convex formulation for WCVaR with copulas. Then, in Section 4.2 we demonstrated how the theory of Section 4.1 can be used to create a robust portfolio of financial assets. Also, in Section 4.2 we introduced two competitive models for portfolio optimisation. In this section we perform back-testing using several datasets to see how our Worst Case Portfolio (WCP) (Problem (4.28)) fares against the other two portfolios, the Gaussian Portfolio (GP) (Problem (4.27)) and the WCM Portfolio (WCMP) (Problem (4.31)). In our comparison tests we also use the Equally Weighted Portfolio (EWP) (equal positions on all assets), since this is a portfolio commonly used by practitioners for benchmarking [11]. The datasets used are introduced in each case considered.

The optimisation problems are solved using the Yalmip Matlab package, together with the CPLEX solver. The PC used for the implementation of the numerical examples is a Intel Core 2 Duo, 2.8GHz with 4 GB of memory. For each of the four copulas, we run 10000 simulations for each of the assets² (Problems (4.27) and (4.28)). Problem (4.31) does not require any simulations.

Simulating data from the Gaussian copula is straightforward with the use of the Cholesky decomposition of the correlation matrix. Simulating data from other copulas can be a difficult task, in general, which makes them less attractive. To simulate data from the three Archimedian copulas we use the algorithms found in Melchiori [27]. Melchiori [27] provides a summary of the results from [12, 26, 30]. For general simulation techniques

²Each dataset contains different number of assets.

regarding copulas see, e.g., [8].

For all our back-testing runs we consider the initial notional value of our portfolio to be $\mathcal{P}_0 = 1$. At every time step t , optimising the portfolio yields a weight vector $\mathbf{w}_t \in \mathbb{R}^n$. The realised return of the portfolio between time t and the next rebalancing time is

$$\mathcal{R}_{t+1} = 1 + \mathbf{w}_t^T \mathbf{r}_{t+1}, \quad (4.35)$$

where $\mathbf{r}_{t+1} = (r_{1,t+1}, \dots, r_{n,t+1})$ is the vector of realised asset returns. The realised return of asset i at time $t + 1$ is defined as $r_{i,t+1} = A_{i,t+1}/A_{i,t} - 1$, where $A_{i,t}$ is the notional value of the asset i at time t . The notional value of the portfolio at time $t + 1$ is

$$\mathcal{P}_{t+1} = \mathcal{P}_0 \prod_{i=1}^{t+1} \mathcal{R}_i. \quad (4.36)$$

Note that it is possible to perform a portfolio rebalancing without re-optimising the weights, i.e. we can have $\mathbf{w}_t = \mathbf{w}_{t+1}$. Particularly in the case of the EWP we have

$$\mathbf{w}_t = \mathbf{w}_{eq} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)^T \quad \forall t. \quad (4.37)$$

For all optimisation problems we define $\underline{\mathbf{w}} = 0$ (see (4.33)).

4.3.1 Fixed estimation window portfolio

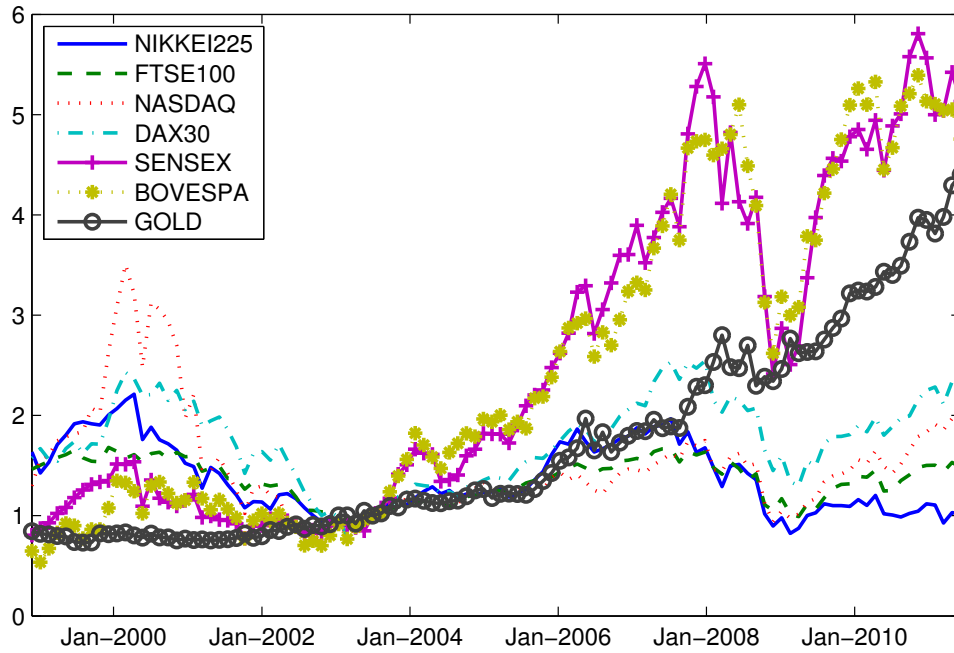
For our first example, we consider a fixed estimation window portfolio in which the weights of the portfolio are calculated only once. A daily rebalancing is performed using the same weights throughout the entire lifespan of the portfolio. The main purpose of this example is to see the general behaviour of the three portfolios, GP, WCMP and WCP, under different levels of π_r (see (4.34)). Since EWP is not dependent on π_r we will omit it from this example.

In this example we use the following seven indices: Nikkei225, FTSE100, Nasdaq, DAX30,

Sensex, Bovespa, Gold index. These represent six different stock exchange markets from different parts of the world and one commodity index. The markets corresponding to the indices are Japan, UK, USA, Germany, India and Brazil. These markets, with the inclusion of the commodity, are intended to lead to a diversified portfolio.

The data used cover the period between November 1998 and July 2011, during which we look at the daily returns. This time line includes the dot-com bubble, South American crisis and Asian crisis. These three events occurred between 1998 and 2002, and they had a large negative impact on the world markets. The data also include the 2008 Global Recession crisis. Both periods of crises can be observed in Figure 4.1.

Figure 4.1: Seven indices from 1998 to 2011. All the indices are normalised to 1 on July 2003



For the computation of the weights we calibrate our copulas using the period between 1998 to 2003 and then solve problems (4.27) and (4.28). To create the set S_+ for WCM we calculate six Σ . In more detail, Σ_1 is calculated from the period between 1998 to 2003. Σ_2 is the diagonal of Σ_1 , with the off-diagonal elements being all equal to zero, to assume independence. Σ_3 is calculated through the relation (3.21) (which holds under normality assumption and in our

case Σ_3 is considerably different to Σ_1). Σ_4 is calculated from a smaller sample concentrated around the peak of the year 2000 bubble, for more extreme effects. Finally, Σ_5 and Σ_6 are calculated from sample periods prior to 2003. The means, variances, and the correlation matrix (used in Gaussian copula) of the seven assets are estimated from the period between 1998 to 2003. Table 4.1 shows the mean, the variance and Σ_1 .

The distribution parameters in Table 4.1 are used for the univariate distributions of the assets as defined by equation (4.25). The univariate distributions together with the four copulas in set $\mathcal{C}$ are used to run Monte Carlo simulations in order to provide the inputs $\mathbf{u}_{[k]}^i$ needed to solve equations (4.20)-(4.24). Problem 4.31 does not require any simulations because it is not a stochastic problem.

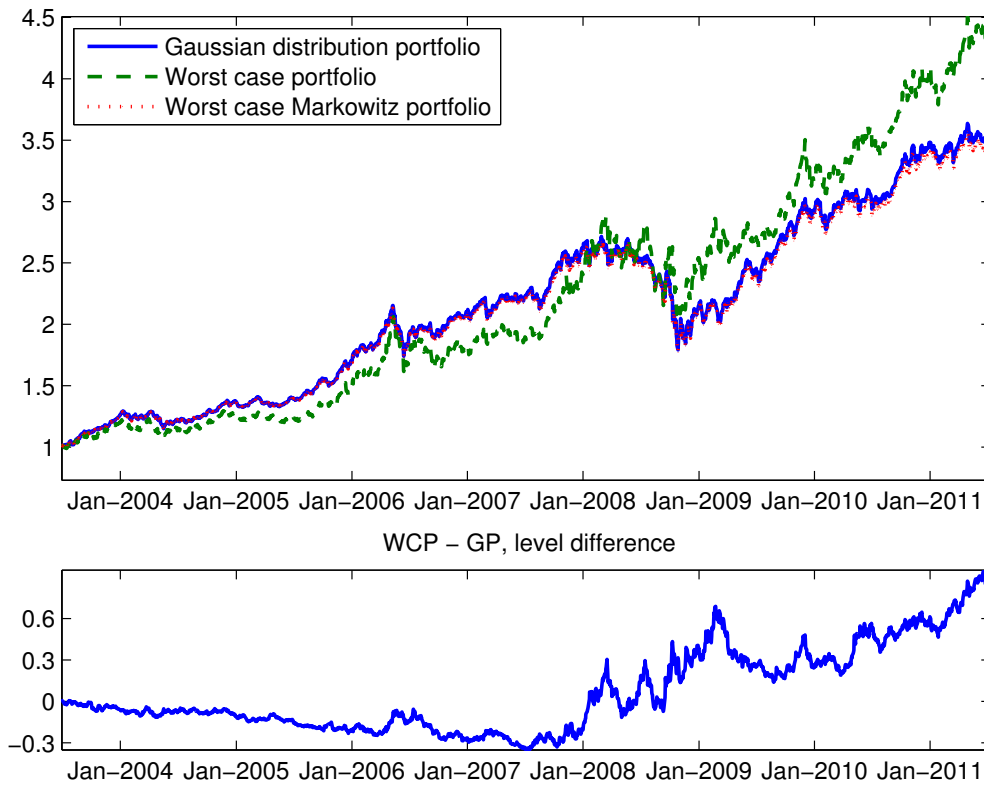
Table 4.1: The mean, variance and covariance of the seven assets daily returns between November 1998 and June 2003

(10 ⁻³)	Mean	Variance		Covariance				
Nikkei225	-0.30	0.22						
FTSE100	-0.22	0.19	0.04					
Nasdaq	0.19	0.82	0.04	0.12				
DAX30	-0.20	0.36	0.04	0.19	0.24			
Sensex	0.31	0.27	0.05	0.03	0.02	0.03		
Bovespa	0.62	0.53	0.02	0.09	0.27	0.14	0.02	
Gold	0.19	0.10	0.01	-0.02	-0.03	-0.03	-0.00	-0.02
			Nikkei225	FTSE100	Nasdaq	DAX30	Sensex	Bovespa

All three problems are solved for $\pi_r = 0$ to 0.0003, where π_r is the required minimum daily expected return (see (4.34)). The results are presented in Figure 4.2, Table 4.2 and Table 4.3. Table 4.2 illustrates the performance of the three portfolios (GP, WCP and WCMP) both for the ‘In Sample’ (November 1998 - June 2003) and ‘Out of Sample’ (June 2003 - July 2011). ‘In Sample’ performance shows that the lower bound π_r is not satisfied for a low expected return, $\pi_r = 0$. This can be verified by AR in Table 4.2. The reason is that the portfolios can satisfy their requirements using the Gold index, which has high expected return and low volatility (see Table 4.1). Hence, the values seen in AR are the results given by the portfolios if constraint (4.34) is not included in the optimisation problem. Thus, for the constraint (4.34) to be binding, π_r has to be bigger than the AR of the unconstrained problem. As a result,

the constraint becomes binding for all portfolios when $\pi_r \geq 0.0002$. Overall (in and out of sample) WCP always has higher volatility (Vol) and CVaR. This is to be expected, since the WCP has to satisfy more constraints in the optimisation problem, i.e., the same constraints that exist for the Gaussian copula in the GP have to be satisfied for all four copulas used in the WCP. Hence, the CVaR obtained for the WCP is the CVaR of the worst case copula, which is the equivalent of the requirement imposed by inequality (4.16). These copulas also have fat tails, and hence higher Vol. On the other hand, WCMP has very similar Vol and CVaR with GP, at least for $\pi_r = 0$ to 0.0002. This similarity no longer holds when higher returns are required.

Figure 4.2: Fixed estimation window portfolios: ‘Out of Sample’ portfolio performance for $\pi_r = 0$ (see equation (4.34))



We expect WCP and WCMP to perform better than the GP, at least under the worst case scenarios. This should apply throughout the out of sample testing period. Hence, we focus on the ‘Out of Sample’ period. Table 4.3 shows that the performance of WCP up to January

Table 4.2: Comparison of the performance of Gaussian optimal, Worst Case optimal and Worst Case Markowitz optimal portfolios

π_r (10^{-3})	GP (I)	In Sample			Out of Sample					
	WCP (II) WCMP (III)	AR ¹ (10^{-3})	Vol ² (10^{-3})	CVaR _{0.95} (10^{-3})	AR ¹ (10^{-3})	Vol ² (10^{-3})	CVaR _{0.95} (10^{-3})	MD1 ³ (%)	MD2 ⁴ (%)	TR ⁵ (%)
0.00	I	0.10	6.51	13.9	0.65	8.89	21.4	-7.27	-32.6	265
	II	0.18	9.36	19.0	0.78	12.2	28.7	-6.97	-26.9	343
	III	0.01	6.53	14.0	0.64	8.86	21.5	-7.39	-33.6	253
0.10	I	0.10	6.51	13.9	0.65	8.89	21.4	-7.27	-32.6	265
	II	0.18	9.36	19.0	0.78	12.2	28.7	-6.97	-26.9	343
	III	0.10	6.53	14.0	0.65	8.89	21.5	-7.34	-33.7	259
0.20	I	0.20	6.86	14.5	0.75	9.43	22.6	-6.77	-33.3	346
	II	0.20	9.19	18.6	0.79	12.0	28.5	-6.85	-27.5	355
	III	0.20	6.86	14.5	0.75	9.42	22.6	-6.79	-33.7	339
0.30	I	0.30	7.80	16.3	0.84	10.1	24.1	-6.35	-36.4	428
	II	0.30	8.90	17.7	0.82	11.1	26.3	-6.61	-32.6	394
	III	0.30	8.52	18.0	0.81	9.85	23.5	-5.83	-39.1	398

1. AR : Average daily return over the period
2. Vol : The volatility defined by the standard deviation
3. MD1 : Maximum draw-down, The worst return between two consecutive days
4. MD2 : Maximum draw-down, The worst return between two days within a period of maximum 6 months
5. TR : The total return from the beginning to the end of the period

2008 (the beginning of the crisis) was always worse than GP with WCMP being almost identical to GP. While π_r increases, the difference between the returns of the WCP and the other two portfolios up to January 2008 increases. This shows that at least, in times of prosperity, an optimistic portfolio performs better.

Table 4.3: Average daily return of Gaussian optimal, Worst Case optimal and Worst Case Markowitz optimal portfolios up to January 2008

π_r		0.00	0.10	0.20	0.30
GP	(10 ⁻³)	0.83	0.83	0.96	1.10
WCP		0.79	0.79	0.82	0.97
WCM		0.83	0.84	0.95	1.11

This behaviour changes from 2008 onwards. In Table 4.2 we see that the performance of GP with respect to Average Return (AR) and Total Return (TR) is worse or similar to WCP for all π_r tested. The reason is the robustness of the performance of the WCP during the 2008 crisis. This change in behaviour from 2008 onwards can also be observed in the bottom Figure 4.2

where we see the difference between the notional value of WCP and GP for $\pi_r = 0$. This difference is quite stable around 0 up to 2008. From 2008 onwards WCP starts to outperform GP. It is surprising that WCMP does not exhibit similar robustness characteristics. On the contrary, its behaviour is similar to the GP. Assessing the risk between the assets through the use of the covariance matrix, similarly to the correlation matrix of problem (4.27), does not carry any information about the characteristics of the tail of the distribution of each asset. Such a model proves inadequate in emulating risk during periods of extreme movements. This is evident from the Maximum Drawdown (MD2). MD2 measures the greatest loss within a 6 month period. For the given dataset this occurs at the time of the crisis. We can see that the WCP MD2 is better than the other two portfolios for all π_r . Another observation from Table 4.2, is that the higher the requirement for π_r is, the more similar the performance of the three portfolios becomes with respect to AR and TR. This is the result of our demand for higher returns, which forces all three portfolios to select high return assets. Thus, the number of assets that can be selected becomes smaller and hence the portfolios become similar. This becomes more apparent in the case of $\pi_r = 0.0003$ where, in addition to the AR and TR, the Vol and CVaR of the three portfolios are much closer.

One inconsistency in the performance of WCP is the fact that Vol and CVaR decrease as π_r increases. The exact opposite occurs in the other two portfolios, which is more in line with our expectations. This inconsistency can be attributed to specific data behaviour and to overall characteristics of the WCP. WCP has the tendency to select the least risky assets on an individual basis, regardless of the correlations between them. Thus, for small π_r , two or three less risky assets that satisfy (4.34) are chosen. This results in a less diversified portfolio which, consequently, more volatile. As π_r increases, the WCP is forced to select a higher number of assets and thus becomes more diversified and less volatile. Someone may think that this implies a riskier portfolio with respect to Vol but, in fact, this does not necessarily translate to higher losses. Higher losses are rather strongly linked to the thickness of the tail of the multivariate distribution, and not so to the volatility.

With the above in mind, we can conclude that an optimistic portfolio, like the GP, is more suitable during times of prosperity. Also, the addition of uncertainty with respect to the values of the covariance matrix, like the WCMP, may improve our assessment of risk during prosperity even further. In contrast, a more robust portfolio like WCP prove to be beneficial during a crisis period.

4.3.2 Rolling estimation window portfolio

We consider the case where the optimal weights of the assets are recomputed every 30 days and then rebalanced daily. In order to calibrate and solve Problems 4.27 and 4.28 we use a moving window of 500 daily returns, i.e. every month we look at the previous 500 daily returns. For Problem 4.31 we use these 500 returns to generate some of the elements of $\mathcal{S}_+$. Σ_1 to Σ_3 are calculated as in Section 4.3.1 using the first 500 returns from the data. We set $\Sigma_4 = 1.5 \cdot \Sigma_1$ and $\Sigma_5 = 1.5 \cdot \Sigma_2$ in order to imply more extreme volatilities and dependencies. Σ_6 is calculated from the moving window of 500 daily returns. Σ_7 is calculated using an expanding window, i.e. at every time step we consider all previous history. The portfolios are rebalanced on a daily basis.

We would like to investigate the adaptability of the three portfolios under changing conditions, using the above setting. This also eliminates the possibility of an one off accurate portfolio selection that may occur in a setting similar to that of Section 4.3.1. We use this setting for all remaining examples.

The moving calibration window forces the estimated values of $\hat{\mu}_i$, used in (4.25), to change every month. Thus, we have to make sure that the constraint (4.34) remains within the feasible set of $\mathbf{w}$. To do so we replace constraint (4.34) with the rolling estimation window constraint

$$E(\mathbf{w}^T \mathbf{F}^{-1}(\mathbf{u})) \geq \mathbf{w}_{eq}^T \hat{\boldsymbol{\mu}}, \quad (4.38)$$

where $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_n)^T$ is the vector of the asset return means, calculated using the calibration

period, and $\mathbf{w}_{eq}$ is defined in (4.37). This way we do not have to individually define (4.34) for each time step. The rest of the constraints remain as in Section 4.3.1.

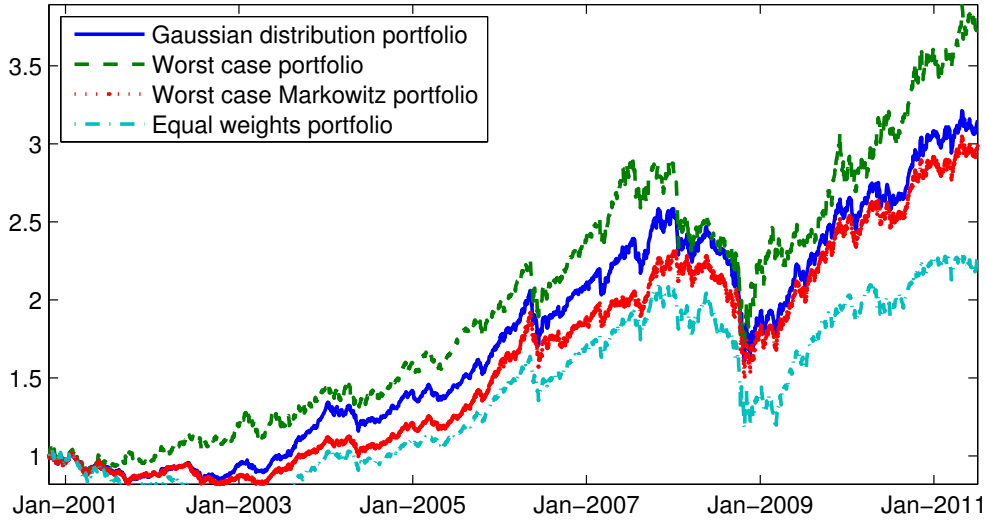
For comparison purposes, we also include a simple portfolio not based on optimisation. The ‘Equally weighted’ portfolio (EWP) has equal positions in all assets, i.e., we always use $\mathbf{w}_{eq}$ as the weights of the portfolio.

The results of this example should not be directly compared to the ones in Section 4.3.1, because here we look into a longer ‘Out of Sample’ period. Using a longer ‘Out of Sample’ window of observation it is possible to observe how the three portfolios behave both in the crises of the period between 1998 and 2003, and 2008. The ‘Out of Sample’ performance of all four portfolios (GP, WCP, WCMP and EWP) is shown in Figure 4.3 and Table 4.4.

The first observation from Figure 4.3, is the difference in the performance between the three optimised portfolios (GP, WCP, WCMP) when compared to the EWP, after their release at the end of 2001. Initially, all four portfolios suffered some losses, but by the end of 2001 all three optimised portfolios stabilised their performance, or in the case of the WCP returned back to gains. The EWP continued to loose value until the end of 2003 and the initial rate of losses, in 2001, was higher than the other three portfolios. Similar observations can be made from Table 4.4 for the 2008 crisis. Looking at the MD2 indicator, we note that the EWP is the worst performing one. Hence, in both periods of crisis, the EWP had the worst performance. In 2008, the WCMP had the smallest depreciation, with the GP and the WCP having very similar performances. After the minimum in October 2008, the WCP managed to outperform the other competitive portfolios.

The overall observations are similar to those of Section 4.3.1. The CVaR and Vol of the WCP are higher than both the GP and the WCMP, but the AR and the TR of the WCP, are significantly better. EWP has the worst performance overall, both with respect to depreciation and returns.

Figure 4.3: Rolling estimation window portfolios performance with $\pi_r = \mathbf{w}_{eq}^T \hat{\mu}$ (see equation (4.38))



4.3.3 Performance analysis

From Sections 4.3.1 and 4.3.2 we have that the performance of the WCP is generally better than the other three portfolios. This is not something that we would expect to happen every time. Since the WCP is a pessimistic portfolio, we would expect that at least in periods of prosperity, a more optimistic portfolio, like the GP, would perform better. In this section, we solve Problems 4.27, 4.28 and 4.31 using different datasets in order to provide more general observations. In particular we focus on areas of dominance, i.e., we investigate which portfolio performs better over intervals of 30 trading days. By identifying the periods where each portfolio is dominant, we can extract conclusions about the underlying characteristics of the three problems. We also look at the worst portfolio depreciation within particular periods. Thus, we can identify the portfolio that is most susceptible to large losses. EWP is still considered for comparison purposes, even though it does not directly encapsulate portfolio characteristics with respect to risk or returns.

In this analysis we optimise the portfolios over three different datasets. In the previous sections we have already used the indices (Indx), which we also include in this section. For

Table 4.4: Comparison of the performance of Gaussian optimal, the Worst Case optimal and Equally Weighted portfolios

	Out of Sample					
	AR ¹ (10 ⁻³)	Vol ² (10 ⁻³)	CVaR _{0.95} (10 ⁻³)	MD1 ³ (%)	MD2 ⁴ (%)	TR ⁵ (%)
GP	0.44	8.10	19.3	-6.07	-34.9	215
WCP	0.53	10.8	24.7	-7.00	-35.1	273
WCMP	0.43	8.40	19.9	-7.37	-33.9	199
EWP	0.34	9.87	22.9	-6.27	-42.9	126
1. AR : Average monthly return over the period 2. Vol : The volatility defined by the standard deviation 3. MD1 : Maximum draw-down, The worst return between two consecutive months 4. MD2 : Maximum draw-down, The worst return between two months within a period of maximum 6 months 5. TR : The total return from the beginning to the end of the period						

our second dataset we take shares from the Mexican stock exchange (Mex), and for the third we take shares from Hong Kong stock exchange (HK). The data for the Mex span from March 2000 to November 2012, and for HK from January 2000 to November 2012. The analytic list of shares can be seen in the figures in Appendix C.1.

Before we proceed on with the monthly performance analysis of the three portfolios we will start with some observations about the general statistics of Mexican and Hong Kong datasets, in a similar manner as Section 4.3.2.

We first start with the Mex. The general results for Mex can be found in Figure 4.4 and Table 4.5. From Figure 4.4 we see that Mex is another example where the WCP under-performs up to 2008 and then outperforms all other three portfolios. Again one of the main reasons is the robust performance of the WCP during the crisis, which by the end of 2008 brings it at level with the other portfolios. From 2009 onwards it looks like it provides a better portfolio choice. We will analyse the performance of the portfolios in more detail later on. Comparing the statistics for Mex in Table 4.5 to the ones of Indx in Table 4.4 we can see a lot of similarities. EWP is the worst performing portfolio with the worst AR, CVaR, MD2 and TR statistics. WCP has higher Vol and CVaR from GP but is the best portfolio of them all with respect to AR, MD2 and TR. The biggest inconsistency that we have between Mex statistics

and Indx is with respect to WCMP. In this instance WCMP has the highest Vol and a very high MD2. We would have not expected a high Vol because WCMP is a worst case volatility minimisation problem (see Problem (4.31)). This may be an indication of the sensitivity of WCMP to the selection of the elements of set $\mathcal{S}_+$. Also, it indicates the inconsistency that may exist between model and realisation.

Figure 4.4: Mexican stock exchange rolling estimation window portfolios performance with $\pi_r = \mathbf{w}_{eq}^T \hat{\mu}$ (see equation (4.38))

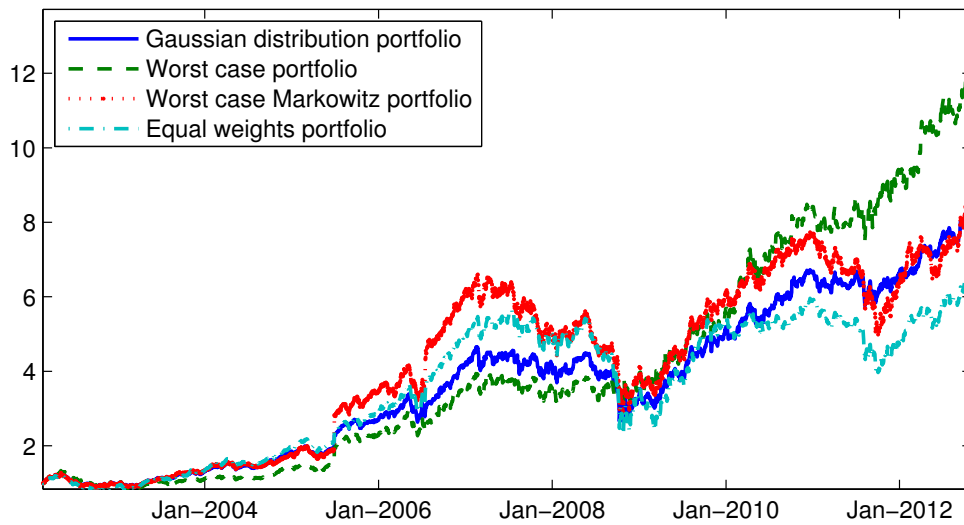


Table 4.5: Comparison of the performance of Gaussian optimal, the Worst Case optimal and Equally Weighted portfolios for Mexican stock exchange.

	Out of Sample					
	AR ¹ (10 ⁻³)	Vol ² (10 ⁻³)	CVaR _{0.95} (10 ⁻³)	MD1 ³ (%)	MD2 ⁴ (%)	TR ⁵ (%)
GP	0.85	14.2	30.8	-10.3	-40.8	722
WCP	1.01	15.8	32.1	-8.55	-34.5	1078
WCMP	0.92	17.9	32.6	-13.2	-48.7	758
EWP	0.80	15.9	35.7	-13.2	-57.8	551

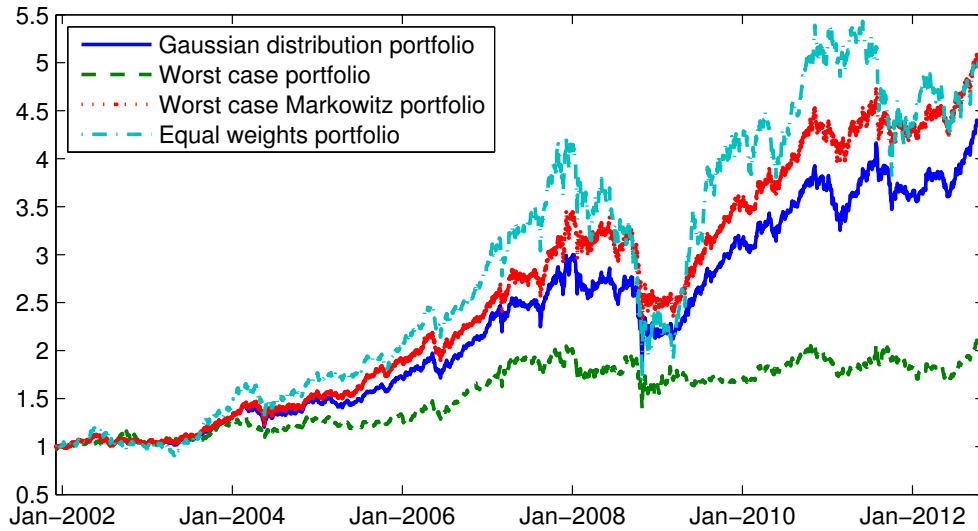
1. AR : Average monthly return over the period
2. Vol : The volatility defined by the standard deviation
3. MD1 : Maximum draw-down, The worst return between two consecutive months
4. MD2 : Maximum draw-down, The worst return between two months within a period of maximum 6 months
5. TR : The total return from the beginning to the end of the period

HK is an example where WCP does not outperform the other portfolios (see Figure 4.5). Also,

from Table 4.6, we see that EWP has the second best performance with respect to returns (see AR and TR). At the same time, EWP remains the most volatile (see Vol, CVaR) and it suffers the worst depreciation among the four portfolios (see MD2). The best TR was obtained by WCMP. WCP, in a similar manner to Indx example (Table 4.4), has higher Vol and CVaR than GP and WCMP but has the best MD2 among all four portfolios. HK provides a good example where a pessimistic portfolio trades high returns for the benefit of robustness.

From the above three examples, Indx, Mex and HK, we can deduce the following general conclusions. EWP can be identified as the riskiest among the portfolios both with respect to volatility and depreciation. WCP is usually more volatile than GP and WCMP but it performs reasonably well under periods of big market movements. GP has relatively low volatility and lies in the middle with respect to returns.

Figure 4.5: Hong Kong stock exchange rolling estimation window portfolios performance with $\pi_r = \mathbf{w}_{eq}^T \hat{\mu}$ (see equation (4.38))



We now move on to the more detailed analysis of the performance of the portfolios. To do so we start with Tables 4.7-4.9, in which we identify the areas where all four portfolios suffered high depreciation. All three examples span the period up to the beginning of 2003 where Latin American, Asian and dot-com crises tailed off as well as the 2008 crisis. Also, in the case of Tables 4.8 and 4.9 we can see the effects of the sovereign crisis in 2011, that followed

Table 4.6: Comparison of the performance of Gaussian optimal, the Worst Case optimal and Equally Weighted portfolios for Hong Kong stock exchange.

	Out of Sample					
	AR ¹ (10 ⁻³)	Vol ² (10 ⁻³)	CVaR _{0.95} (10 ⁻³)	MD1 ³ (%)	MD2 ⁴ (%)	TR ⁵ (%)
GP	0.58	10.4	23.6	-12.2	-36.7	346
WCP	0.34	11.5	25.7	-11.5	-29.4	118
WCMP	0.63	10.4	24.1	-11.7	-37.8	416
EWP	0.68	14.4	33.5	-15.0	-56.0	410
1. AR	: Average monthly return over the period					
2. Vol	: The volatility defined by the standard deviation					
3. MD1	: Maximum draw-down, The worst return between two consecutive months					
4. MD2	: Maximum draw-down, The worst return between two months within a period of maximum 6 months					
5. TR	: The total return from the beginning to the end of the period					

the 2008 crisis. Out of the 12 incidences of high depreciation observed in Tables 4.7-4.9, WCP had in 6 occasions the smallest depreciation, GP in 3, EWP in 2 and WCMP 1. We have only 2 instances where WCP had the worst depreciation, in 2006 (Table 4.8) and in 2010 (Table 4.9). In the case of the former, the depreciation was of similar order of magnitude as the other three portfolios. In general, WCP had the smallest depreciation or had a depreciation of similar orders of magnitude as the best of the other three portfolios. EWP had the most extreme incidences of depreciation, e.g., in 2008 and in 2011 (Table 4.6). It consistently had the worst performance during 2000-2003 and 2008. GP and WCMP gave mixed results.

Table 4.7: Indices rolling estimation window portfolios: Worst periods of depreciation (Highest point/ Lowest point %).

Period	WCP	GP (./.) ¹	WCMP (./.) ¹	EWP (./.) ¹
Nov. 2000 - Sept. 2001	-17.6	-20.4 (-13.8)	-21.0 (-16.3)	-34.7 (-49.4)
May 2006 - Jun. 2006	-15.8	-17.0 (-7.0)	-18.1 (-12.7)	-17.3 (-8.6)
Oct. 2007 - Oct. 2008	-42.3	-37.8 (11.2)	-34.8 (21.4)	-44.6 (-5.2)
1 : Relative difference of WCP depreciation against the other portfolios (WCP/.). A decrease in depreciation indicates better performance.				

Having analysed the performance of the portfolios with respect to depreciation, we now move on to the areas of dominance. EWP is excluded from this analysis, since it does not make use of any information related to the risk or the returns of the assets used in the portfolios.

Table 4.8: Mexican stock exchange rolling estimation window portfolios: Worst periods of depreciation (%).

Period	WCP	GP ($\cdot/\cdot$) ¹	WCMP ($\cdot/\cdot$) ¹	EWP ($\cdot/\cdot$) ¹
May 2002 - Oct. 2002	-33.3	-30.5 (9.3)	-31.6 (5.3)	-37.5 (-11.3)
May 2006 - Jun. 2006	-23.4	-22.3 (4.9)	-22.6 (3.5)	-21.3 (9.8)
Feb. 2007 - Nov. 2008	-22.7	-42.9 (-47.1)	-56.4 (-59.7)	-59.6 (-61.9)
Dec. 2010 - Aug. 2011	-14.1	-14.5 (-3.0)	-35.5 (-60.2)	-33.4 (-57.8)
1 : Relative difference of WCP depreciation against the other portfolios (WCP/.). A decrease in depreciation indicates better performance.				

Table 4.9: Hong Kong stock exchange rolling estimation window portfolios: Worst periods of depreciation (%).

Period	WCP	GP ($\cdot/\cdot$) ¹	WCMP ($\cdot/\cdot$) ¹	EWP ($\cdot/\cdot$) ¹
May 2002 - Apr. 2003	-12.9	-10.2 (26.3)	-10.2 (25.9)	22.6 (-43.0)
Feb 2004 - May 2004	-18.4	-16.5 (11.4)	-16.6 (11.3)	-22.4 (-17.9)
Oct. 2007 - Oct. 2008	-33.1	-39.7 (-16.6)	-38.8 (-14.6)	-60.8 (-45.6)
Nov. 2010 - Feb. 2011	-17.7	-17.1 (3.4)	-12.2 (45.8)	-9.9 (79.5)
Jun. 2011 - Jan. 2012	-8.7	-16.6 (-47.7)	-13.1 (-33.7)	-29.9 (-70.9)
1 : Relative difference of WCP depreciation against the other portfolios (WCP/.). A decrease in depreciation indicates better performance.				

By areas of dominance we imply the 30 trading days intervals, where the compounded returns of one portfolio was greater than the others, i.e. dominated, outperformed the other portfolios. To evaluate the performance of the three portfolios, GP, WCP and WCMP, we use Figures 4.6-4.8 (the figures can be found at the end of the chapter). In Figures 4.6a-4.8a, we highlight the intervals where each portfolio dominated the other two simultaneously. In Figures 4.6b-4.8b, we evaluate the performance of WCP against GP, and in Figures 4.6c-4.8c the performance of WCP against WCMP.

Looking into Figures 4.6a-4.8a we observe that WCP consistently outperformed both GP and WCMP during the period around the beginning of 2009. This can be further verified from Figures 4.6b-4.8c. WCP also outperformed the other two portfolios during the period prior to 2003 for Indx, and mid 2002 for Mex and HK. These periods correspond to the first column of Tables 4.7-4.9. For the case of Mex we have the first half of 2011 and for HK mid 2011 when the market crashed probably because of the sovereign crisis. Looking at the individual cases separately, other areas that WCP was relatively dominant were the period before the

end of 2006 to the mid 2007 for Indx and HK, a period of sudden and big changes in the corresponding markets, prior to the burst of the subprime mortgage bubble at the beginning of 2008.

On the other side of the spectrum the periods where the other two portfolios dominated WCP, at the majority of the 30 day intervals, were 2003-2006 and around mid 2009. The period between 2003 and 2006 was a time of prosperity. Also, 2009 was a year where the markets believed that the worst of 2008 was past and were not aware yet of the beginning of the sovereign crisis.

From the above observations we can conclude that the WCP fares better than the other two portfolios during periods of high volatility and big changes in the market conditions. Also, it underperforms during periods of tranquillity.

4.3.4 Considering transaction cost

Up to this point we did not consider any transaction cost for the rebalancing of the portfolios. In practice, transaction costs should be considered, since big changes in the asset positions of a portfolio can result in transaction costs bigger than the profitability of the portfolio. De Miguel et al. [11] use the above argument to say that because of the transaction cost, it is very difficult to outperform the EWP. To prove this argument, they apply a large number of portfolio optimisation strategies to a large number of datasets and they compare their performance against EWP. Their results validate their statement. The advantages of the EWP are that it does not make any assumption on the risk and returns of the assets, which can be considered as having no systematic error in the modelling perspective, and by having fixed weights, the adjustment after rebalancing only depends on the movements of the assets, implying small changes which imply small transaction costs.

In this section, we look into the *turnover* of each portfolio for all three datasets, Indx, Mex and HK. Using (4.35) and (4.36), we define the difference between existing position of the

portfolio due to market changes and the position given by the weights at time t as

$$\mathcal{DI}_t = \sum_{i=1}^n |\mathcal{P}_{t-1} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t w_{i,t}|, \quad (4.39)$$

and the turnover as the percentage of $\mathcal{DI}_t$ when compared to value of the portfolio at the time,

$$\mathcal{T}_t = \frac{\mathcal{DI}_t}{\mathcal{P}_t}. \quad (4.40)$$

The higher the turnover, the higher the transaction cost that the portfolio incurs at each rebalancing day.

The results in Table 4.10 agree with the ones in [11]. The EWP has the smallest turnover among all four portfolios for all datasets. The portfolio with the second smallest turnover is consistently the WCMP. In two out of the three cases, the WCP has the highest turnover.

Table 4.10: Average turnover when the portfolios are rebalanced on daily basis

Dataset	GP	WCP	WCMP	EWP
Indx	1.30	0.97	1.21	0.89
Mex	1.53	1.81	1.35	1.14
HK	1.80	2.05	1.37	1.17

This raises the following questions. Is it worth considering a portfolios other than the EWP? Does the high turnover of the WCP instate it not worthy of consideration? In Sections 4.3.2 and 4.3.3 we saw that for Indx and Mex all portfolios outperformed the EWP and the difference in the performance was significant. In these two cases, the WCP had the best total return, which was approximately 100% better than the one of EWP, and 26%-50% better than the other two portfolios. In the case of the HK only the WCMP outperformed the EWP. Hence, there are cases where considering the WCP or an alternative portfolio to the EWP seems viable.

In order to analyse the impact of transactions on portfolios, we have to introduce it in our

calculations. We define the portfolio value, after transaction costs, at time t as

$$\mathcal{P}_t^{tr} = \mathcal{P}_{t-1}^{tr} \mathcal{R}_t - cost \cdot \sum_{i=1}^n |\mathcal{P}_{t-1}^{tr} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t^{tr} w_{i,t}|, \quad (4.41)$$

where $\mathcal{R}_t$ and $w_{i,t}$ are as defined in (4.35). Here, we take $cost$ to be proportional to the transaction. Usually $cost$ is of the order 10^{-3} , but this highly depends on how big of a market player the portfolio holder is. The values of $\mathcal{P}_{t-1}^{tr}$ are known from the preceding step, the weights $w_{i,t}$ are obtained by solving an optimisation problem, $cost$ is given and $r_{i,t}$ are the observed realised returns. Hence, the only remaining value to be found is $\mathcal{P}_t^{tr}$. From a quick inspection of (4.41) we see that $\mathcal{P}_t^{tr}$ appears both on the left and right hand side and cannot be separated. Thus, in order to calculate the value of $\mathcal{P}_t^{tr}$, we have to solve the following linear programming problem,

$$\begin{aligned} & \underset{\theta \in \mathbb{R}, \mathcal{P}_t^{tr} \in \mathbb{R}, \mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^n}{\text{minimize}} && \theta \\ & \text{subject to} && \mathcal{P}_{t-1}^{tr} \mathcal{R}_t - \mathcal{P}_t^{tr} - cost \cdot \mathbf{1}^T (\mathbf{a} + \mathbf{b}) \leq \theta \\ & && \mathcal{P}_{t-1}^{tr} \mathcal{R}_t - \mathcal{P}_t^{tr} - cost \cdot \mathbf{1}^T (\mathbf{a} + \mathbf{b}) \geq 0 \\ & && \mathcal{P}_{t-1}^{tr} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t^{tr} w_{i,t} - (a_i - b_i) = 0 \quad i = 1, \dots, n \\ & && \mathbf{a} \geq 0 \\ & && \mathbf{a} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t, \\ & && \mathbf{b} \geq 0 \\ & && \mathbf{b} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t \\ & && \mathcal{P}_t^{tr} \geq 0 \\ & && \mathcal{P}_t^{tr} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t, \end{aligned} \quad (4.42)$$

where $\mathbf{1}$ is the constant vector of ones of size n . The derivation of $\mathcal{P}_t^{tr}$ into a linear programming problem can be found in Appendix C.2.

Since the scope of our research is the theoretical introduction of copulas in robust portfolio optimisation and to understand the characteristics of the model introduced in this chapter, we will not proceed with the implementation of the portfolios with transaction cost. We

leave this as an open research problem for later.

4.4 Summary, future work and conclusions

In this chapter we demonstrate one way of using copulas in a portfolio optimisation framework, where the worst-case copula is considered. In particular, we focus on the derivation of the CVaR and the WCVaR for copulas. In the case of the WCVaR we show how a mixture copula can be used in order to obtain a convex optimisation problem.

By introducing copulas in the CVaR framework we allow more flexibility in the selection of the distribution. The most commonly used distribution for modelling multivariate dependencies is the Gaussian copula. This is due to the simplicity of its construction and the availability of efficient methods of simulation. Its disadvantages are its symmetry and its ability to only describe linear dependencies via the use of linear correlation in its structure. However, symmetric behaviour and linear dependencies among assets are unrealistic [1, 2, 18, 19]. We discuss alternative distribution functions in the form of copulas, that can exhibit asymmetric behaviour and utilise monotonic measures of dependence in their formulation. These are the three Archimedean copulas introduced in Section 3.3.3, from which we can easily simulate data using the algorithms given by Melchiori [27]. Archimedean copulas can also describe tail risk, i.e. risk described at the extremes of the distribution, something that the Gaussian copula lacks.

The advantage of using non-symmetric distribution functions is demonstrated by the numerical examples of Section 4.3. In Section 4.3 we provide a comparison between four competitive models, WCVaR, Gaussian copula CVaR, WCM and equal weights using three different datasets, Indx, Mex and HK.

By applying a fixed estimation window portfolio on Indx, where the weights are optimised only once (Section 4.3.1), we show that for low minimum expected portfolio returns, the

WCP outperforms both the GP and the WCMP in every sector except in the Vol and the CVaR. In particular during the 2008 crisis the WCP performed more robustly than the other two portfolios, and that was true even for a high minimum expected return requirement. This shows that the assumption of symmetry and the use of the correlation or covariance matrix as a measure of dependence provides insufficient information for assessing the risk during periods of crisis. We also find that despite the fact that for higher expected portfolio returns the GP and WCMP outperformed WCP, the margin was not very significant. This stems from the fact that as we require higher expected portfolio returns, the number of assets that satisfy the constraint reduces, which causes the different optimisation problems to make similar asset selections. On the other hand, if the required expected portfolio return is significantly lower than the expected returns of the individual assets, it is possible for the constraint not to be satisfied, thus obtaining the weights corresponding to the unconstrained problem.

In Sections 4.3.2 and 4.3.3 we follow a more rolling estimation window strategy with the incorporation of a moving observation window. In Section 4.3.2 we introduce and apply the strategy on Indx and we find that we get similar results to Section 4.3.1 regardless of the differences in the two strategies. In Section 4.3.3 we apply the rolling estimation window strategy on Mex and HK and we proceed with performance analysis for all three datasets. The performance analysis involves identifying areas of dominance, i.e., intervals during which the compounded returns of a portfolio dominate those of others. We also look at the behaviour of the portfolios during sudden market depreciation. From the three datasets, the WCP was the most profitable in Indx and Mex, and the least profitable was HK. Regardless of the differences in the overall performance of the three datasets, the WCP shared very similar characteristics in all three of them. In fact, the WCP was the dominant portfolio around the periods of 2000-2003 and 2009 crises. It had the best performance around the minima of the crises. Also, in most cases it suffered the smallest depreciation and in those cases that it had not, the depreciation was comparable to that of the best performing portfolio. Generally speaking, the WCP fared better than the other portfolios around areas of big volatility and

sudden changes in market conditions. The periods during which it was most evidently outperformed by the other portfolios were between 2003-2006 and 2009, times of prosperity. This is the behaviour that one would expect from a pessimistic portfolio.

One of the portfolios used was the EWP. This portfolio had consistently the most extreme behaviour with respect to volatility and tail risk. It suffered the worst depreciation in 8 out of the 12 cases that are shown in Tables 4.7-4.6. Particularly, in the cases of Indx and Mex, the EWP was the worst performing portfolio with respect to returns, which were significantly lower than those of the other portfolios (Tables 4.4 and 4.5). Only in the case of HK it came second out of four. The advantage of the EWP however, is the turnover. In Section 4.3.4 we saw that the turnover of the EWP was significantly lower than that of the other three portfolios. This implies that in practice, EWP will pay the smallest amount of transaction fees. WCP was the worst with respect to turnover in two out of the three cases.

It is reasonable to conclude that there are certain characteristics that make WCP attractive. Furthermore, a non risk minimisation portfolio like the EWP is extremely volatile but carries small transaction costs. We have to investigate further how this transaction cost impacts the performance of the WCP.

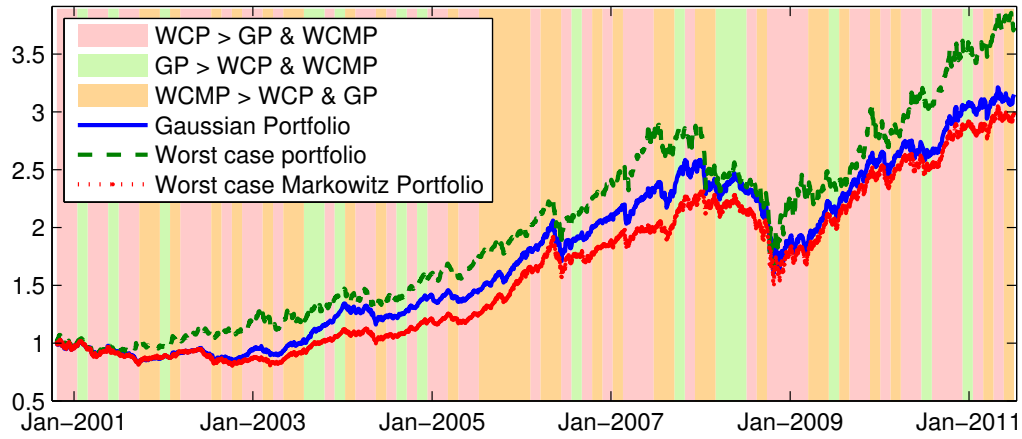
Another topic of investigation is the introduction of a riskless asset in the portfolios. We believe, because of the asset selection process that WCP applies, it will be significantly more robust than the other portfolios. The WCP will clearly show its preference to the non-risky assets, in contrast to the GP and the WCMP which will use correlations in order to offset the risk.

Lastly, we would like to investigate how one can use the information that the WCP provides around crises and incorporate it within a more optimistic strategy. As we saw in the HK, the WCP significantly underperformed compared to the other portfolios but showed good behaviour during crises. Thus, if we want to improve the performance we want to use an optimistic portfolio and somehow make use of the information that the WCP provides in time of crisis. WCP can serve as a "crisis signal indicator".

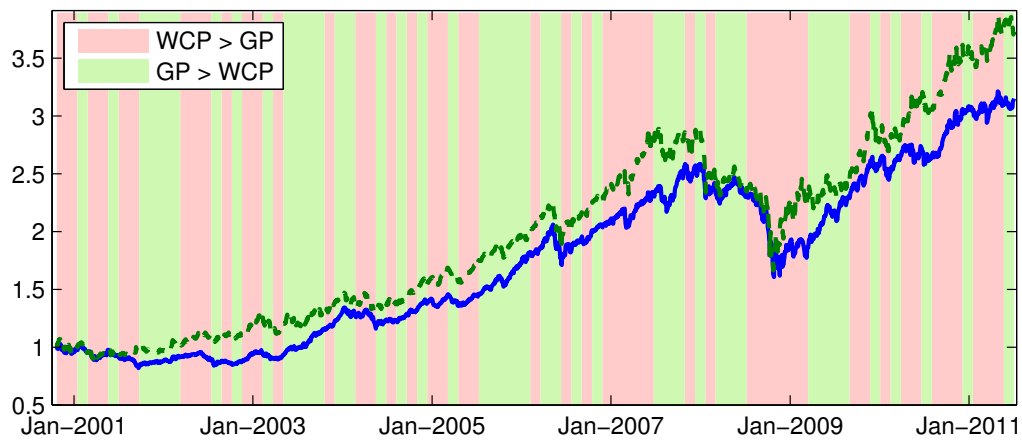
The numerical examples in this chapter, prove that the theory developed can be very useful in practical applications. There is evidence that copulas can be beneficial as modelling tools and in deriving a worst case strategy. Further empirical analysis is required to establish the advantages of the WCP characteristics, and expand the ways that this theory can be used and applied.

Figure 4.6: Indices rolling estimation window portfolios: 30 trading days performance analysis (compounded return).

(a) WCP Vs. All



(b) WCP Vs. GP



(c) WCP Vs. WCMP

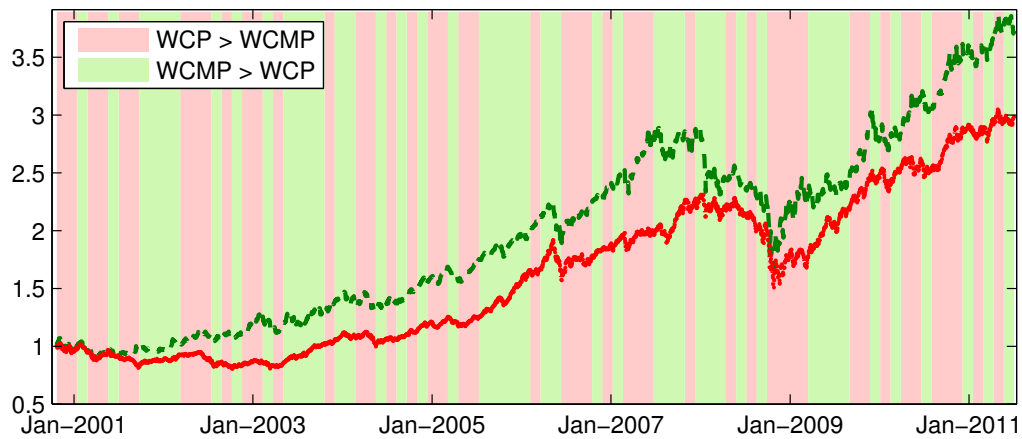
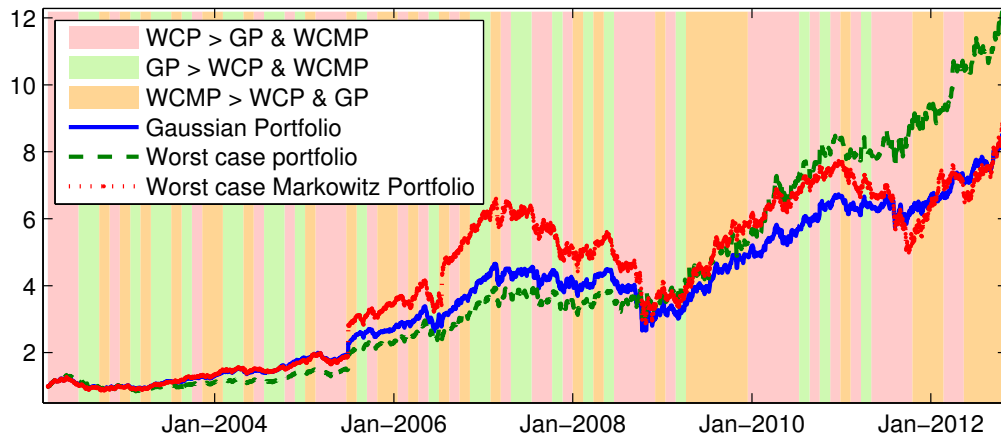
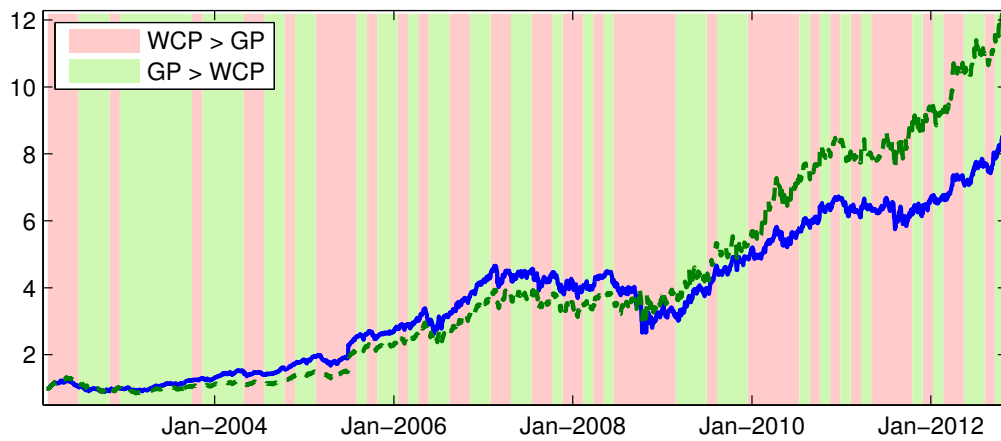


Figure 4.7: Mexican stock exchange rolling estimation window portfolios: 30 trading days performance analysis (compounded return).

(a) WCP Vs. All



(b) WCP Vs. GP



(c) WCP Vs. WCMP

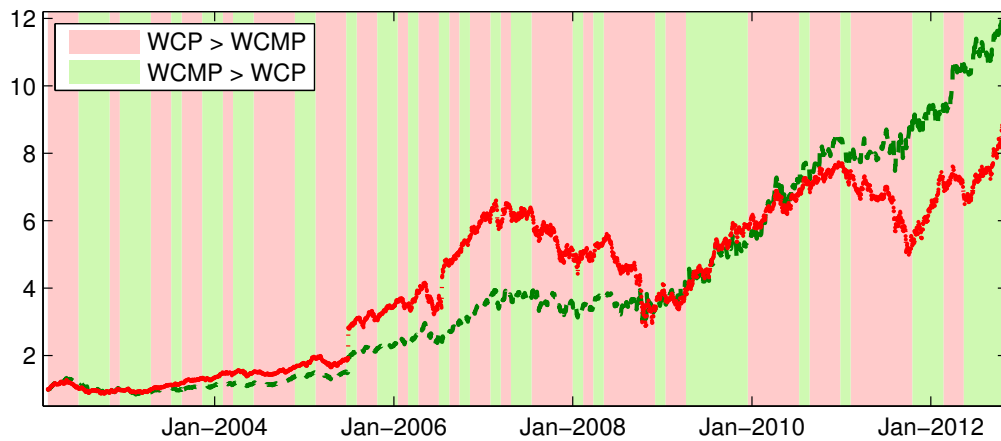
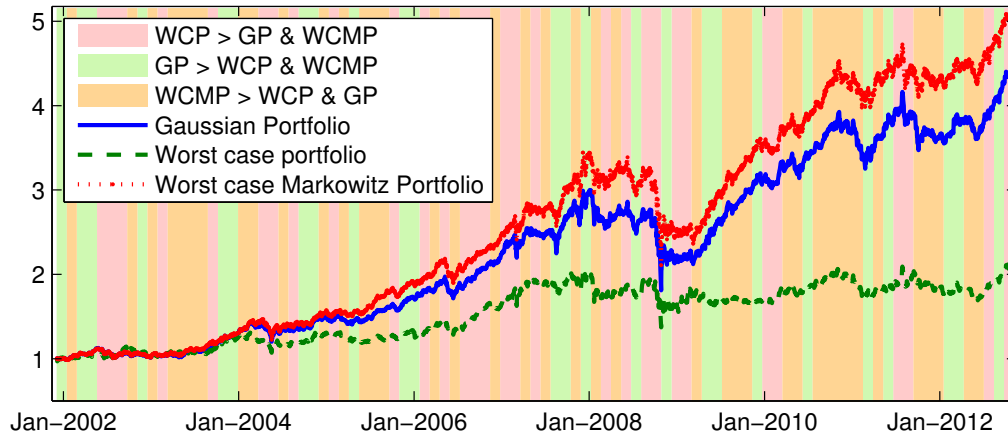
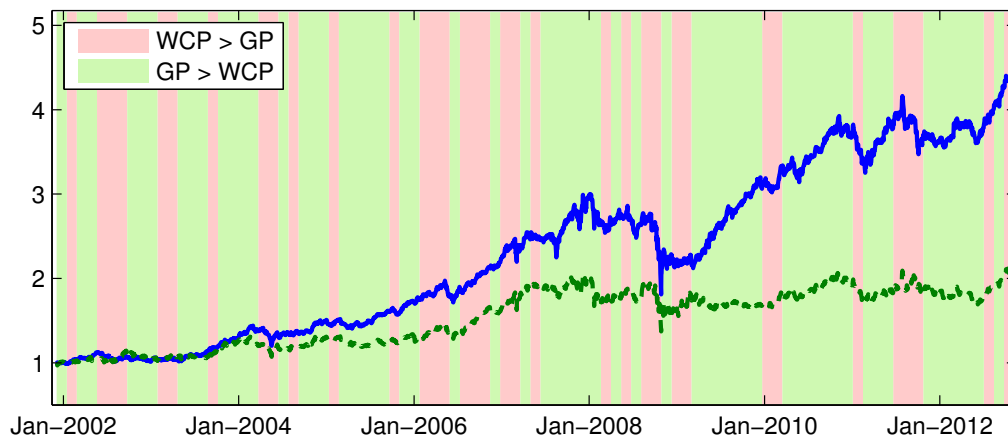


Figure 4.8: Hong Kong rolling estimation window portfolios: 30 trading days performance analysis (compounded return).

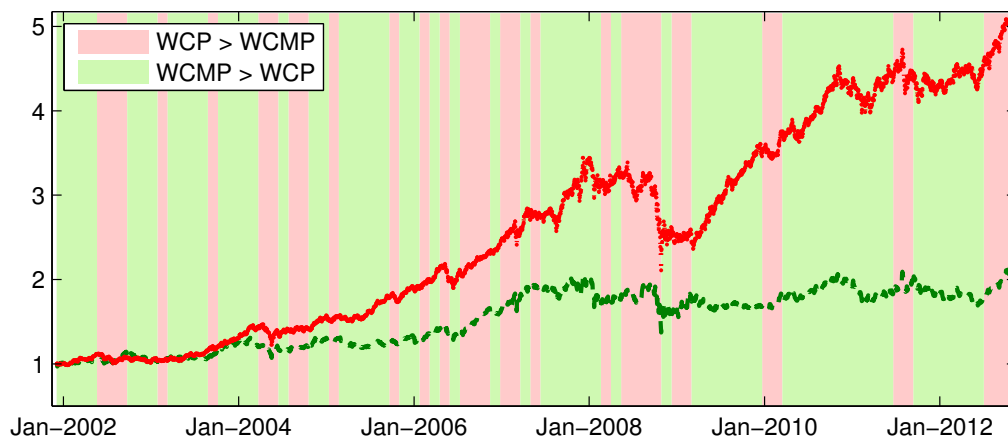
(a) WCP Vs. All



(b) WCP Vs. GP



(c) WCP Vs. WCMP



Chapter 5

Non-symmetric convex uncertainty sets

Making assumptions about the probability distribution of a dataset can introduce systematic errors. Considering an uncertainty set that covers $p\%$ of all possible realisations is a much simpler and a more factual assumption. Hence, we want to use some of the properties of probability theory but not the theory itself.

Ellipsoidal uncertainty (2.11) is an example of transition between assuming a distribution or an uncertainty set. Consider a dataset that is assumed to be described by the multivariate normal distribution,

$$\phi(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right\}. \quad (5.1)$$

Although multivariate normal distribution provides a reasonable description of where the mass of the data lies, it may not accurately represent the actual probability density function of the data. Hence, we prefer to consider the uncertainty set that covers $p\%$ of all possible realisations. Thus, we choose a constant γ such that

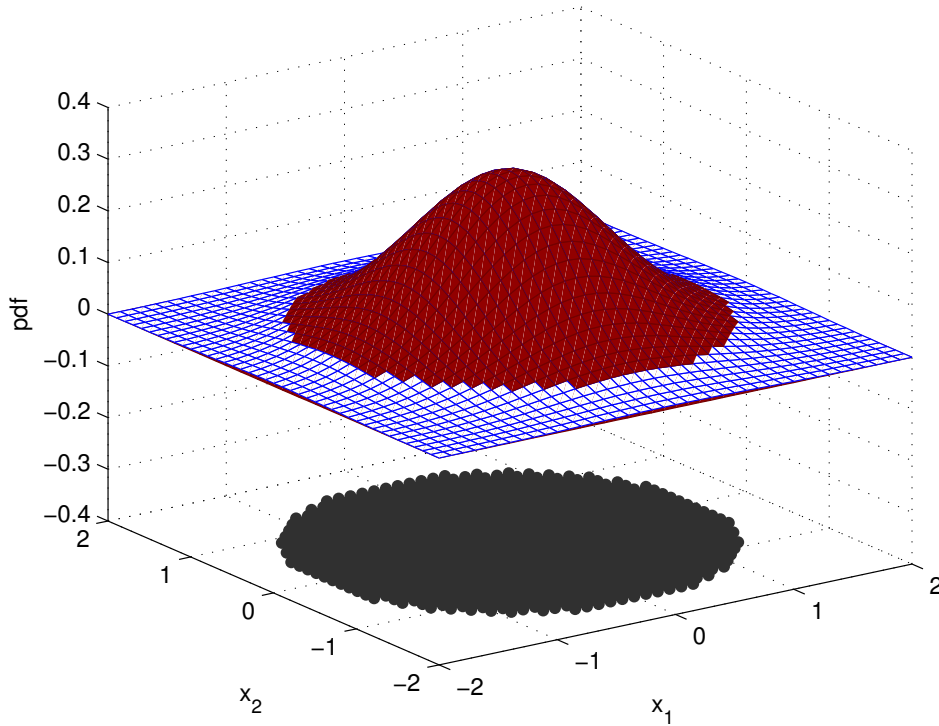
$$\int_{\phi(\mathbf{x}) \geq \gamma} \phi(\mathbf{x}) d\mathbf{x} = p\%. \quad (5.2)$$

Once we have chosen γ , we can derive the uncertainty set by the following series of operations

$$\begin{aligned}
 & \phi(\mathbf{x}) \geq \gamma \\
 \Leftrightarrow & \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\} \geq \gamma \\
 \Leftrightarrow & (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq -2 \log \left\{ \gamma (2\pi)^{n/2} |\Sigma|^{1/2} \right\} \\
 \Leftrightarrow & (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq \log \left\{ \gamma^{-2} (2\pi)^{-n} |\Sigma|^{-1} \right\}.
 \end{aligned} \tag{5.3}$$

We thus obtain the ellipsoidal uncertainty set with $\delta^2 = \log \left\{ \gamma^{-2} (2\pi)^{-n} |\Sigma|^{-1} \right\}$. Hence, set $S_x(\gamma) = \{\mathbf{x} \in \mathbb{R}^n : \phi(\mathbf{x}) \geq \gamma\}$ is the same as (2.11). An illustration of this argument can be found in Figure 5.1, where the part of the probability density function, and the projection of the set $S_x(\gamma) \subseteq \mathbb{R}^2$ that satisfies $\phi(\mathbf{x}) \geq \gamma$, appear in dark red and grey respectively. $S_x(\gamma)$ is our uncertainty set.

Figure 5.1: The dark red area satisfies $\phi(x_1, x_2) \geq \gamma$. The grey area in $\mathbb{R}^2$ represents the corresponding uncertainty set, $S_x(\gamma) = \{\mathbf{x} \in \mathbb{R}^n : \phi(\mathbf{x}) \geq \gamma\}$.



The problem with $\phi(\mathbf{x}) \geq \gamma$ is that, although it defines a convex set, the function $\phi(\cdot)$ is not convex or concave. Hence, $\phi(\cdot)$ cannot be used within a convex optimisation problem as defined in P2.2. This is why we use (2.11) instead, because $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \delta^2$ is a convex function. Then, we can apply duality theory and reformulate the problem into a simpler convex problem (see E2.3.3).

The most common uncertainty sets are the box and ellipsoidal, which are both convex. These sets, however, are very structured. For example, the symmetry that exists in a normal distribution is carried over to the ellipsoidal uncertainty set. In order to alleviate the issue, the same approach, described in (5.2) and (5.3), is used to generate uncertainty sets using copulas. In Section 5.1 we show how convex sets for the case of Clayton copula are created. In Section 5.2 we formulate a robust portfolio optimisation problem using the uncertainty set from Section 5.1. We conclude this chapter with a summary and discussion on future work.

5.1 Defining uncertainty sets through the use of copulas

Continuing in the spirit of portfolio optimisation of financial assets, we are more interested in minimising the losses of our portfolio. Thus, we want to make accurate assumptions about the lower tail risk. A copula that gives more weight on lower tail risk is Clayton (see Definition 3.3.7 and Figure 3.2). Hence, our analysis evolves around Clayton copula.

Take the bivariate Clayton copula,

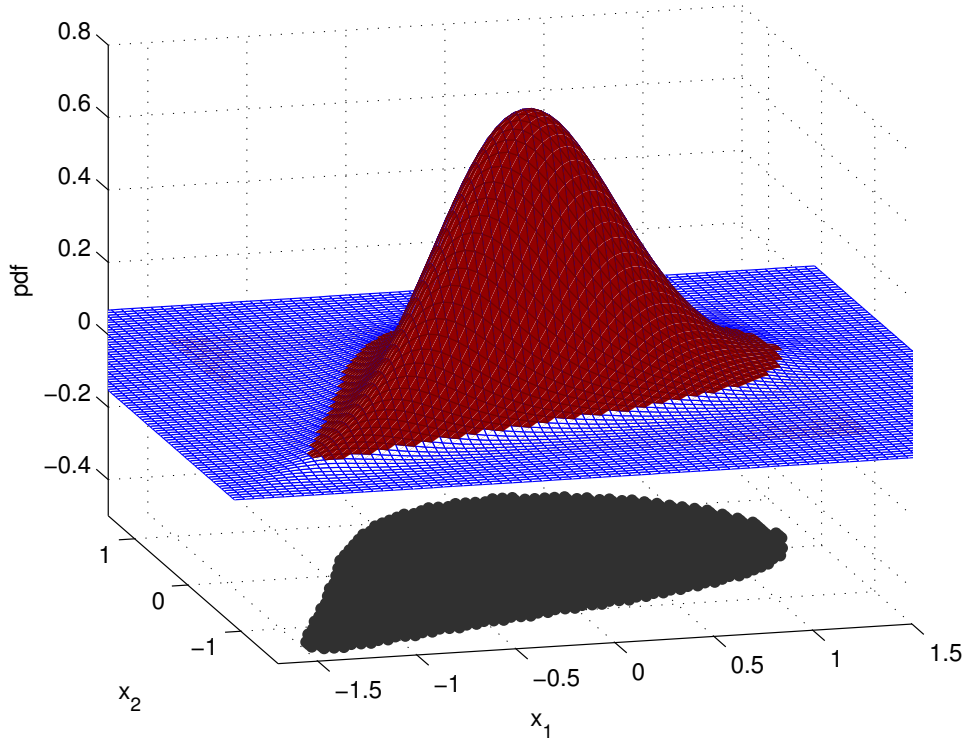
$$C_{Cl}(u_1, u_2) = \max[(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha}, 0], \quad (5.4)$$

with copula density function,

$$c_{Cl}(u_1, u_2) = (1 + \alpha) u_1^{-\alpha-1} u_2^{-\alpha-1} ((u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha})^{-2-1/\alpha}. \quad (5.5)$$

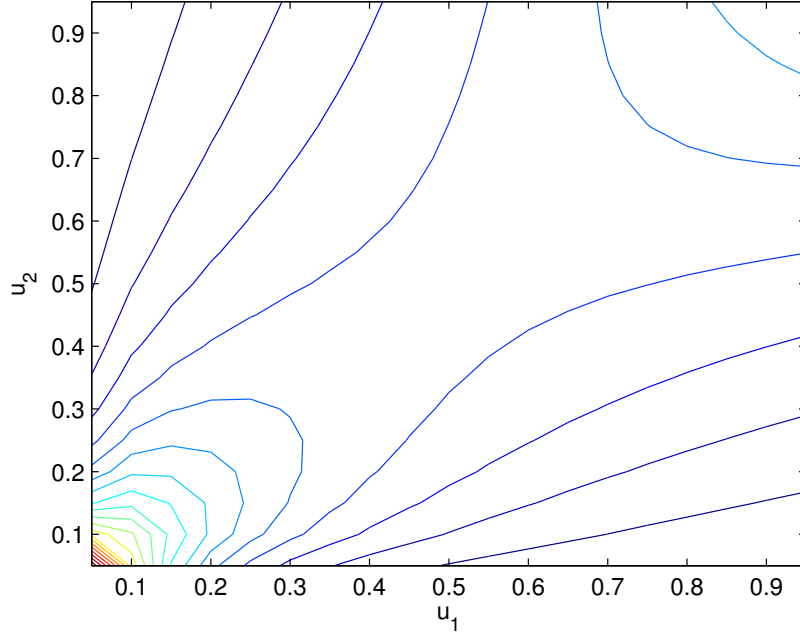
As explained in previous chapters, we can set $u_i = F_i(x_i)$, with $F_i(\cdot)$ being any cumulative distribution function of our choice. Using (3.5) we can obtain the equivalent probability density function $f(\mathbf{x})$. Hence, we can apply the argument illustrated in (5.3) and (5.8) to our newly created $f(\mathbf{x})$ in order to define a set $S_{\mathbf{x}}(\gamma)$. Figure 5.2 illustrates an example where $u_i = \Phi_i(x_i)$. It can be clearly seen that Clayton copulas place emphasis on the lower tail risk of the probability density function, a fact which is also reflected on set $S_{\mathbf{x}}(\gamma)$.

Figure 5.2: The dark red area satisfies $c_{Cl}(\Phi_1(x_1), \Phi_2(x_2))/(f_1(x_1)f_2(x_2)) \geq \gamma$. The grey area in $\mathbb{R}^2$ represents the corresponding uncertainty set.



$S_{\mathbf{x}}(\gamma)$ can be used within a robust optimisation framework. To reformulate the problem into a convex one, we require the inequality that defines $S_{\mathbf{x}}(\gamma)$ to be described by a convex function (see P2.2). We take the simplest case where x_i is uniformly distributed. Hence, $f(\mathbf{x}) = c_{Cl}(\mathbf{x})$. Given $u_i \in \mathbb{I}$, u_i^a is convex for all $a \in \mathbb{R}$. From (5.5) we have that c_{Cl} is the product of convex functions, thus convexity is not guaranteed [6]. We can verify this by a visual inspection of Figure 5.3. The area within each level curve represents $S_{\mathbf{x}}(\gamma)$ for different values γ . The same holds for $u_i = \Phi_i(x_i)$. To make the visual inspection easier, Figure 5.4 shows the contour plot

Figure 5.3: Contour plot of $c_{Cl}(u_1, u_2)$. The different level curves represent different values of γ , (5.3).



of $\log(f(\mathbf{x}))$ ¹.

The question that we want to answer is the following: how can we create a convex set that carries the information from $S_x(\gamma)$ given by the Clayton copula?

5.1.1 Constructing a convex set

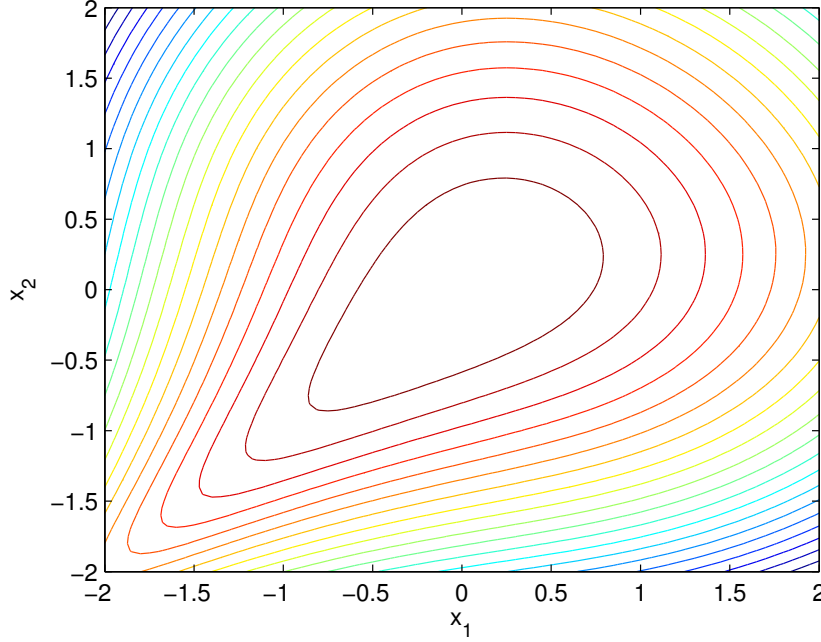
Above we saw that for a constraint like $f(\mathbf{x}) \geq \gamma$ to be usable in a convex setting, we have to define a convex set $S_x(\gamma)$ and it should be possible to be reformulated into an inequality constraint as defined in P2.2. These requirements are given in the following proposition.

Proposition 5.1.1. *If a function $f(\mathbf{x})$ can be described by*

- I. $h(-g(\mathbf{x}))$, where $h(\cdot)$ is a non decreasing function, and $g(\mathbf{x})$ is convex or
- II. $h(g(\mathbf{x}))$, where $h(\cdot)$ is a non increasing function, and $g(\mathbf{x})$ is convex,

¹A logarithmic transformation to a convex function preserves the convexity of the domain.

Figure 5.4: Contour plot of $\log(f(\mathbf{x})) = \log(c_{Cl}(\Phi_1(x_1), \Phi_2(x_2))/(f_1(x_1)f_2(x_2)))$. The different level curves represent different values of $\log(\gamma)$, (5.3).



then $S_x(\gamma) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq \gamma\}$ is a convex set and the inequality can be reformulated to be described by a convex function.

Proof. For both *I* and *II* we start with the inequality.

I.

$$\begin{aligned} h(-g(\mathbf{x})) &\geq \gamma \\ \Leftrightarrow -g(\mathbf{x}) &\geq h^{-1}(\gamma) \\ \Leftrightarrow g(\mathbf{x}) &\leq -h^{-1}(\gamma), \end{aligned} \tag{5.6}$$

where $g(\mathbf{x})$ is convex, hence $g(\mathbf{x}) \leq -h^{-1}(\gamma)$ defines a convex. Hence, the set $S_x(\gamma)$ is convex.

II.

$$\begin{aligned} h(g(\mathbf{x})) &\geq \gamma \\ \Leftrightarrow g(\mathbf{x}) &\leq h^{-1}(\gamma), \end{aligned} \tag{5.7}$$

where $g(\mathbf{x})$ is convex and thus $g(\mathbf{x}) \leq h^{-1}(\gamma)$ defines a convex set. Therefore, the set $S_x(\gamma)$ is convex. $\square$

As we have already seen, the probability density function of Clayton copula does not satisfy the conditions in Proposition 5.1.1. We have already said that we aim at creating a convex set that does not lose information from the original set, or at least carry most of it. Since Clayton copula is used to describe lower tail risk and in portfolio optimisation we are interested in minimising that risk, the information that we want to preserve is contained in the left lower quartile. To do so, we intend to use the copula itself, with the help of the following theorem and corollary.

Theorem 5.1.2. *Let $S_x \subseteq \mathbb{R}^n$ be a compact set. Also, define $\mathbb{P} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq p, p \in \mathbb{R}, \text{dom} f(\mathbf{x}) \supseteq S_x\}$. Then, there exist $p = p_{\mathbb{X}}$ such that $S_x \subseteq \mathbb{P}$.*

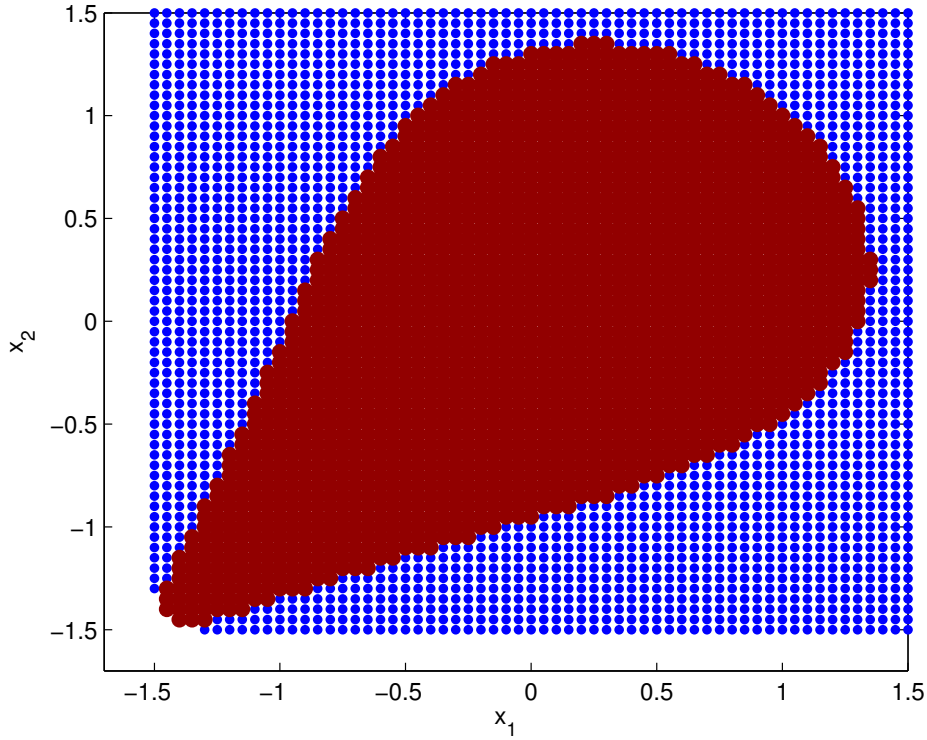
Proof. Since S_x is a compact set we can define $p_{\mathbb{X}} = \min \{f(\mathbf{x}) \in \mathbb{R} : \mathbf{x} \in S_x, \text{dom} f(\mathbf{x}) \supseteq S_x\}$. Let $\mathbb{P}_{\mathbb{X}} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq p_{\mathbb{X}}\}$. Now take $\mathbf{x}_+ \in S_x \cap \mathbf{x}_+ \notin \mathbb{P}$. This implies that $f(\mathbf{x}_+) \leq p_{\mathbb{X}}$. This is a contradiction. Hence, $S_x \subseteq \mathbb{P}_{\mathbb{X}}$. $\square$

Corollary 5.1.3. *Let $S_x(\gamma) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq \gamma, \int_{-\infty}^{\infty} f(\mathbf{x}) d\mathbf{x} = 1, h > 0\}$. Also, define $\mathbb{P} = \{\mathbf{x} \in \mathbb{R}^n : F(\mathbf{x}) \geq p, F(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} f(\mathbf{x}) d\mathbf{x}, p \in \mathbb{I}\}$. Then, there exist $p = p_{\mathbb{X}}$ such that $S_x(\gamma) \subseteq \mathbb{P}$.*

The corollary can be used to bound set $S_x(\gamma)$ with $\mathbb{P}$. As we can see in Figure 5.5, $\mathbb{P}$ preserves the information of the lower extreme of the diagonal. Hence the important information about extreme tail risk is preserved. What we have to find now is whether Clayton copula satisfies Proposition 5.1.1. Let us first begin with the simple case where x_i is uniformly distributed, i.e. $x_i = u_i$. Consider the inequality,

$$\begin{aligned}
 C(\mathbf{u}) &\geq p \\
 \Leftrightarrow (\sum_{i=1}^n u_i^{-\alpha} - n + 1)^{-1/\alpha} &\geq p \\
 \Leftrightarrow \sum_{i=1}^n u_i^{-\alpha} - n + 1 &\leq p^{-\alpha} \\
 \Leftrightarrow \sum_{i=1}^n u_i^{-\alpha} - p^{-\alpha} - n + 1 &\leq 0.
 \end{aligned} \tag{5.8}$$

In order for the left hand side of the last inequality to be convex we need $u_i^{-\alpha}$ to be convex

Figure 5.5: The set $\mathbb{P}$ in blue, bounding $S_x(\gamma)$ in dark red.

[6]. We verify the convexity of $u_i^{-\alpha}$ through the use of second order condition (see (2.5)),

$$\frac{d^2 u^{-\alpha}}{du^2} = -\alpha(-\alpha - 1)u^{-\alpha-2} \geq 0, \quad (5.9)$$

and (5.9) because $u \in \mathbb{I}$. Hence,

$$\mathbb{P}_{Cl} = \{\mathbf{u} \in \mathbb{I}^n : C_{Cl}(\mathbf{u}) \geq p, p \in \mathbb{I}\}$$

is a convex set and it can be equivalently represented as

$$\mathbb{P}_{Cl} = \left\{ \mathbf{u} \in \mathbb{I}^n : \sum_{i=1}^n u_i^{-\alpha} - p^{-\alpha} - n + 1 \leq 0, p \in \mathbb{I} \right\}. \quad (5.10)$$

For the general case where $u_i = F_i(x_i)$, $F_i(\cdot)$ needs to be a concave function, otherwise further assumptions will be required. In the following section we will see examples and possible

formulations within an optimisation problem.

$\mathbb{P}_{CI}$ or any $\mathbb{P}$ using any copula, can be used as a bound for the corresponding S_x or it can be used independently. $\mathbb{P}$ can be considered as an uncertainty set on its own right, where the instances of $P(X \leq \mathbf{x} | \mathbf{x} \text{ satisfies } C(\mathbf{x}) = p) \leq p$ are excluded from the set.

5.2 Formulating a convex optimisation problem

In the previous section we investigated ways for creating uncertainty sets through the use of copulas. Now we want to proceed by utilising these sets within a robust optimisation framework. The optimisation problems in this section are of similar nature to E2.3.2-E2.3.4, where a robust VaR optimisation is performed.

We first start with the general description of the optimisation problem,

$$\begin{aligned} & \underset{\theta \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \theta \\ & \text{subject to} && \sup_{\mathbf{x} \in \mathbb{P}} -\mathbf{w}^T \mathbf{x} \leq \theta. \end{aligned} \tag{P 5.2.1}$$

This problem is practically the same as E2.3.3. Using $\mathbb{P}_{CI}$ with $u_i = x_i$, i.e. the random variable X_i is uniformly distributed, the problem becomes

$$\begin{aligned} & \underset{\theta \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \theta \\ & \text{subject to} && \sup_{\mathbf{u} \in \mathbb{I}^n} -\mathbf{w}^T \mathbf{u} \leq \theta \\ & && \sum_{i=1}^n u_i^{-\alpha} - p^{-\alpha} - n + 1 \leq 0. \end{aligned} \tag{P 5.2.2}$$

Since the constraints in P 5.2.2 are convex, this constitutes a convex minmax optimisation problem. For the general case where $\mathbb{P}_{CI}$ with $u_i = F_i(x_i)$ is used, convexity is not guaranteed.

The general problem formulation is

$$\begin{aligned}
& \underset{\theta \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \theta \\
& \text{subject to} && \sup_{\mathbf{x} \in \mathbb{R}^n} -\mathbf{w}^T \mathbf{x} \leq \theta \\
& && \sum_{i=1}^n F_i(x_i)^{-\alpha} - p^{-\alpha} - n + 1 \leq 0.
\end{aligned} \tag{P 5.2.3}$$

For P 5.2.3 to be convex we need $F_i(x_i)^{-\alpha}$ to be convex. Since $u_i^{-\alpha}$ is convex (see 5.9), $F_i(\cdot)$ has to be concave [6].

In the case where $F_i(x_i)$ is not concave, we can consider one of the following solutions. The simplest approach would be to use a global optimisation solver. Another way would be to incorporate Monte Carlo Importance sampling techniques to populate the uncertainty set $\mathbb{P}_{Cl}$. In detail, we simulate random vector $\mathbf{u}$ from the underlying copula used to define the set, where $u_i, i = 1, \dots, n$ are independent. If the vector $\mathbf{u}$ satisfies $\sum_{i=1}^n u_i^{-\alpha} - p^{-\alpha} - n + 1 \leq 0$ we keep it. The process is repeated until the sample of vectors $\mathbf{u}$ that satisfy the constraint is large enough, e.g. size m . For every vector $\mathbf{u}$ that we keep, we have the corresponding vector $\mathbf{x}$ obtained by $x_i = F_i^{-1}(u_i)$. Then we can reformulate the optimisation problem as,

$$\begin{aligned}
& \underset{\theta \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n}{\text{minimize}} && \theta \\
& \text{subject to} && -\mathbf{w}^T \mathbf{x}_{[k]} \leq \theta \quad k = 1, \dots, m,
\end{aligned} \tag{P 5.2.4}$$

where $\mathbf{x}_{[k]}$ is the k^{th} sample vector. The above methods can, however, prove to be computationally expensive, especially as the dimensions of $\mathbf{x}$ increase. Further investigation is required to improve the tractability of P 5.2.3.

5.3 Summary, future work and conclusions

In this chapter we considered an alternative way of using copulas within a robust optimisation framework. Instead of using the probabilistic properties of copulas, we showed how

they can be used to define uncertainty sets.

The main focus of our analysis was Clayton copula. We used Clayton copula, because of its ability to describe lower tail risk, making it better suited to robust portfolio optimisation. Throughout Section 5.1 we explained how either the copula or the copula density function can be used to describe an uncertainty set. We particularly explained how and under which conditions the copula can be used to approximate the copula density. In the case of Clayton copula we identified the scenarios under which the uncertainty set is convex.

In Section 5.2 we used the uncertainty sets from Section 5.1.1, to formulate a robust VaR minimisation problem. We showed the general formulation and a convex formulation. We finally ended with a brief discussion on possible solutions for non convex problems.

Further investigation is required into the conditions that guarantee convexity of P 5.2.3. Duality theory could also be used to reformulate P 5.2.3 into a simple minimization problem. Non-convexity cases for which P 5.2.4 formulation is used, should be tested for numerical tractability as dimensions increase.

The theory from this chapter should also be extended and tested on other copulas. Simple copulas like the Archimedians from Section 3.3.3, may yield interesting results.

Chapter 6

Summary and Conclusions

In this thesis we discussed *copula applications in robust optimisation*. In particular, we introduced copulas in the robust optimisation framework and evaluated their advantages and their applicability. Chapters 2 and 3 laid the foundations for this study, while Chapters 4 and 5 developed the theory.

Chapter 2 gave a brief account of *optimisation*, and covered the main concepts and definitions. The focus was on the theory of convex optimisation and duality, which were later used in Chapters 4 and 5.

In Chapter 3, we provided an overview of the copula literature. After formally defining copulas and giving the necessary theorems that associate copulas with probability distributions, we continued with the measures of dependence, calibration techniques and the definitions used in the following chapters. Of particular interest were mixture copulas and a special family of copulas called Archimedian. Archimedian copulas have interesting dependence structures (see Sections 3.3.3), some of which place strong emphasis on tail dependence. Mixture copulas enhance the modelling flexibility, and provide means for introducing copulas into the robust optimisation framework. This flexibility and the need to describe tail dependence, particularly in modelling financial assets, was demonstrated through various

examples in Section 3.5.

Our first theoretical contribution was analysed in Chapter 4. Here we showed how mixture copulas can be used to write a convex robust optimisation problem. Firstly, we formally derived the VaR and CVaR of a portfolio using copulas. Then, with the use of mixture copulas we derived a convex WCVaR. In Section 4.2 we described how WCVaR can be used for robust portfolio optimisation. Together with the WCVaR portfolio model we introduced two competitive optimisation models. The first was a simple CVaR minimisation model using the Gaussian copula for describing the returns of the financial assets. The Gaussian copula is the most common model used by practitioners for describing multivariate dependencies. Thus, it can be considered as benchmark for our WCVaR model. The second competitive model, WCM, is a reformulation of the classical Markowitz portfolio optimisation problem, in which the covariance matrix belongs within an uncertainty set. Hence, the portfolios is optimised over the worst case covariance matrix within the given set. We also compared our model with a deterministic portfolio using equal weights, the EWP.

The theory of Chapter 4 was tested, together with the competitive models, in Section 4.3, through a series of numerical examples. For our empirical analysis we used 3 datasets, spanning the period between 2000 and 2013. The three datasets were, Indx, using stock indexes of international stock exchanges, Mex which included shares from the Mexican stock exchange and HK which included shares from Hong Kong stock exchange. We used the datasets to perform back-testing on the competitive models.

At the outset a fixed estimation window portfolio strategy is considered. The weights of the portfolios were optimised only once and then they were rebalanced on a daily basis. The main purpose of this exercise was to assess their behaviour under different minimum required expected returns (see (4.34)). Our findings revealed that given a relatively low returns requirement, none of the portfolios satisfied the constraint. Furthermore, for high minimum expected returns, the portfolios began to behave the same. This was because the number of assets with high enough expected returns satisfying the constraints reduced,

resulting to portfolios with similar positions.

The performance analysis of the competitive portfolios was carried out in four steps. We first looked into the statistics of the portfolios, like CVaR, Vol, average returns etc.. Then, we assessed the performance during periods of strong depreciation. We then proceeded with identifying periods where different portfolios dominated each other. It transpires that although WCP usually had high overall volatility and CVaR, this does not necessarily translate to bad performance. On the contrary, WCP was the most resilient during periods of high volatility and sudden markets defaults. Usually, it had the smallest or among the smallest depreciation and it had a better performance before and after the local minima. The periods during which the WCP was under-performing, compared to the other portfolios, were periods of market tranquillity. EWP was the portfolio with most extreme changes in its value and usually very high Vol and CVaR. The final test for the portfolios involved checking their turnover. Initially, we did not consider transaction costs, which can significantly affect the performance of a portfolio. Hence, by calculating the turnover, we looked at the percentage change in the portfolio's asset positions before and after rebalancing, which is an indicator of how stable the portfolios were with respect to these positions. EWP was consistently the portfolio with the smallest turnover. WCP had the highest turnover in two out of the three examples. We conclude the section by proposing an additional optimisation step, that allows us to take transaction costs into account, but details are left for further investigation.

In Chapter 5 we examined an alternative way of using copulas within a robust optimisation framework. We developed a method which allows us to create uncertainty sets, using probability density functions and copulas. Our main objective was to create convex asymmetric uncertainty sets. The asymmetry was achieved by choosing the right copula. If the copula density function or the corresponding probability density function are asymmetric, then they yield an asymmetric uncertainty set. In order to illustrate our theory, we used the Clayton copula. From there, we supplied some conditions under which the corresponding uncer-

tainty set is convex. Our analysis was concluded by reformulating robust VaR minimisation problems, which make use of the copula defined uncertainty sets.

Future work was discussed at the end of Chapters 4 and 5, where we explained what remains to be done and what are the possible extensions for each Chapter's topic.

Furthermore, despite the fact that two portfolio optimisation problems were examined in this thesis, there are many other application areas where copulas and optimisation theory can be combined. For example, in Section 3.3.2 the shuffles of the Fréchet-Hoeffding upper bound can be used to approximate any copula. Maybe a model fitting optimisation algorithm can be developed around this idea.

We demonstrated that copulas are useful within optimisation. As seen in Chapter 4, several characteristics that copulas possess, such as the modelling flexibility and robustness for example, prove beneficial in modelling uncertainty. Moreover, as demonstrated in Chapter 5, copulas can be utilised in more abstract ways. But there still is a lot of work to be done in order to realise copulas' full potential within optimisation.

Bibliography

- [1] ANG, A., AND CHEN, J. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63 (2002), 443–494.
- [2] ARTZNER, P., DELBAEN, F., ERBER, J., AND HEATH, D. Coherent measures of risk. *Mathematical Finance* 9 (1998), 203–228.
- [3] BEN-TAL, A., AND NEMIROVSKI, A. Robust convex optimization. *Mathematics of Operations Research* 23, 4 (1998), 769–805.
- [4] BEN-TAL, A., AND NEMIROVSKI, A. Robust solutions of uncertain linear programs. *Operational Research Letters* 25 (1999), 1–13.
- [5] BLACK, F., AND SCHOLES, M. The pricing of options and corporate liabilities. *The Journal of Political Economy* 81, 3 (1973), 637–654.
- [6] BOYD, S., AND VANDENBERGHE, L. *Convex Optimization*. Cambridge University Press, 2008.
- [7] CHAN, J. C. C., AND KROESE, D. P. Efficient estimation of large portfolio loss probabilities in t-copula models. *European Journal of Operational Research* 205, 2 (2010), 361 – 367.
- [8] CHERUBINI, U., LUCIANO, E., AND VECCHIATO, W. *Copula Methods in Finance*. Finance Series. John Wiley and Sons, 2004.
- [9] CLEMENTE, A. D., AND ROMANO, C. Measuring and optimizing portfolio credit risk: a copula-based approach. *Economic Notes* 33 (2004), 325–357.

- [10] DANIELSSON, J., JORGENSEN, B., SAMORODNITSKY, G., SARMA, M., AND DE VRIES, C. G. Subadditivity re-examined: the case for value-at-risk. *www.RiskResearch.org* (2005).
- [11] DEMIGUEL, V., GARLAPPI, L., AND UPPAI, R. Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *Review of Financial Studies* 22, 5 (2010), 1915–1953.
- [12] DEVROYE, L. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986.
- [13] EMBRECHTS, P., LINDSKOG, F., AND MCNEIL, A. Modelling dependence with copulas and applications to risk management. Department of Mathematics ETHZ CH-8092 Zurich Switzerland, 2001.
- [14] FAN, K. Minimax theorems. *Proceedings of National Academy of Science* 39, 1 (1953), 42–47.
- [15] GHAOUI, L. E., OKS, M., AND OUSTRY, F. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operational Research* 51, 4 (2003), 543–556.
- [16] HASSELBLAD, V. Estimation of parameters for a mixture of normal distributions. *Technometrics* 8, 3 (1966), 431–444.
- [17] HASSELBLAD, V. Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association* 64, 328 (1969), 1459–1471.
- [18] HU, L. Dependence patterns across financial markets: methods and evidence. Department of Economics Ohio State University, 2002.
- [19] HU, L. Dependence patterns across financial markets: a mixed copula approach. *Applied Financial Economics* 16 (2006), 717–729.
- [20] HULL, J., AND WHITE, A. Pricing interest-rate derivative securities. *The Review of Financial Studies* 3, 4 (1990), 573–592.
- [21] ITO, K. Stochastic integral. *Proceedings of Imperial Academy* 20 (1944), 519–524.

- [22] ITO, K. On stochastic differential equations. *Memoirs, American Mathematical Society* 4 (1951), 1–51.
- [23] KONTOGHIORGHES, E. J., RUSTEM, B., AND SIOKOS, S. *Computational Methods in Decision-Making, Economics and Finance*. Kluwer Academic Publishers, 2002.
- [24] LING, C. H. Representation of associative functions. *Publication Mathematics Debrecen* 12 (1965), 189–212.
- [25] MARKOWITZ, H. Portfolio selection. *The Journal of Finance* 7, 1 (1952), 77–91.
- [26] MARSHALL, A. W., AND OLKIN, I. Families of multivariate distributions. *Journal of the American Statistical Association* 83, 403 (1988), 834–841.
- [27] MELCHIORI, M. R. Tools for sampling multivariate archimedean copulas. Tech. rep., University Nacional del Litoral, Santa Fe, Argentina, 2006.
- [28] MODIGLIANI, F., AND MILLER, M. The cost of capital, corporation finance and the theory of investment. *American Economic Review* 48, 3 (1958), 261–297.
- [29] NELSEN, R. B. *An Introduction to Copulas*, 2 ed. Series in Statistics. Springer, 2006.
- [30] NOLAN, J. P. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston, 2012. In progress, Chapter 1 online at academic2.american.edu/~jpnolan.
- [31] PREKOPA, A. *Stochastic Programming*. Kluwer Academic Publishers, 1995.
- [32] RENYI, A. On measures of dependence. *Acta Mathematica Hungarica* 10, 3-4 (1959), 441–451.
- [33] ROCKAFELLAR, R. T., AND URYASEV, S. Optimization of conditional value-at-risk. *Journal of Risk* 2, 3 (2000).
- [34] ROCKAFELLAR, R. T., AND URYASEV, S. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26, 7 (2002), 1443 – 1471.

- [35] ROMANO, C. Applying copula function to risk management. Part of author's Ph.D. Thesis 'Extreme Value Theory and coherent risk measures: applications to risk management'.
- [36] RUSTEM, B., BECKER, R. G., AND MARTY, W. Robust min-max portfolio strategies for rival forecast and risk scenarios. *Journal of Economics Dynamics & Control* 24 (2000), 1591–1621.
- [37] SANCETTA, A., AND SATCHELL, S. Bernstein approximations to the copula function and portfolio optimization. Tech. rep., Faculty of Economics and Politics, Trinity College, Cambridge, UK, 2001.
- [38] SCHWEIZER, B., AND SKLAR, A. Operations on distribution functions not derivable from operations on random variables. *Studia Mathematica* (1974).
- [39] SCHWEIZER, B., AND SKLAR, A. *Probabilistic Metric Spaces*. North-Holland, 1983.
- [40] SCHWEIZER, B., AND WOLFF, E. F. On non-parametric measures of dependence for random variables. *The Annals of Statistics* 9, 4 (1981), 879–885.
- [41] SHERSTNEV, A. N. On a probabilistic generalization of metric spaces. *Kazanskii Gosudarstvennyi Universitet Uchenye Zapiski* 124 (1964).
- [42] SKLAR, A. Fonctions de repartition a n dimensions et leurs marges. *Pupl. Inst. Statist. Univ. Paris* 8 (1959), 229–231.
- [43] SKLAR, A. Random variables, joint distribution functions, and copulas. *Kybernetika* 9, 6 (1973), 449–460.
- [44] SKLAR, A. Random variables, distribution functions and copulas: a personal look backward and forward. In *Distributions with Fixed Marginals and Related Topics* (1996), vol. 28 of *IMS Lecture Notes - Monograph Series*, pp. 1–14.
- [45] SMILLIE, A. *New copula models in quantitative finance*. PhD thesis, Imperial College of Science, Technology and Medicine University of London, 2007.
- [46] SZEGO, G. Measures of risk. *European Journal of Operational Research* 163 (2005), 5–19.

-
- [47] WOLFF, E. F. N-dimensional measures of dependence. *Stochastica* 4, 3 (1980), 175–188.
- [48] ZHU, S., AND FUKUSHIMA, M. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research* 57, 5 (2009), 1155–1168.
- [49] ZYMLER, S., KUHN, D., AND RUSTEM, B. Worst-case value at risk of nonlinear portfolios. *Managements Science* (2012), 1–12. Article in advance (online).
- [50] ZYMLER, S., KUHN, D., AND RUSTEM, B. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* 137, 1-2 (2013), 167–198.

Appendix A

A.1 Definitions for sets

Definition A.1.1. The affine hull (**aff**) of a set $S \subseteq \mathbb{R}^n$, is the set of all affine combinations using points from set S :

$$\mathbf{aff} S = \{\theta_1 x_1 + \dots + \theta_n x_n : x_1, \dots, x_n \in S, \theta_1 + \dots + \theta_n = 1\}.$$

Definition A.1.2. The relative interior (**relint**) of a set $S \subseteq \mathbb{R}^n$ is defined as

$$\mathbf{relint} S = \{\mathbf{x} \in S : B(\mathbf{x}, r) \cup \mathbf{aff} S \subseteq S, r > 0\},$$

where $B(\mathbf{x}, r) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq r\}$, the ball of radius r and centre $\mathbf{x}$ in the norm $\|\cdot\|$ ($\|\cdot\|$ can be any norm).

Appendix B

B.1 Theorem 3.3.6

Proof. For the proof we use Definition 3.3.4.

1. Symmetry:

$$\begin{aligned} C(u_1, \dots, u_i, \dots, u_j, \dots, u_n) &= \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_i) + \dots + \varphi(u_j) + \dots + \varphi(u_n)) \\ &= \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_j) + \dots + \varphi(u_i) + \dots + \varphi(u_n)) \\ &= C(u_1, \dots, u_j, \dots, u_i, \dots, u_n) \end{aligned}$$

2. Associativity:

$$\begin{aligned} C(C(\mathbf{u}_{-i}, u_i), \mathbf{w}_{-j}) &= \varphi^{-1}(\varphi(w_1) + \dots + \varphi(w_{j-1}) + \varphi(C(\mathbf{u}_{-i}, u_i)) + \varphi(w_{j+1}) + \dots + \varphi(w_n)) \\ &= \varphi^{-1}(\varphi(w_1) + \dots + \varphi(w_{j-1}) + \\ &\quad \varphi(\varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_{i-1}) + \varphi(u_i) + \varphi(u_{i+1}) + \dots + \varphi(u_n))) + \\ &\quad \varphi(w_{j+1}) + \dots + \varphi(w_n)) \\ &= \varphi^{-1}(\varphi(w_1) + \dots + \varphi(w_{j-1}) + \\ &\quad \varphi(u_1) + \dots + \varphi(u_{i-1}) + \varphi(u_i) + \varphi(u_{i+1}) + \dots + \varphi(u_n) + \\ &\quad \varphi(w_{j+1}) + \dots + \varphi(w_n)) \end{aligned}$$

$$\begin{aligned}
&= \varphi^{-1}\left(\varphi(u_1) + \dots + \varphi(u_{i-1}) + \right. \\
&\quad \left. \varphi\left(\varphi^{-1}(\varphi(w_1) + \dots + \varphi(w_{j-1}) + \varphi(u_i) + \varphi(w_{j+1}) + \dots + \varphi(w_n))\right) + \right. \\
&\quad \left. \varphi(u_{i+1}) + \dots + \varphi(u_n)\right) \\
&= \varphi^{-1}\left(\varphi(u_1) + \dots + \varphi(u_{i-1}) + \varphi(C(\mathbf{w}_{-j}, u_i)) + \varphi(u_{i+1}) + \dots + \varphi(u_n)\right) \\
&= C(C(\mathbf{w}_{-j}, u_i), \mathbf{u}_{-i})
\end{aligned}$$

3. Preserving generator properties under scalar multiplication:

Lets consider generator $\varphi_1(u) = \lambda\varphi_2(u)$ where $\varphi_2(u)$ is an Archimedian generator and λ is a positive constant as define in Theorem 3.3.6. Then we have that

$$u = \varphi_1^{-1}(\lambda\varphi_2(u)) = \varphi_2^{-1}\left(\frac{1}{\lambda}\varphi_1(u)\right)$$

Hence, we have

$$\begin{aligned}
C_{\varphi_1}(\mathbf{u}) &= \varphi_1^{-1}(\varphi_1(u_1) + \varphi_1(u_2) + \dots + \varphi_1(u_n)) \\
&= \varphi_2^{-1}\left(\frac{1}{\lambda}(\lambda\varphi_2(u_1) + \lambda\varphi_2(u_2) + \dots + \lambda\varphi_2(u_n))\right) \\
&= \varphi_2^{-1}(\varphi_2(u_1) + \varphi_2(u_2) + \dots + \varphi_2(u_n)) \\
&= C_{\varphi_2}(\mathbf{u}),
\end{aligned}$$

which implies that $\varphi_1(u)$ is an Archimedian generator as well and has the same copula as $\varphi_2(u)$. □

Appendix C

C.1 Shares used from Mexican and Hong Kong stock exchanges

Figure C.1: Mexican stock exchange shares from 2000 to 2011. All the indices are normalized at the beginning of the period

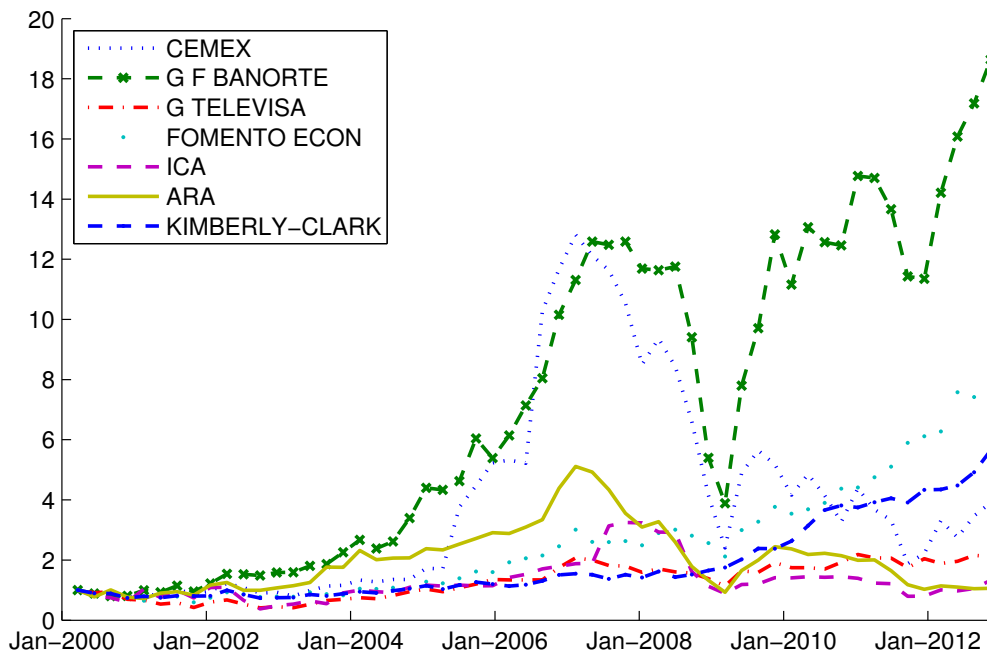
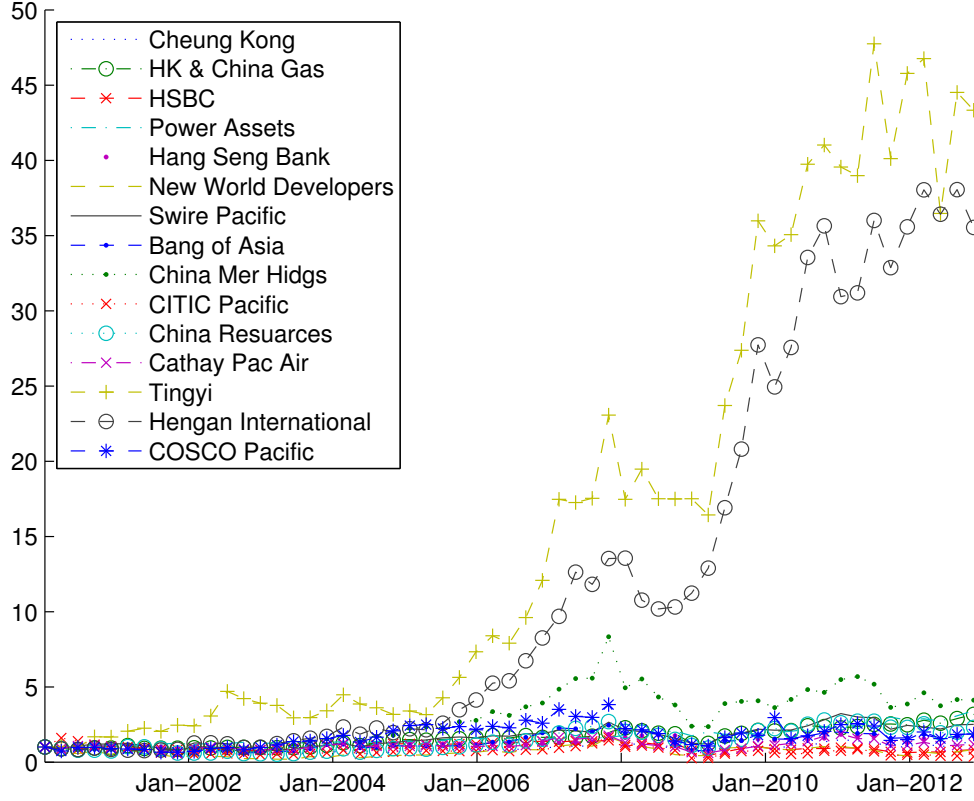


Figure C.2: Hong Kong stock exchange shares from 2000 to 2011. All the indices are normalized at the beginning of the period



C.2 $\mathcal{P}_t^{tr}$ linear programming reformulation

The original problem for the calculating $\mathcal{P}_t^{tr}$ in (4.41) is a rearrangement of (4.41)

$$\begin{aligned}
 & \underset{\mathcal{P}_t^{tr} \in \mathbb{R}}{\text{minimize}} && \mathcal{P}_t^{tr} + cost \cdot \sum_{i=1}^n \left| \mathcal{P}_{t-1}^{tr} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t^{tr} w_{i,t} \right| - \mathcal{P}_{t-1}^{tr} \mathcal{R}_t \\
 & \text{subject to} && \mathcal{P}_t^{tr} + cost \cdot \sum_{i=1}^n \left| \mathcal{P}_{t-1}^{tr} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t^{tr} w_{i,t} \right| - \mathcal{P}_{t-1}^{tr} \mathcal{R}_t \geq 0 \\
 & && \mathcal{P}_t^{tr} \geq 0 \\
 & && \mathcal{P}_t^{tr} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t,
 \end{aligned} \tag{C.1}$$

which is a convex optimisation problem. Because linear programming solvers are more broadly available we want to reformulate it into a linear program. First we start with the use

of the epigraph formulation which makes the objective linear [6],

$$\begin{aligned}
& \underset{\substack{\mathcal{P}_t^{tr} \in \mathbb{R} \theta \in \mathbb{R}}}{\text{minimize}} && \theta \\
& \text{subject to} && \mathcal{P}_t^{tr} + \text{cost} \cdot \sum_{i=1}^n \left| \mathcal{P}_{t-1}^{tr} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t^{tr} w_{i,t} \right| - \mathcal{P}_{t-1}^{tr} \mathcal{R}_t \leq \theta \\
& && \mathcal{P}_t^{tr} + \text{cost} \cdot \sum_{i=1}^n \left| \mathcal{P}_{t-1}^{tr} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t^{tr} w_{i,t} \right| - \mathcal{P}_{t-1}^{tr} \mathcal{R}_t \geq 0 \\
& && \mathcal{P}_t^{tr} \geq 0 \\
& && \mathcal{P}_t^{tr} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t.
\end{aligned} \tag{C.2}$$

Then we introduce the slack variables $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$ in order to replace the sum of absolute values and we obtain the final problem,

$$\begin{aligned}
& \underset{\substack{\theta \in \mathbb{R}, \mathcal{P}_t^{tr} \in \mathbb{R}, \mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^n}}{\text{minimize}} && \theta \\
& \text{subject to} && \mathcal{P}_{t-1}^{tr} \mathcal{R}_t - \mathcal{P}_t^{tr} - \text{cost} \cdot \mathbf{1}^T (\mathbf{a} + \mathbf{b}) \leq \theta \\
& && \mathcal{P}_{t-1}^{tr} \mathcal{R}_t - \mathcal{P}_t^{tr} - \text{cost} \cdot \mathbf{1}^T (\mathbf{a} + \mathbf{b}) \geq 0 \\
& && \mathcal{P}_{t-1}^{tr} w_{i,t-1} (1 + r_{i,t}) - \mathcal{P}_t^{tr} w_{i,t} - (a_i - b_i) = 0 \quad i = 1, \dots, n \\
& && \mathbf{a} \geq 0 \\
& && \mathbf{a} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t, \\
& && \mathbf{b} \geq 0 \\
& && \mathbf{b} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t \\
& && \mathcal{P}_t^{tr} \geq 0 \\
& && \mathcal{P}_t^{tr} \leq \mathcal{P}_{t-1}^{tr} \mathcal{R}_t,
\end{aligned} \tag{C.3}$$

where $\mathbf{1}$ is the constant vector of ones of size n .

Appendix D

Notation and Acronyms

D.1 Notation

$C(\cdot)$: copula function

$c(\cdot)$: copula density function

$F(\cdot)$: cumulative distribution function

$\mathbf{F}(\cdot)$: the vector of cumulative distribution functions, $(F_1(\cdot), \dots, F_n(\cdot))$

$f(\cdot)$: probability density function

$\phi(\cdot)$: Gaussian probability density function

$\Phi(\cdot)$: Gaussian cumulative density function

$\varphi(\cdot)$: Archimedian copula generator

∇ : partial derivative operator $\left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_2}\right)$

$P(\cdot)$: probability

$\mathbf{x}^*$	: optimal solution of optimisation problem
$\mathbf{x}$	: lower case bold letter generally denote a vector
$\mathbf{x}_1, \mathbf{x}_2$	: vector 1, vector 2 etc.
x_i	: the i^{th} element of vector $\mathbf{x}$
$\mathbf{M}$	: upper case bold letters generally denote a matrix
$M_{i,j}$	: the element in the i^{th} row and j^{th} column of matrix $\mathbf{M}$
$\mathbf{x}^T$ or $\mathbf{M}^T$	: the vector $\mathbf{x}$ or matrix $\mathbf{M}$ transposed
$sgn_B(\mathbf{z})$	: signature of vector $\mathbf{z}$ based on n-box B , see Definition 3.1.2
S	: a general set
$\mathbb{R}$	: the set of real numbers, $(-\infty, \infty)$
$\mathbb{R}^n$	: the set of real numbers in n dimensions
$\mathbb{I}$	: the set of real numbers between and including 0 and 1, $[0, 1]$
$\mathcal{C}$	: set of copulas
$\mathbb{W}$	: set of decision/ weight vector $\mathbf{w}$
$\mathbb{N}$	: set of natural numbers
$\mathcal{P}_t$	: portfolio value at time t
$\mathcal{P}_t^{tr}$	: portfolio value at time t when transaction cost is considered
$\mathcal{R}_t$	: portfolio return at time t
$\mathcal{DI}_t$	: the difference between existing positions of the portfolio due to market changes and the positions implied by the weights at time t
$\mathcal{T}_t$	: portfolio turnover at time t
μ	: the mean of a random variable
Σ	: the covariance matrix
$\sum_{i=0}^n a_i$	: the sum of a series $\{a_i\}_{i=0}^n$ from 0 to n
$\prod_{i=0}^n a_i$	: the product of a series $\{a_i\}_{i=0}^n$ from 0 to n

D.2 Acronyms

dom	: domain of a function
ran	: the range of a function
relint	: the relative interior of a set
KKT	: Karush-Kuhn-Tucker conditions, see (2.9)
VaR	: Value at Risk
CVaR	: Conditional Value at Risk
WCVaR	: Worst Case Conditional Value at Risk
GP	: Gaussian Portfolio
WCP	: Worst Case Portfolio
WCM	: Worst Case Markowitz
WCMP	: Worst Case Markowitz Portfolio
EWP	: Equally Weighted Portfolio
Mex	: Dataset using shares from the Mexican stock exchange
Ind	: Dataset using Indexes from International stock exchanges
HS	: Dataset using shares from Hong Kong stock exchange
MLE	: Maximum Likelihood Estimator
IFM	: Inference Function for Marginals
CML	: Canonical Maximum Likelihood