Essays on optimal portfolio choice and model risk assessment

De Vecchi, Corrado

Publication date:
2022

License:
Other

Document Version:
Final published version

Link to publication

Citation for published version (APA):
De Vecchi, C. (2022). *Essays on optimal portfolio choice and model risk assessment*.

FACULTY OF SOCIAL SCIENCES
AND SOLVAY BUSINESS SCHOOL

Essays on Optimal Portfolio Choice and Model Risk Assessment

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Economics
at Vrije Universiteit Brussel

Corrado De Vecchi

2022

Doctoral Committee

Advisors	Prof. dr. Carole Bernard	Vrije Universiteit Brussel
	Prof. dr. Steven Vanduffel	Vrije Universiteit Brussel
Members	Prof. dr. Valeria Bignozzi	Università degli Studi di Milano-Bicocca
	Prof. dr. Lieven De Moor	Vrije Universiteit Brussel
	Prof. dr. Julien Trufin	Université libre de Bruxelles
Chair	Prof. dr. Kris Boudt	Vrije Universiteit Brussel

Acknowledgements

I would like to thank all the people who have helped and supported me throughout my PhD.

First, I want to express my gratitude toward my PhD supervisors, Carole Bernard and Steven Vanduffel. It has been a real privilege to learn from them during these years. I wish to thank them not only for their skilful guidance through the most technical aspects of my research activity, and for their constant optimism, but also for the openness and interest they always showed toward my point of view. Having two PhD supervisors that encourage a student to express his own opinions and ideas is not common and it has been a strong source of motivation that significantly helped me to grow as a researcher.

A big thank goes to the colleagues of the Vrije Universiteit Brussel, for the discussions and the companionship that have been an important part of my doctoral path.

I will always be grateful to my parents, Sergio and Luisa, and to my brother, Pietro, for their strong encouragement during all these years. I would have not been able to reach this point without them.

My warmest appreciation goes to all the friends that helped during this journey. In particular, to my long-time friends, Giorgio and Riccardo, for always being there in the bad and in the good times, to the friends from the Vincent's house, Hakan, Despina, Anna and Sara, because meeting them has made Brussels a great place to live, and to my Ph.D mate Rodrigue, for always being more than a colleague.

Last, but indeed not least, I am deeply grateful to my girlfriend Chiara, for her unconditional support that helped me to overcome the toughest moments of my PhD, and for all the cherished moments we spent together. I wish to dedicate this thesis to her.

Contents

1 Introduction **1**

1.1 This work . . . 3

1.2 Related publications . . . 4

2 When do two- or three-fund separation theorems hold? **7**

2.1 Introduction . . . 7

2.2 Problem formulation and preferences . . . 9

2.3 Two-fund separation theorems . . . 10

2.4 Three-fund separation theorems . . . 15

2.5 Application . . . 23

2.6 Final remarks . . . 28

2.7 Appendix . . . 30

3 The impact of correlation on (Range)Value-at-Risk **35**

3.1 Introduction . . . 35

3.2 Risk bounds with two risks . . . 38

3.3 RVaR risk bounds with n risks . . . 57

3.4 Final remarks . . . 69

3.5 Appendix . . . 70

4 Robust assessment of life insurance products **77**

4.1 Introduction 77
4.2 Problem formulation 78
4.3 Computing premium bounds 82
4.4 Examples of robust assessment for annuities 87
4.5 Unimodality constraint 91
4.6 Comparison with Pichler (2014) 93
4.7 Final remarks 96
4.8 Appendix 97

5 Discussion and outlook **105**

Bibliography **109**

Chapter 1

Introduction

This thesis is a collection of three essays that answer questions related to Optimal Portfolio Choice and Model Risk Assessment, two topics that have long-standing tradition in the financial and actuarial literature.

The main contribution of this thesis related to Optimal Portfolio Choice is to derive general sufficient conditions leading to a two- or a three-fund separation for the investors' optimal portfolio. Consider a static portfolio choice problem in a market with a risk-free asset and n risky assets having random returns. Given a certain investor, let $\boldsymbol{\omega}^* \in \mathbb{R}^n$ be the vector that describes the optimal amount that she should invest in each risky asset. The two-fund separation theorem holds if there exists a vector $\boldsymbol{\omega}_M \in \mathbb{R}^n$, whose composition depends solely on the joint distribution of the assets returns, such that for each investor we can write

$$\boldsymbol{\omega}^* = \alpha\, \boldsymbol{\omega}_M, \text{ for some } \alpha \in \mathbb{R}.$$

When this is the case, each investor's optimal choice is to put part of the initial budget in the market fund described by $\boldsymbol{\omega}_M$ and the rest of the initial budget in the risk-free asset. The three-fund separation theorem holds if there exist two vectors $\boldsymbol{\omega}_{M_1}, \boldsymbol{\omega}_{M_2} \in \mathbb{R}^n$ whose composition depends solely on the joint distribution of the assets returns, such that for each investor we can write

$$\boldsymbol{\omega}^* = \alpha_1\, \boldsymbol{\omega}_{M_1} + \alpha_2\, \boldsymbol{\omega}_{M_2}, \text{ for some } \alpha_1, \alpha_2 \in \mathbb{R}.$$

In other words, under the three-fund separation theorem, the optimal choice for any investor is to allocate part of the initial wealth in the market funds described by $\boldsymbol{\omega}_{M_1}$ and $\boldsymbol{\omega}_{M_2}$, while the residual initial budget is allocated in the risk-free asset.

The relevance of the fund theorems in the economic literature is justified by the two following observations. First, the fund separation theorems significantly simplify the optimal portfolio selection problem. Under two-fund resp. three-fund separation, the original n-dimensional portfolio optimization boils down to a one-dimensional resp. two-dimensional problem. Second, the fund separation theorems are intimately connected with the mean-variance efficient portfolio, and possible deviations from it. In fact, the characterization of mean-variance efficient portfolios proposed in Markowitz (1952) and Tobin (1958) is the first result obtained in the literature concerning the fund separation theorems. Further important references in this stream of literature include Owen and Rabinovitch (1983), Chamberlain (1983), Levy and Levy (2004), Mencía and Sentana (2009), Pirvu and Schulze (2012) and Birge and Chavez-Bedoya (2016). More details on these and other references related to the fund separation theorems can be found in the introductory section of Chapter 2.

The second research area in which the present thesis gives a contribution is Model Risk Assessment. Many business decisions are based on a law invariant functional that assigns a number to each probability distribution. Standard examples are the amount of capital that needs to be allocated to a risky portfolio (Embrechts et al. (2013)) or the net premium that an insurance company sets for an annuity (Dickson et al. (2009)). These two examples will receive a specific treatment in Chapter 3 and Chapter 4, respectively.

The underlying probability distribution is usually estimated starting from certain model assumptions. Therefore, making the wrong model assumptions implies that the value of any law invariant functional of interest is miscalculated. This observation justifies a current tendency to study what happen when some model assumptions are weakened. In most cases, having weaker model assumptions implies that the probability distribution of interest is not completely specified, but instead it is only known to belong to a certain family of probability distributions. Model Risk Assessment studies what are the best- and worst-case scenarios for the quantity of interest, when the underlying probability distribution belongs to a certain set of probability distributions. In the literature related to Model Risk Assessment, the best- and worst-case scenarios are sometimes called risk bounds. Equivalently, the knowledge of the best- and worst-case scenarios allow to obtain an assessment for the quantity of interest that is robust with respect to changes in the underlying model assumptions. Thus, the output of this study can be used to make well-informed business decisions that do not rely too heavily on the specific model considered.

Historically, early results in this field of research can be dated back at least to the seminal article Cantelli (1910). In this paper, the author points out that the probability bounds studied for example by P. Tchebichef and I.-J. Bienaymé could be of interest for the development of risk theory with insurance applications, and not solely from a purely mathematical point of view. Other notable historical results in the literature related to Model Risk Assessment include, but are not limited to, Makarov (1981) and Rüschendorf (1982), that dealt with risk aggregation under dependence uncertainty, and Jansen et al. (1986), Kaas and Goovaerts (1986), Hürliman (1998), Hürlimann (2002) and De Schepper and Heijnen (2010) that studied risk bounds in those situations in which only some moments of a certain probability distribution of interest are given. For more recent results concerning risk bounds based on the moments of a distribution, we cite Bernard et al. (2018b), Cornilly et al. (2018) and Bernard et al. (2020a). An additional literature review can be found in the introductory sections of Chapter 3 and Chapter 4.

Recent events have motivated further studies focused on the management of model risk. For instance, according to Salmon (2009), the over-reliance on the David Li formula to price credit derivates was one of the events that trigged the 2008 financial crisis. This formula assumes a Gaussian dependence structure among the assets of interest, and thus it is not able to properly describe the tail-dependence that is sometimes observed in the financial markets.

The relevance of Model Risk Assessment in the modern financial and insurance regulation emerges also from several documents. As for the banking sector, a clear statement is made in Board of Governors of the Federal Reserve System (2011):

"*The expanding use of models in all aspects of banking reflects the extent to which models can improve business decisions, but models also come with costs. There is the direct cost of devoting resources to develop and implement models properly. There are also the potential indirect costs of relying on models, such as the possible adverse consequences (including financial loss) of decisions based on models that are incorrect or misused. Those consequences should be addressed by active management of model risk.*"

For considerations concerning the insurance context, there is a growing effort toward the development of practical procedures that can be used by actuaries to identify and manage the impact of model risk. See, e.g., American Academy of Actuaries (2019), Black et al. (2018) and The Deloitte Center for Regulatory Strategy (2018).

For some concrete examples of documents that underline the role of model risk in the context of risk aggregation, the interested reader can consult Embrechts et al. (2014) in which the authors illustrate how the study of risk aggregation problems under dependence uncertainty can offer a contribution to the debate regarding the choice of the risk measures that should be considered for the computation of solvency capital requirements. As for the life insurance pricing, EIOPA (2016) highlights that any life insurance evaluation is intrinsically exposed to model risk. For example, referring to the well-know Lee-Carter model, EIOPA (2016) state the following.

"*It should be noted that the model does neither take into account uncertainty with respect to parameters nor with regard to the model. The future deviations from the best estimate may be larger or smaller because mortality trends may occur which cannot be predicted at present. These include for instance the effects on future mortality of changes in behavioural factors, socio-economic developments and developments in ethics. Which unknown viruses and bacteria may still have an effect on mortality? How will the resistance of antibiotics develop and what medical developments can be expected? All these factors may result in a situation where the future distribution of mortality around the best estimate may differ from the distribution on the basis of historic data calculated in accordance with the model.*"

1.1 This work

The second Chapter of this thesis deals with Optimal Portfolio Choice and derives sufficient conditions on the joint distribution of asset returns such that a two- or three-fund theorem holds. In Chapter 2, we show that when asset returns satisfy a location-scale property (possibly conditionally as e.g., for a multivariate generalized hyperbolic distribution) and the investor has law-invariant and increasing preferences, the optimal investment portfolio always exhibits two-fund or three-fund separation. As a consequence, we recover many of the three-fund (and two-fund) separation theorems that have been derived in the literature under very specific assumptions on preferences or distributions. These are thus merely special cases from the general characterization result of optimal portfolios that we provide. Moreover, we illustrate how having a two- or a three-fund separation significantly simplify the portfolio optimization problem and allow to study possible deviations of optimal portfolios from the mean-variance efficiency frontier.

Chapter 3 and Chapter 4 deal with Model Risk Assessment and its application in the context of risk aggregation and life insurance pricing, respectively.

The assessment of portfolio risk is often explicitly (e.g., the square root formula under Basel III) or implicitly (e.g., credit risk portfolio models) driven by the marginal distributions of the risky components and the correlations amongst them. In Chapter 4, we e assess the extent by which such practice is prone to model error.

In the case of a sum of $n = 2$ risks, we investigate under which conditions the unconstrained Value-at-Risk (VaR) bounds (which are the maximum and minimum VaR for $S = \sum_{i=1}^{n} X_i$ when only the marginal distributions of the X_i are known) coincide with the (constrained) VaR bounds when in addition one has information on some measure of dependence (e.g., Pearson correlation or Spearman's rho). We find that both bounds coincide if the measure of dependence takes value in an interval that we explicitly determine. For probability levels used in risk management practice, we show that using correlation information has typically no effect on the highest possible VaR whereas it can affect the lowest possible VaR.

In the case of a general sum of $n \geqslant 2$ risks, we derive Range Value-at-Risk (RVaR) bounds under an average correlation constraint (in addition to the knowledge of the marginal distributions). While these bounds are not best-possible in general, we show that they are in the case of a sum of $n \geqslant 3$ standard uniformly distributed risks. As far as we know, this result is the first that provides a best-possible bound on RVaR for a general sum of $n \geqslant 3$ risks (uniformly distributed) under a correlation constraint.

Survival probabilities are required in many actuarial evaluations, such as the computation of net premiums in the life insurance business. As the output of a statistical procedure, their estimation is subject to uncertainty. In Chapter 4, we propose a robust assessment for the net premium of a standard life insurance contract with respect to the uncertainty on the estimated residual lifetime distribution function. Specifically, we provide a method to derive the range of values that the net premium of a given contract can attain when considering all residual lifetime distribution functions that satisfy an L^2 distance constraint with a reference distribution function. The results obtained in this chapter can be used to obtain a conservative evaluation of the net premium that an insurance company should charge for a life insurance contract in order to avoid financial losses due to unexpected changes in longevity trends. A key point of our analysis is showing that the net premium bounds obtained using an L^2 distance constraint be easily computed, a feature that is not common in the literature related to Model Risk Assessment but that is important to encourage industrial applications of this research line.

Efforts have been undertaken to make the notation as homogeneous as possible. However, due to the different approaches considered in the chapters, different levels of mathematical notation may be required, leading to the necessity to redefine specific notations.

1.2 Related publications

The following three chapters are based on three research papers that I co-authored:

1. Bernard C., De Vecchi C. and Vanduffel S. (2021). When do two- or three-fund separation theorems hold? Quantitative Finance, 21(11):1869-1883.

2. Bernard C., De Vecchi C. and Vanduffel S. (2022). The impact of correlation on (Range) Value-at-Risk. Submitted.

3. Bernard C., De Vecchi C. and Vanduffel S. (2022). Robust assessment of life insurance products. Submitted.

Chapter 2

When do two- or three-fund separation theorems hold?

2.1 Introduction

James Tobin and Harry Markowitz laid out the foundation of modern portfolio theory. Specifically, Tobin (1958) was the first to provide a (two-fund) separation theorem. He argued that in a world with one safe asset and a large number of risky assets, investors should combine cash with a single portfolio of risky assets. Owen and Rabinovitch (1983) point out that Tobin's separation result holds for any stochastic return generating process if the investor's utility function is quadratic and for any concave increasing utility function if the returns are multivariate normally distributed. Chamberlain (1983) extends this last result to the class of elliptical distributions.

The assumption of quadratic utility to justify a two-fund separation is however highly problematic. Indeed, quadratic utility implies increasing absolute risk-aversion, which has an unrealistic behavioral implication in that an increase in available wealth leads to lower investments in risky assets and not more (Huang and Litzenberger (1988)). Moreover, utility theory itself has been criticized for not being consistent with real-world decision making and a series of alternative decision theories have emerged. The most prominent amongst these is the so-called Cumulative Prospect Theory (CPT) from Tversky and Kahneman (1992). Levy and Levy (2004) show that when returns are normally distributed, optimal portfolios for CPT-investors are to be found in the set of mean-variance efficient portfolios. This result was generalized by Pirvu and Schulze (2012) who show that a two-fund separation theorem holds under elliptically distributed returns. However, whilst each of these alternative decision theories has its own features and is of interest, none of them is deemed suitable for accommodating all possible investors' preferences. Therefore, in this analysis we do not make specific assumptions on the choice of the behavioral theory, rather we only assume some properties that are rarely disputed. For instance, most adopted theories agree that more is better than less (compliance with first-order stochastic dominance) and that in addition, a certain income is better than an uncertain one with the same mean (compliance with second-order stochastic dominance). The results we derive hold for all the preferences that satisfy at least one of these two key properties.

As for the assumption that returns can be described by a multivariate elliptical model, some discussion is needed. As yearly returns are in essence sums of daily returns, one may expect that they display a Gaussian pattern; see Cesari and Cremonini (2003) for formal empirical evidence. However, studies based on daily returns show that asset returns typically exhibit skewness; see for instance Eberlein and Keller (1995), Küchler et al. (1999), and Carr et al. (2002), amongst others. The effect of skewness on optimal portfolio choice (under various theories of choice under risk) has been explored in a series of papers. Assuming that asset returns are distributed according to a location-scale mixture of normals, Mencía and Sentana (2009) show that mean-variance-skewness efficient portfolios exhibit three-fund separation. Assuming a generalized hyperbolic (GH) skewed Student t-distribution for the returns, Birge and Chavez-Bedoya (2016) show that three-fund separation also holds for exponential utility maximizers and they obtain explicit solutions in various cases of interest. Vanduffel and Yao (2017) extend this result by characterizing optimal portfolios for risk averse expected utility maximizers when returns follow a so-called multivariate generalized hyperbolic (MGH) distribution[1] (which includes the GH skewed Student t-distribution as a special case), that is they find that all risk-averse expected utility maximizers invest in three funds only. Kwak and Pirvu (2018) also obtain three-fund separation but model preferences with Cumulative Prospect Theory (CPT).

In this analysis we make very weak assumptions on preferences and derive two-fund theorems assuming the returns have a location-scale distribution and three-fund theorems when the returns are assumed to have a conditional location-scale property. Specifically, we show first that when returns have a location-scale property (equivalently, they are elliptically distributed), all decision theories that are compliant with First-order Stochastic Dominance (FSD) yield optimal portfolios that exhibit two-fund separation. So, while these theories may be very different they essentially lead to similar portfolio compositions. Second, when the returns only need to satisfy a conditional location-scale property (e.g., when they follow a MGH distribution) a three-fund separation can still be obtained. The proofs of these results are rather straightforward and generalize all mentioned specific results. For instance, Cumulative Prospect Theory preferences are consistent with FSD and hence the results of Pirvu and Schulze (2012) (see also Levy and Levy (2004)) in the elliptical case are a consequence of ours.

The remainder of the chapter is organized as follows. In Section 2.2 we formulate the optimal portfolio problem and provide our assumptions relative to the preferences. In Section 2.3 we prove a two-fund separation theorem when returns have a location scale distribution. In Section 2.4 we prove a three-fund separation theorem when returns have conditional location-scale property (MGH distribution). Section 2.5 illustrates the theoretical results with a numerical application. We conclude in Section 2.6.

[1] The MGH distribution was introduced in the literature by Barndorff-Nielsen (1978), Barndorff-Nielsen (1997) and Blaesild and Jensen (1981) and has shown to be useful for modeling asset returns (Barndorff-Nielsen (1997), McNeil et al. (2010)).

2.2 Problem formulation and preferences

In this chapter we study the optimal allocation of wealth among n assets for an investor under fairly weak assumptions on his preferences. We first describe the market setting and then discuss the weak assumptions that we make on the investors' preferences.

We consider a single period economy in which there are $n+1$ assets available for investment. There is one risk-free asset yielding a fixed return $r > 0$ and there are n risky assets yielding stochastic returns that are described by the vector $\mathbf{X} = (X_1, \ldots, X_n)^T$ and have a joint distribution $F_{\mathbf{X}}$. We denote by $\boldsymbol{m} = (m_1, ..., m_n)^T$ the vector of expected returns and by $\boldsymbol{\Delta}$ their positive definite covariance matrix. In what follows, we tacitly assume that all these quantities exist and are finite.

Let W_0 denote the total fixed initial wealth and $\boldsymbol{\omega} := (\omega_1, ..., \omega_n)^T$ be the vector of amounts invested in the n different risky assets (the remaining amount is thus invested in the risk-free asset). We call $\boldsymbol{\omega}$ a portfolio. The final wealth $W_{\boldsymbol{\omega}}$ of the portfolio writes as

$$\begin{aligned} W_{\boldsymbol{\omega}} &= \sum_{i=1}^{n} \omega_i(1+X_i) + \left(W_0 - \sum_{i=1}^{n} \omega_i\right)(1+r) \\ &= W_0(1+r) + \sum_{i=1}^{n} \omega_i(X_i - r). \end{aligned} \tag{2.2.1}$$

Denote by $\mathcal{W}$ the set of final wealths that can be purchased with initial wealth W_0 and denote by $V(\cdot)$ the investor's objective function. The investor's goal is to determine the optimal portfolio $\boldsymbol{\omega}^*$ (equivalently, the optimal terminal wealth $W_{\omega^*} \in \mathcal{W}$) by solving the following optimization problem

$$\max_{\boldsymbol{\omega}} V(W_{\boldsymbol{\omega}}). \tag{2.2.2}$$

In this chapter, we do not explicitly specify[2] the objective function $V(\cdot)$. Nevertheless, we state some properties that appear very natural for "reasonable" objective functions to satisfy. In what follows we denote by F_W the distribution function of a random terminal wealth $W \in \mathcal{W}$ and by F_W^{-1} its quantile function (defined as the left inverse of F_W).

Definition 2.2.1. *Let $W_1, W_2 \in \mathcal{W}$. We say that W_1 is first-order stochastically dominated by W_2, denoted as $W_1 \prec_{FSD} W_2$, if for all $p \in (0,1)$, $F_{W_1}^{-1}(p) \leqslant F_{W_2}^{-1}(p)$.*

It is intuitive that investors when choosing between W_1 and W_2 will prefer W_2 whenever $W_1 \prec_{FSD} W_2$.

Assumption 2.2.1 (FSD-consistency on $\mathcal{W}$)**.** *Preferences $V(\cdot)$ are consistent with first-order stochastic dominance (FSD) on $\mathcal{W}$. That is, for $W_1, W_2 \in \mathcal{W}$, $W_1 \prec_{FSD} W_2$ implies $V(W_1) \leqslant V(W_2)$ and equality only holds when W_1 and W_2 have the same distribution.*

[2]It could for instance be an expected utility, i.e., $V(W_{\boldsymbol{\omega}}) := \mathbb{E}[U(W_{\boldsymbol{\omega}})]$ in which $U(x)$ is some specific utility function. It could also refer to a non-expected utility setting such as the decision theories of Yaari (1987), Tversky and Kahneman (1992) or Quiggin (1993).

It is known that Assumption 2.2.1 is also equivalent to having a law-invariant and increasing objective function $V(\cdot)$ (see Theorem 1 in Bernard et al. (2015)). So, being consistent with FSD is equivalent to assuming that "more is preferred to less" ($a < b \Rightarrow V(a) < V(b)$) and that optimal choices are only driven by the distribution of final wealth and not by the states in which cash-flows are received (law-invariance). Clearly, Assumption 2.2.1 is completely natural and most decision theories comply with it. In fact, many economists consider a violation of the FSD property as grounds for refuting a particular model; see, for example, Birnbaum (1997), Birnbaum and Navarrette (1998) for more discussions. Recall also that although the original prospect theory by Kahneman and Tversky (1979) provides explanations for phenomena that were unexplained before, it violates first-order stochastic dominance. To overcome this potential issue, Tversky and Kahneman (1992) have developed the cumulative prospect theory. In what follows, investors with preferences that comply with Assumption 2.2.1 are called *FSD-investors*.

Definition 2.2.2. *Let $W_1, W_2 \in \mathcal{W}$. We say that W_1 is second-order stochastically dominated by W_2, denoted as $W_1 \prec_{SSD} W_2$, if for all $p \in (0,1)$, $\int_0^p F_{W_1}^{-1}(q)dq \leqslant \int_0^p F_{W_2}^{-1}(q)dq$.*

Assumption 2.2.2 (SSD-consistency)**.** *Preferences $V(\cdot)$ are consistent with second-order stochastic dominance (SSD). That is, $W_1 \prec_{SSD} W_2$ implies $V(W_1) \leqslant V(W_2)$ and equality only holds when W_1 and W_2 have the same distribution.*

For instance, expected utility maximizers that employ an increasing and concave utility function to make decisions have preferences $V(\cdot)$ that are SSD-consistent. That is, $W_1 \prec_{SSD} W_2$ implies $\mathbb{E}(u(W_1)) \leqslant \mathbb{E}(u(W_2))$, where $u(x)$ is an increasing and concave utility function. Clearly, SSD-consistency implies FSD-consistency but the opposite does not hold true in general. In general, being SSD-consistent is quite a strong assumption. For instance, preferences that are consistent with rank dependent utility theory exhibits FSD-consistency but not SSD-consistency (Ryan (2006)) and the same holds true for cumulative prospect theory (see e.g., Baucells and Heukamp (2006)). In what follows, investors with preferences that comply with Assumption 2.2.2 are called *SSD-investors*.

2.3 Two-fund separation theorems

We first study a market model in which the joint distribution $F_{\mathbf{X}}$ of the vector of asset returns $\mathbf{X}$ is assumed to belong to a so-called location-scale family of distributions. We then derive a two-fund theorem for FSD-investors. We point out that various two-fund theorems that have been derived in the literature under specific assumptions on preferences (e.g., preferences according to the cumulative prospect theory Tversky and Kahneman (1992)) and on distributions (e.g., following an elliptical model) comply with this setting and are merely particular cases of the characterization result we provide. Furthermore, we show that in this market setting, SSD-investors cannot be distinguished from FSD-investors, i.e., being SSD-consistent is equivalent to being FSD-consistent.

2.3.1 Distributional assumption on returns

Definition 2.3.1 (Location-scale property of $F_{\mathbf{X}}$). *Let Z be a real-valued random variable taking values having zero mean and unit variance. We say that $F_{\mathbf{X}}$ of $\mathbf{X} = (X_1, X_2, ..., X_n)$ has the location-scale property associated with Z if for any vector $\mathbf{a} = (a_1, \ldots, a_n)^T$ it holds that*

$$\mathbf{a}^T\boldsymbol{X} \overset{d}{=} \mathbf{a}^T\boldsymbol{m}+\sqrt{\mathbf{a}^T\boldsymbol{\Delta}\mathbf{a}}Z, \tag{2.3.1}$$

where "$\overset{d}{=}$" denotes the equality in distribution and where we recall that $\mathbf{m}$ is the vector of expected returns of $\mathbf{X}$ and $\boldsymbol{\Delta}$ is their positive definite covariance matrix.

The family $\mathcal{F}$ of all multivariate distributions that have this location-scale property is then called the location-scale family of distributions associated with Z.

Specifically, if the joint distribution $F_{\mathbf{X}}$ of $\mathbf{X} = (X_1, X_2, ..., X_n)$ is a member of $\mathcal{F}$, then for all i, there exist $m_i \in \mathbb{R}$ and $\delta_i > 0$ such that $X_i \overset{d}{=} m_i + \delta_i Z$, where m_i is the mean and δ_i is the standard deviation of X_i. The distributional constraint (2.3.1) is fairly restrictive, as it imposes a condition on the distribution of all linear combinations. In fact, Chamberlain (1983) provides a characterization result that makes it possible to conclude that the only family $\mathcal{F}$ satisfying the condition in Definition 2.3.1 is the multivariate elliptical family, that is when Z is an elliptically distributed random variable. The equivalence of (i) and (ii) in Proposition 2.3.2 can be found in Theorem 1 of Chamberlain (1983).

Proposition 2.3.2 (Chamberlain (1983)). *Let $\mathbf{X}$ be a random vector of $\mathbb{R}^n$ with invertible covariance matrix $\boldsymbol{\Delta}$ (with Cholesky decomposition $\boldsymbol{\Delta} = \mathbf{L}\mathbf{L}^T$) and mean $\boldsymbol{m}$. The three following statements are equivalent:*

(i) For all $\mathbf{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, for all $c \in \mathbb{R}$, the distribution of $\mathbf{a}^T\mathbf{X} + c$ is determined by its mean $\mathbf{a}^T\boldsymbol{m} + c$ and its variance $\mathbf{a}^T\boldsymbol{\Delta}\mathbf{a}$.

(ii) $F_{\mathbf{X}}$ belongs to the multivariate elliptical family with covariance matrix $\boldsymbol{\Delta} = \mathbf{L}\mathbf{L}^T$ and mean vector $\boldsymbol{m}$, i.e., $\mathbf{Z} = \mathbf{L}^{-1}(\mathbf{X} - \boldsymbol{m})$ is spherically distributed.

(iii) $F_{\mathbf{X}}$ belongs to a location-scale family $\mathcal{F}$ associated to a random variable $Z := Z_1$ (where $\mathbf{Z} := \mathbf{L}^{-1}(\mathbf{X} - \boldsymbol{m})$) as described in Definition 2.3.1.

Proof. **(iii)⇒(i):**. Let the distribution of $\mathbf{X}$ be in $\mathcal{F}$, as defined in Definition 2.3.1. Then, for any $\mathbf{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, the mean and the standard deviation of $\mathbf{a}^T\mathbf{X}$ are equal to $\mathbf{a}^T\boldsymbol{m}$ and $\sqrt{\mathbf{a}^T\boldsymbol{\Delta}\mathbf{a}}$, respectively. Moreover, the cdf of $\mathbf{a}^T\mathbf{X}$ can be expressed as $F_{\mathbf{a}^T\mathbf{X}}(x) = F_Z\left(\frac{x-\mathbf{a}^T\boldsymbol{m}}{\sqrt{\mathbf{a}^T\boldsymbol{\Delta}\mathbf{a}}}\right)$. Therefore, the distribution of $\mathbf{a}^T\mathbf{X}$ is completely specified by its mean $\mathbf{a}^T\boldsymbol{m}$, and its standard deviation $\sqrt{\mathbf{a}^T\boldsymbol{\Delta}\mathbf{a}}$.

(i)⇒(ii): Assume that the distribution of $\mathbf{a}^T\mathbf{X}+c$ is characterized by its mean and its variance for all $\mathbf{a} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Using the Cholesky decomposition, there exists a triangular invertible matrix $\mathbf{L}$ such that $\boldsymbol{\Delta} = \mathbf{L}\mathbf{L}^T$. Define $\mathbf{T} = \mathbf{L}^{-1}$ and $\mathbf{Z} = \mathbf{T}(\mathbf{X} - \boldsymbol{m})$. Then $\mathbb{E}[\mathbf{Z}] = \mathbf{0}$ and the covariance matrix of $\mathbf{Z}$ is the identity matrix $\mathbf{I}_n$ (because $\text{cov}(\mathbf{Z}) = \text{cov}(\mathbf{L}^{-1}\mathbf{X}) =$

$\mathbf{L}^{-1}\text{cov}(\mathbf{X})(\mathbf{L}^{-1})^T = \mathbf{L}^{-1}\boldsymbol{\Delta}(\mathbf{L}^{-1})^T = \mathbf{L}^{-1}\mathbf{L}\mathbf{L}^T(\mathbf{L}^{-1})^T = \mathbf{I}_n$). Let $\mathbf{R}$ be an orthogonal matrix, i.e. $\mathbf{R}\mathbf{R}^T = \mathbf{I}$. Define $\mathbf{w} = \mathbf{R}\mathbf{Z}$ then $\mathbb{E}[\mathbf{w}] = \mathbf{0}$ and the covariance matrix of $\mathbf{w}$ is the identity matrix $\mathbf{I}_n$ (because $\text{cov}(\mathbf{R}\mathbf{Z}) = \mathbf{R}\text{cov}(\mathbf{Z})\mathbf{R}^T$). For any $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a}^T\mathbf{Z} = \mathbf{a}^T\mathbf{T}(\mathbf{X}-\boldsymbol{m}) = \mathbf{a}_1^T\mathbf{X}+c_1$ and $\mathbf{a}^T\mathbf{w} = \mathbf{a}^T\mathbf{R}\mathbf{T}(\mathbf{X}-\boldsymbol{m}) = \mathbf{a}_2^T\mathbf{X}+c_2$ are two portfolios with the same mean 0 and the same variance $\mathbf{a}^T\mathbf{a}$. But by (i), the distributions of $\mathbf{a}_1^T\mathbf{X}+c_1$ and $\mathbf{a}_2^T\mathbf{X}+c_2$ are characterized by their means and variance, thus $\mathbf{a}^T\mathbf{Z}$ and $\mathbf{a}^T\mathbf{w}$ have the same distribution for all $\mathbf{a}$. Thus we conclude that $\mathbf{Z}$ and $\mathbf{w}$ have the same distribution (a distribution is characterized by the distribution of all linear combinations as then for all $\mathbf{t} \in \mathbb{R}^n$, $\mathbb{E}[e^{i\mathbf{t}^T\mathbf{X}}] = \mathbb{E}[e^{i\mathbf{t}^T\mathbf{Y}}]$, i.e. the vectors $\mathbf{X}$ and $\mathbf{Y}$ have the same characteristic function and thus must have the same distribution). Hence $\mathbf{Z}$ is spherically distributed about $\mathbf{0}$.

Proof of (ii)⇒(iii): It is well-known that the elliptical family of distributions satisfies (iii). See, for example, Section in 3.3 McNeil et al. (2010). □

2.3.2 Characterization of optimal portfolios for FSD-investors

When the joint distribution $F_{\mathbf{X}}$ of the random return vector $\mathbf{X} = (X_1, X_2, ..., X_n)$ has the location-scale property associated with Z, we obtain from (2.3.1) that the terminal wealth $W_{\boldsymbol{\omega}}$ defined by (2.2.1) satisfies

$$W_{\boldsymbol{\omega}} \stackrel{d}{=} m_{\boldsymbol{\omega}} + \delta_{\boldsymbol{\omega}} Z \tag{2.3.2}$$

with parameters $m_{\boldsymbol{\omega}}$ and $\delta_{\boldsymbol{\omega}}$ given as

$$\begin{cases} m_{\boldsymbol{\omega}} := \mathbb{E}[W_{\boldsymbol{\omega}}] = W_0(1+r) + \boldsymbol{\omega}^T(\boldsymbol{\mu}-r\mathbf{1}) \\ \delta_{\boldsymbol{\omega}} := \text{std}[W_{\boldsymbol{\omega}}] = \sqrt{\boldsymbol{\omega}^T\boldsymbol{\Delta}\boldsymbol{\omega}}, \end{cases} \tag{2.3.3}$$

where $\mathbf{1}$ is a vector of ones. The terminal wealth W_ω that arises from the portfolio allocation $\boldsymbol{\omega}$ is thus characterized by the coefficients $m_{\boldsymbol{\omega}}$ and $\delta_{\boldsymbol{\omega}}$ given in (2.3.3) and we can thus reformulate our optimization problem (2.2.2) as

$$\max_{(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})\in\mathcal{A}} V(m_{\boldsymbol{\omega}} + \delta_{\boldsymbol{\omega}} Z) \tag{2.3.4}$$

where $\mathcal{A}$ is the set of all couples $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$, as in (2.3.3), i.e., $\mathcal{A}$ is given as

$$\mathcal{A} := \left\{ \left(\sqrt{\boldsymbol{\omega}^T\boldsymbol{\Delta}\boldsymbol{\omega}}, W_0(1+r) + \boldsymbol{\omega}^T(\boldsymbol{m} - r\mathbf{1}) \right) \right\}_{\boldsymbol{\omega}\in\mathbb{R}^n} \tag{2.3.5}$$

Given that the optimization (2.3.4) only deals with the mean $m_{\boldsymbol{\omega}}$ and the standard deviation $\delta_{\boldsymbol{\omega}}$ of portfolios $\boldsymbol{\omega}$, it becomes apparent that a connection to the mean-variance analysis developed by Markowitz (1952) holds.

Definition 2.3.3 (Mean-variance efficiency frontier)**.** *Consider a portfolio $\boldsymbol{\omega}$ with terminal wealth $W_{\boldsymbol{\omega}}$ having mean $m_{\boldsymbol{\omega}}$ and variance $\delta_{\boldsymbol{\omega}}^2$. The portfolio $\boldsymbol{\omega}$ is mean-variance efficient if there is no portfolio yielding a terminal wealth with the same variance but a strictly larger mean. The set $\mathcal{A}^*$ $\subseteq \mathcal{A}$ containing all pairs $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$ for which $\boldsymbol{\omega}$ is a mean-variance efficient portfolio is called the mean-variance efficiency frontier.*

Proposition 2.3.4. *The mean-variance efficiency frontier $\mathcal{A}^*$ is explicitly given as*

$$\mathcal{A}^* = \left\{ (\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}}) \mid \delta_{\boldsymbol{\omega}} > 0 \text{ and } m_{\boldsymbol{\omega}} = W_0(1+r) + \delta_{\boldsymbol{\omega}} \sqrt{h} \right\}_{\boldsymbol{\omega} \in \mathbb{R}^n}, \tag{2.3.6}$$

where

$$h = (\boldsymbol{m} - \boldsymbol{r1})^T \boldsymbol{\Delta}^{-1} (\boldsymbol{m} - \boldsymbol{r1}) > 0. \tag{2.3.7}$$

Proof. Given $\delta_{\boldsymbol{\omega}} = \delta$, let us build the mean-variance efficient portfolio, i.e. the portfolio that solves the optimization problem

$$\max_{\boldsymbol{\omega}} m_{\boldsymbol{\omega}} \text{ subject to } \delta_{\boldsymbol{\omega}} = \delta. \tag{2.3.8}$$

This is a standard problem and using Lagrange multipliers one readily obtains that the optimal portfolio is given as

$$\boldsymbol{\omega}_* := \frac{\delta}{\sqrt{h}} \boldsymbol{\Delta}^{-1} (\boldsymbol{m} - r\boldsymbol{1}). \tag{2.3.9}$$

Using the fact that $\boldsymbol{\Delta}^{-1}$ is symmetric, the expected return of this portfolio is given by

$$\begin{aligned} m_{\boldsymbol{\omega}_*} &= W_0(1+r) + {\boldsymbol{\omega}_*}^T (\boldsymbol{m} - r\boldsymbol{1}) \\ &= W_0(1+r) + \frac{\delta_{\boldsymbol{\omega}_*}}{\sqrt{h}} (\boldsymbol{m} - r\boldsymbol{1})^T \boldsymbol{\Delta}^{-1} (\boldsymbol{m} - r\boldsymbol{1}) \\ &= W_0(1+r) + \delta_{\boldsymbol{\omega}_*} \sqrt{h}. \end{aligned}$$

□

It is clear from the characterization of the set $\mathcal{A}^*$ that for each $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}}) \in \mathcal{A}^*$, there exists exactly one portfolio $\boldsymbol{\omega}$ yielding this specific mean m_w and variance δ_w^2. In what follows, we sometimes identify such portfolio $\boldsymbol{\omega}$ with the pair $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$ and correspondingly call $\mathcal{A}^*$ also the set of mean-variance efficient portfolios.

Proposition 2.3.5 (Two-fund theorem)**.** *When $V(\cdot)$ satisfies Assumption 2.2.1 and when Problem* (2.2.2) *has a solution $\boldsymbol{\omega}_*$, then $(\delta_{\boldsymbol{\omega}_*}, m_{\boldsymbol{\omega}_*}) \in \mathcal{A}^*$. Furthermore, an optimum $\boldsymbol{\omega}_*$ is of the form*

$$\boldsymbol{\omega}_* = \frac{\delta}{\sqrt{h}} \boldsymbol{\Delta}^{-1} (\boldsymbol{m} - r\boldsymbol{1}) \tag{2.3.10}$$

for some $\delta > 0$ that is such that $V(\boldsymbol{\omega})$ is maximum when $\boldsymbol{\omega} = \boldsymbol{\omega}_*$.

Proof. Let $\boldsymbol{\omega}_*$ be an optimal solution to Problem (2.2.2). Note that $\boldsymbol{\omega}_*$ must be mean-variance efficient, i.e., it must maximize the mean for a given standard deviation δ (Problem (2.3.8)). Indeed, if $\boldsymbol{\omega}_*$ is not mean-variance efficient one can find another portfolio $\boldsymbol{\omega}$ which is dominating in the sense of FSD and thus yields a higher objective value. Its expression was already derived above in (2.3.9), which ends the proof. □

There are several important implications from Proposition 2.3.5. *First*, regardless of their specific objective function, the optimal allocation in risky assets of FSD-investors is always *proportional* to $\boldsymbol{\Delta}^{-1}(\boldsymbol{m} - r\mathbf{1})$ and the remaining funds are invested in the risk-free asset. We label $\boldsymbol{\Delta}^{-1}(\boldsymbol{m} - r\mathbf{1})$ as the market-fund. Hence, the optimal portfolio of FSD-investors ultimately always translate in an optimal proportion that is allocated to the market-fund. This observation is important for practical investment advice, as eliciting the proportion (a single number) that an investor is prepared to allocate to the market-fund is much easier than eliciting his preferences (i.e., the objective function $V(\cdot)$ that he aims at maximizing). *Second*, solving the high-dimensional optimal portfolio problem in (2.2.2) amounts to solving a one-dimensional problem in which the only unknown is the parameter δ in (2.3.10), which maximizes the objective function. *Third*, this proposition shows that various contributions in the literature are merely special cases of the general characterization we provide. Levy and Levy (2004) show that under normally distributed returns, CPT-investors select their portfolio on the mean-variance efficient frontier. This result was generalized by Pirvu and Schulze (2012) for elliptically distributed returns. Bertsimas et al. (2004) obtain under an elliptical model that a two-fund theorem holds when investors minimize the expected shortfall for a given desired expected return. However, all these results are immediately implied by Proposition 2.3.5. In addition, from the proposition it also follows that under the distributional assumption we make, two-fund separation holds for investors with preferences described by Rank Dependent Utility Theory (Quiggin (1993)).

2.3.3 Characterization of optimal portfolios for SSD-investors

Recall that every SSD-investor is also an FSD-investor. Hence, Proposition 2.3.5 also applies to SSD-investors and their optimal portfolios thus also exhibit two-fund separation. However, by exploiting the specific characteristics of SSD-investors it might be possible to obtain a more refined characterization of their optimal portfolio. In this regard, it can be shown that the optimal portfolio of an SSD-investor must also solve the problem

$$\min_{\boldsymbol{\omega}} \ \delta_{\boldsymbol{\omega}} \text{ subject to } m_{\boldsymbol{\omega}} = m. \tag{2.3.11}$$

for some given $m > W_0(1+r)$[3]. Observe next that for every couple $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$ in $\mathcal{A}^*$ the corresponding portfolio $\boldsymbol{\omega}$ must be a solution to a problem of the form (2.3.11), since otherwise there exists $\boldsymbol{\omega}'$ such that $\delta_{\boldsymbol{\omega}'} < \delta_{\boldsymbol{\omega}}$ and $m_{\boldsymbol{\omega}'} = m_{\boldsymbol{\omega}}$, but this is not possible as we proved that the maximum attainable expected value is a strictly decreasing function of the standard deviation and $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$ is mean-variance efficient by assumption. This means that using the extra information that optimal portfolios must also solve a problem of the form (2.3.11) does not lead to a reduction of the set $\mathcal{A}^*$.

Remark 2.3.1 (Rationalization of portfolios)**.** *In general, one cannot expect that for every mean-variance efficient portfolio ω there exists a utility maximizer for whom this portfolio is optimum. However, when the distribution $F_{\mathbf{X}}$ of the random return vector $\mathbf{X} = (X_1, X_2, ..., X_n)$ has the*

[3]Consider two portfolios $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ such that $m_{\boldsymbol{\omega}} = m_{\boldsymbol{\omega}'}$ and $\delta_{\boldsymbol{\omega}} \leqslant \delta_{\boldsymbol{\omega}'}$. Under the location-scale distributional assumption, we have $W_{\boldsymbol{\omega}} \leqslant_{cx} W_{\boldsymbol{\omega}'}$. This implies that for any SSD-consistent objective function $V(W_{\boldsymbol{\omega}}) \geqslant V(W_{\boldsymbol{\omega}'})$. Therefore, a portfolio that does not satisfy (2.3.11) cannot be optimal for an SSD-investor.

location-scale property this holds true. Indeed, recall that in this case we consider the problem

$$\max_{(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}}) \in \mathcal{A}} V(m_{\boldsymbol{\omega}} + \delta_{\boldsymbol{\omega}} Z) \tag{2.3.12}$$

where $\mathcal{A}$ is the set that contains all couples $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$ as in (2.3.3). Consider now expected utility preferences $V(.) = \mathbb{E}(u(.))$ in which the utility function is given as $u(x) = x - \alpha x^2$ for some $\alpha > 0$. In this case, our maximization problem reads as

$$\max_{(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}}) \in \mathcal{A}} m_{\boldsymbol{\omega}} - \alpha(\delta_{\boldsymbol{\omega}}^2 + m_{\boldsymbol{\omega}}^2) \tag{2.3.13}$$

Clearly, if there is a solution $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$ it must be such that $\delta_{\boldsymbol{\omega}}$ is minimum for the given value of $m_{\boldsymbol{\omega}}$. As this property holds true for all $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}}) \in \mathcal{A}^*$, we only need to show that for every $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}}) \in \mathcal{A}^*$ there exists $\alpha > 0$ such that $(\delta_{\boldsymbol{\omega}}, m_{\boldsymbol{\omega}})$ solves problem (2.3.13). On $\mathcal{A}^*$ the optimization problem writes as

$$\max_{m_{\boldsymbol{\omega}} > W_0(1+r)} m_{\boldsymbol{\omega}} - \alpha \left[\frac{(m_{\boldsymbol{\omega}} - W_0(1+r))^2}{h} + m_{\boldsymbol{\omega}}^2 \right] \tag{2.3.14}$$

in which h is as in (2.3.7)*. Differentiation with respect to $m_{\boldsymbol{\omega}}$ and equating to zero yields that $\alpha = \frac{1}{2\left(m_{\boldsymbol{\omega}} + \frac{m_{\boldsymbol{\omega}} - W_0(1+r)}{h}\right)} > 0$. Note that for any $\alpha \geqslant \frac{1}{2W_0(1+r)}$ the risk-free investment is optimal (i.e., $m_{\boldsymbol{\omega}} = W_0(1+r)$ and $\delta_{\boldsymbol{\omega}} = 0$).*

Remark 2.3.2 (Two-fund theorems without distributional assumptions)**.** *In Section 3.1 it was shown that two-fund separation holds for general preferences under the key assumption that the joint distribution $F_{\mathbf{X}}$ of the return vector $\mathbf{X}$ has a location-scale property. Several contributions in the literature also derive a two-fund theorem without making distributional assumptions on asset returns. In this case, however, one requires specific preferences in that these solely balance the expected return ("reward") and the variance ("risk") of the terminal wealth. Specifically, De Giorgi et al. (2011) (Theorem 1) provide a two-fund theorem when the investor preferences can be described by*

$$V(W) = f(m(W), \rho(W)),$$

where f is monotonically decreasing in the risk $\rho(W)$ and monotonically increasing in the reward $m(W)$.

2.4 Three-fund separation theorems

In this section, we significantly relax the assumption of location-scale invariance for the joint distribution $F_{\mathbf{X}}$ of the return vector $\mathbf{X} = (X_1, X_2, ..., X_n)$. We derive a three-fund theorem and discuss the implications thereof. Specifically, we point out that our characterization of optimality implies various three-fund theorems that were derived in the literature under more restrictive assumptions.

Definition 2.4.1 (Location-scale mixture property of $F_{\mathbf{X}}$)**.** *Let Z be a random variable taking values in $\mathbb{R}$ and having zero mean and unit variance. Let $Y \geqslant 0$ a.s. be a positive random variable* *that is independent of Z. We say that the joint distribution $F_{\mathbf{X}}$ of $\mathbf{X} = (X_1, X_2, ..., X_n)$ has the location-scale mixture property associated with the random variables Z and Y if for any vector $\mathbf{a} = (a_1, \ldots, a_n)^T$ it holds that*

$$\mathbf{a}^T\mathbf{X} \stackrel{d}{=} \mathbf{a}^T\boldsymbol{\mu} + Y\mathbf{a}^T\boldsymbol{\gamma} + \sqrt{Y}\sqrt{\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a}}Z, \tag{2.4.1}$$

for some vectors $\boldsymbol{\mu}$, and $\boldsymbol{\gamma}$ and for the positive definite symmetric matrix $\boldsymbol{\Sigma}$, which can be interpreted as parameters.

The family $\mathcal{G}$ of all multivariate distributions that have this location-scale mixture property is then called the location-scale family of distributions associated with Z and Y.

If the joint distribution $F_{\mathbf{X}}$ of $\mathbf{X} = (X_1, X_2, ..., X_n)$ is a member of the location scale multivariate family associated with the random variables Z and Y, then for all i, X_i belongs to the same location-scale mixture family associated with Z and Y, i.e., there exist $\mu_i \in \mathbb{R}$, $\gamma_i \in \mathbb{R}$, and $\sigma_i > 0$ such that $X_i = \mu_i + Y\gamma_i + \sqrt{Y}\sigma_i Z$. The same is true for all univariate affine transformations of $\mathbf{X}$.

A prominent example of Definition 2.4.1 arises when $F_{\mathbf{X}}$ belongs to the so-called multivariate generalized hyperbolic (MGH) family of distributions (Section 3.2 in McNeil et al. (2010)), which is a natural extension of the multivariate elliptical family of distributions. In this case, we find that $\mathbf{X} = (X_1, ..., X_n) \sim F_{\mathbf{X}}$ can be represented as

$$(X_1, ..., X_n) \stackrel{d}{=} \boldsymbol{\mu} + Y\boldsymbol{\gamma} + \sqrt{Y}\mathbf{A}\mathbf{Z},$$

where $\mathbf{Z}$ is a random vector that follows a multivariate normal distribution $MVN_k(0, I_k)$, $\mathbf{A} \in \mathbb{R}^{n\times k}$ is a matrix to be chosen taking into account that $\mathbf{AZ} \sim MVN_n(0, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$, and the scalar factor in the mixture, Y, is a generalized inverse Gaussian distribution with parameters λ, χ and ψ (Section 3.2 in McNeil et al. (2010)).

2.4.1 Characterization of optimal portfolios for FSD-investors

When the joint distribution $F_{\mathbf{X}}$ of the random return vector $\mathbf{X} = (X_1, X_2, ..., X_n)$ has the location-scale mixture property (2.4.1) associated with the random variables Z and Y, then we obtain that the final wealth $W_{\boldsymbol{\omega}}$ satisfies

$$W_{\boldsymbol{\omega}} \stackrel{d}{=} \mu_{\boldsymbol{\omega}} + Y\gamma_{\boldsymbol{\omega}} + \sqrt{Y}\sigma_{\boldsymbol{\omega}}Z \tag{2.4.2}$$

and has a distribution depending on the following three parameters

$$\begin{cases} \mu_{\boldsymbol{\omega}} = W_0(1+r) + \boldsymbol{\omega}^T(\boldsymbol{\mu} - r\mathbf{1}) \\ \gamma_{\boldsymbol{\omega}} = \boldsymbol{\omega}^T\boldsymbol{\gamma} \\ \sigma_{\boldsymbol{\omega}} = \sqrt{\boldsymbol{\omega}^T\boldsymbol{\Sigma}\boldsymbol{\omega}} \end{cases}. \tag{2.4.3}$$

In this regard, we point out that unlike the case of a distribution $F_{\mathbf{X}}$ with a location-scale property, the parameters $\mu_{\boldsymbol{\omega}}$ and $\sigma_{\boldsymbol{\omega}}$ can no longer be readily interpreted as the expected value and standard deviation of the terminal wealth W_ω. It is straightforward to show that

$$\begin{cases} \mathbb{E}(W_\omega) = \mu_{\boldsymbol{\omega}} + \mathbb{E}(Y)\gamma_{\boldsymbol{\omega}}, \\ var(W_\omega) = \gamma_{\boldsymbol{\omega}}^2 var(Y) + \sigma_{\boldsymbol{\omega}}^2 \mathbb{E}(Y), \\ skew(W_{\boldsymbol{\omega}}) = \frac{\gamma_{\boldsymbol{\omega}}^3 \mathbb{E}\left[(Y-\mathbb{E}(Y))^3\right] + 3\gamma_{\boldsymbol{\omega}}\sigma_{\boldsymbol{\omega}}^2 var(Y) + \sigma_{\boldsymbol{\omega}}^3 \mathbb{E}\left(Y^{\frac{3}{2}}\right)\mathbb{E}(Z^3)}{var(W_\omega)^{\frac{3}{2}}}, \end{cases} \tag{2.4.4}$$

where we used the following formula for the skewness, $skew(W_{\boldsymbol{\omega}}) = \frac{\mathbb{E}[(W_{\boldsymbol{\omega}} - \mathbb{E}(W_{\boldsymbol{\omega}}))^3]}{var(W_{\boldsymbol{\omega}})^{\frac{3}{2}}}$.

In order to solve the portfolio optimization problem (2.2.2) under the new assumption on the return distributions, we can reformulate this problem now as

$$\max_{(\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \in \mathcal{B}} V(W_{\boldsymbol{\omega}}) \tag{2.4.5}$$

where the set of triplets $\mathcal{B}$ is given as

$$\mathcal{B} := \left\{ \left(W_0(1+r) + \boldsymbol{\omega}^T(\boldsymbol{\mu} - r\mathbf{1}), \sqrt{\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}}, \boldsymbol{\omega}^T \boldsymbol{\gamma} \right) \right\}_{\boldsymbol{\omega} \in \mathbb{R}^n} \tag{2.4.6}$$

Definition 2.4.2 ("Mean-skewness-variance" efficiency frontier)**.** *Consider a portfolio $\boldsymbol{\omega}$ with terminal wealth $W_{\boldsymbol{\omega}}$ having parameters $\mu_{\boldsymbol{\omega}}$, $\gamma_{\boldsymbol{\omega}}$ and $\sigma_{\boldsymbol{\omega}}$. The portfolio $\boldsymbol{\omega}$ is said to be "mean-skewness-variance" efficient if there is no portfolio that has the same value for $\sigma_{\boldsymbol{\omega}}$ while having values for $\mu_{\boldsymbol{\omega}}$ and $\gamma_{\boldsymbol{\omega}}$ that are at least as big. In particular, the set $\mathcal{B}^*$ containing all triplets $(\mu_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}})$ for which $\boldsymbol{\omega}$ is a "mean-skewness-variance" efficient portfolio is called the "mean-skewness-variance" efficient frontier.*

Proposition 2.4.3. *The "mean-skewness-variance" efficiency frontier $\mathcal{B}^*$ is explicitly given as*

$$\mathcal{B}^* = \left\{ (\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \;\middle|\; \begin{array}{l} \mu_{\boldsymbol{\omega}} = W_0(1+r) + \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{hg-k^2}\sqrt{\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2}}{g}, \\ \sigma_{\boldsymbol{\omega}} > 0, \gamma_{\boldsymbol{\omega}} \in \left[\frac{k}{\sqrt{h}}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}}\right) \end{array} \right\} \tag{2.4.7}$$

where h, g and k are defined as follows

$$h = (\boldsymbol{\mu} - \boldsymbol{r}\mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{r}\mathbf{1}) > 0, \quad g = \boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma} > 0, \quad k = (\boldsymbol{\mu} - \boldsymbol{r}\mathbf{1})^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}. \tag{2.4.8}$$

Proof. We first consider the problem:

$$\max_{\boldsymbol{\omega}} \ \mu_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \sigma, \gamma_{\boldsymbol{\omega}} = \gamma. \tag{2.4.9}$$

The well-posedness of Problem (2.4.9) is discussed in Appendix 2.7.1, where we show that the constraints σ and γ must satisfy $\gamma^2 < g\sigma^2$. Equivalently, we first consider the maximization of

$$\sum_{i=1}^n \omega_i(\mu_i - r) + W_0(1+r) - \lambda_1 \left(\sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \sigma_{ij} - \sigma^2 \right) - \lambda_2 \left(\sum_{i=1}^n \omega_i \gamma_i - \gamma \right)$$

in which λ_1 and λ_2 are Lagrange multipliers and we have used the expression in (2.4.3). After differentiation with respect to ω_i and equating to 0, we can rewrite the n equations in the following condensed way

$$\boldsymbol{\omega}_* = \frac{1}{2\lambda_1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{1}) - \frac{\lambda_2}{2\lambda_1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$$

where λ_1 and λ_2 are such that $\boldsymbol{\omega}_*^T\boldsymbol{\gamma} = \gamma$ and $\boldsymbol{\omega}_*^T\boldsymbol{\Sigma}\boldsymbol{\omega}_* = \sigma^2$. After rewriting this as a quadratic equation in λ_2 we obtain after some calculation that

$$\lambda_1 = \frac{\sqrt{hg - k^2}}{2\sqrt{\sigma^2 g - \gamma^2}}, \quad \lambda_2 = \frac{k}{g} - \frac{\gamma}{g}\frac{\sqrt{hg - k^2}}{\sqrt{\sigma^2 g - \gamma^2}}.$$

The portfolio $\boldsymbol{\omega}^*_{\sigma,\gamma}$ that solves Problem (2.4.9) is thus given as

$$\boldsymbol{\omega}^*_{\sigma,\gamma} = \frac{\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{\sqrt{hg - k^2}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mathbf{r}\mathbf{1}) - \left(\frac{k\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{g\sqrt{hg - k^2}} - \frac{\gamma_{\boldsymbol{\omega}}}{g}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}. \tag{2.4.10}$$

To obtain the value for $\mu_{\boldsymbol{\omega}}$ that corresponds to the portfolio $\boldsymbol{\omega}^*_{\sigma,\gamma}$, note that

$$\mu_{\boldsymbol{\omega}} = W_0(1 + r) + \left(\boldsymbol{\omega}^*_{\sigma,\gamma}\right)^T(\boldsymbol{\mu} - r\mathbf{1}).$$

We then obtain that

$$\begin{aligned}
\left(\boldsymbol{\omega}^*_{\sigma,\gamma}\right)^T(\boldsymbol{\mu} - r\mathbf{1}) &= \frac{\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{\sqrt{hg - k^2}}(\boldsymbol{\mu} - \mathbf{r}\mathbf{1})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{1}) - \left(\frac{k\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{g\sqrt{hg - k^2}} - \frac{\gamma_{\boldsymbol{\omega}}}{g}\right)\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{1}) \\
&= \frac{\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{\sqrt{hg - k^2}}h - \left(\frac{k\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{g\sqrt{hg - k^2}} - \frac{\gamma_{\boldsymbol{\omega}}}{g}\right)k = \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{\sqrt{hg - k^2}}\left(\frac{hg - k^2}{g}\right) \\
&= \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}\sqrt{hg - k^2}}{g},
\end{aligned}$$

where we used the fact that $k = k^T = \boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{1})$. We denote by $\mathcal{B}_1^*$ the set of portfolios that have maximum value for $\mu_{\boldsymbol{\omega}}$ given $\sigma_{\boldsymbol{\omega}}$ and $\gamma_{\boldsymbol{\omega}}$. This set is thus explicitly given as

$$\mathcal{B}_1^* = \left\{(\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \;\middle|\; \begin{array}{l} \mu_{\boldsymbol{\omega}} = W_0(1 + r) + \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{hg - k^2}\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{g}, \\ \sigma_{\boldsymbol{\omega}} > 0, \gamma_{\boldsymbol{\omega}} \in \left(-\sqrt{g}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}}\right) \end{array}\right\} \tag{2.4.11}$$

Clearly, any "mean-skewness-variance" efficient portfolio belongs to $\mathcal{B}_1^*$.

Next, we study for a given value for $\sigma_{\boldsymbol{\omega}}$, the functional relationship between $\mu_{\boldsymbol{\omega}}$ and $\gamma_{\boldsymbol{\omega}}$ on the set $\mathcal{B}_1^*$. We compute the following first and second derivatives

$$\frac{\partial\mu_{\boldsymbol{\omega}}}{\partial\gamma_{\boldsymbol{\omega}}} = \frac{k}{g} - \frac{\sqrt{hg - k^2}}{g}\frac{\gamma_{\boldsymbol{\omega}}}{\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}} \tag{2.4.12}$$

and

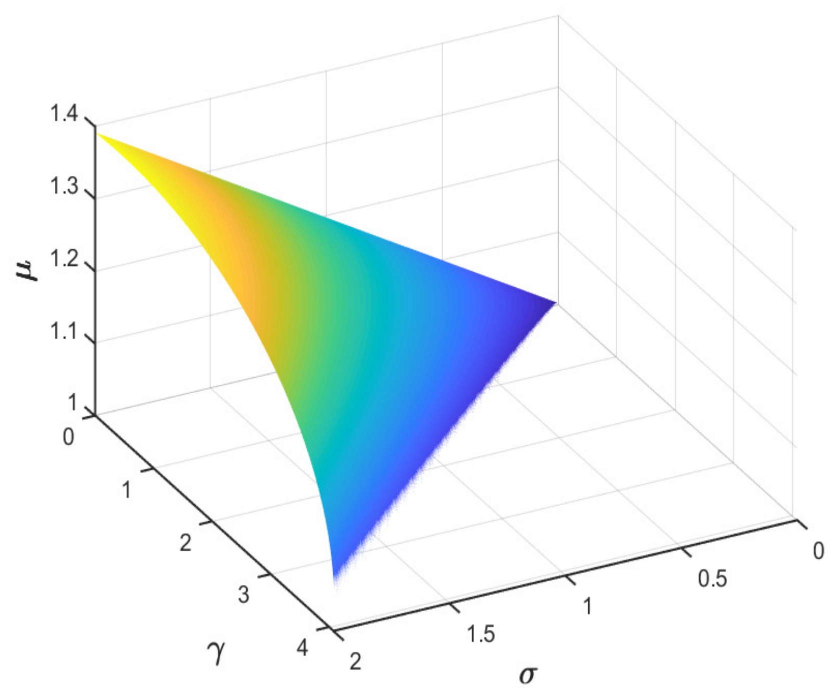
$$\frac{\partial^2 \mu_{\boldsymbol{\omega}}}{\partial^2 \gamma_{\boldsymbol{\omega}}} = -\frac{\sqrt{hg-k^2}}{g}\left(\frac{\gamma_{\boldsymbol{\omega}}^2}{\left(\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2\right)^{\frac{3}{2}}} + \left(\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2\right)^{-\frac{1}{2}}\right). \tag{2.4.13}$$

Clearly, the second-order derivative is always strictly negative for $\gamma_{\boldsymbol{\omega}} \in (-\sigma_{\boldsymbol{\omega}}\sqrt{g}, \sigma_{\boldsymbol{\omega}}\sqrt{g})$. Hence, we conclude that for a given value for $\sigma_{\boldsymbol{\omega}}$, $\mu_{\boldsymbol{\omega}}$ is a strictly concave function of $\gamma_{\boldsymbol{\omega}}$, and an easy calculation shows it attains its maximum when $\gamma_{\boldsymbol{\omega}} = \sigma_{\boldsymbol{\omega}}\frac{k}{\sqrt{h}}$. Let now $\mathcal{B}^*$ be the subset of $\mathcal{B}_1^*$ in which we restrict $\gamma_{\boldsymbol{\omega}}$ to the interval $\left[\frac{k}{\sqrt{h}}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}}\right)$. Namely,

$$\mathcal{B}^* = \left\{ (\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \;\middle|\; \begin{array}{l} \mu_{\boldsymbol{\omega}} = W_0(1+r) + \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{hg-k^2}\sqrt{\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2}}{g}, \\ \sigma_{\boldsymbol{\omega}} > 0, \gamma_{\boldsymbol{\omega}} \in \left[\frac{k}{\sqrt{h}}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}}\right) \end{array} \right\}. \tag{2.4.14}$$

A graphical illustration of the set $\mathcal{B}^*$ is presented in Figure 2.1.

Figure 2.1: Set $\mathcal{B}^*$. Using the market parameters given in Table 2.3, this graph shows the shape of the "mean-skewness-variance" frontier $\mathcal{B}^*$.



Any "mean-skewness-variance" portfolio must strictly belong to $\mathcal{B}^*$. Indeed it cannot belong to $\mathcal{B}_1^* \setminus \mathcal{B}^*$, as in this case one can always find a portfolio in $\mathcal{B}^*$ with the same $\sigma_{\boldsymbol{\omega}}$, and higher values for $\gamma_{\boldsymbol{\omega}}$ and $\mu_{\boldsymbol{\omega}}$. Conversely, every portfolio in $\mathcal{B}^*$ is "mean-skewness-variance" efficient (note that $\mu_{\boldsymbol{\omega}}$ is decreasing in $\gamma_{\boldsymbol{\omega}}$). □

Proposition 2.4.4 (Three-fund theorem for FSD-investors)**.** *When $V(\cdot)$ satisfies Assumption 2.2.1 and when Problem* (2.4.5) *has a solution $\boldsymbol{\omega}^*$, then $(\mu_{\boldsymbol{\omega}^*}, \sigma_{\boldsymbol{\omega}^*}, \gamma_{\boldsymbol{\omega}^*})$ must be in the set $\mathcal{B}^*$. Furthermore, the optimum $\boldsymbol{\omega}^*$ is given as*

$$\boldsymbol{\omega}^* := \boldsymbol{\omega}^*_{\sigma_{\boldsymbol{\omega}^*},\gamma_{\boldsymbol{\omega}^*}} = \frac{\sqrt{\sigma^2_{\boldsymbol{\omega}^*} g - \gamma^2_{\boldsymbol{\omega}^*}}}{\sqrt{hg-k^2}} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu} - r\mathbf{1}\right) - \left(\frac{k\sqrt{\sigma^2_{\boldsymbol{\omega}^*} g - \gamma^2_{\boldsymbol{\omega}^*}}}{g\sqrt{hg-k^2}} - \frac{\gamma_{\boldsymbol{\omega}^*}}{g} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma},$$

where $\sigma_{\boldsymbol{\omega}^*} > 0$ and $\gamma_{\boldsymbol{\omega}^*} \in \left[\frac{k}{\sqrt{h}}\sigma_{\boldsymbol{\omega}^*}, \sqrt{g}\sigma_{\boldsymbol{\omega}^*}\right)$ are chosen such that $V(W_{\boldsymbol{\omega}^*})$ is maximum.

Proof. Let $\boldsymbol{\omega}_*$ be the solution to Problem (2.4.5), i.e., the optimal portfolio for the objective function $V(\cdot)$, that is consistent with FSD. Let $(\mu_{\boldsymbol{\omega}^*}, \sigma_{\boldsymbol{\omega}^*}, \gamma_{\boldsymbol{\omega}^*})$ be the parameters of the terminal wealth determined by $\boldsymbol{\omega}_*$. If $(\mu_{\boldsymbol{\omega}^*}, \sigma_{\boldsymbol{\omega}^*}, \gamma_{\boldsymbol{\omega}^*})$ is not in $\mathcal{B}_1^*$, then it is possible to construct a portfolio $\boldsymbol{\omega}'$, that has the same parameters $\sigma_{\boldsymbol{\omega}^*}$ and $\gamma_{\boldsymbol{\omega}^*}$, but a strictly higher parameter $\mu_{\boldsymbol{\omega}'} > \mu_{\boldsymbol{\omega}^*}$. The portfolio $\boldsymbol{\omega}'$ strictly dominates $\boldsymbol{\omega}^*$ in FSD. As $V(\cdot)$ is consistent with FSD, this implies $V(W_{\boldsymbol{\omega}'}) > V(W_{\boldsymbol{\omega}^*})$, which violates the hypothesis of optimality of $\boldsymbol{\omega}^*$. If $\boldsymbol{\omega}_*$ is in $\mathcal{B}_1^* \setminus \mathcal{B}^*$ then we can find a portfolio in $\mathcal{B}^*$ with the same $\sigma_{\boldsymbol{\omega}}$, a higher $\gamma_{\boldsymbol{\omega}}$ and a higher or equal $\mu_{\boldsymbol{\omega}}$, which violates again the hypothesis of optimality of $\boldsymbol{\omega}^*$. Finally, the expression of a "mean-skewness-variance" efficient portfolio $\boldsymbol{\omega}^*$ is given in (2.4.10), which ends the proof. □

From Proposition 2.4.4, the optimal portfolio thus consists in investing part of the initial wealth in the risk-free asset and another part in a linear combination of two funds, $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mathbf{r1})$ and $\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$. Moreover, the composition of these two funds does not depend on the investor's preferences. The preferences of the investor translate into optimal weights that are allocated to both funds. Thus, the proposition also implies that solving the high-dimensional optimal portfolio problem in (2.2.2) can be reduced to solving a two-dimensional problem in $\mathbb{R}^2$ (i.e., the two optimal weights to be determined). Moreover, Proposition 2.4.4 shows that a three-fund theorem holds in any setting where the distribution of asset returns exhibits a location-scale mixture property and where preferences are FSD-consistent. Therefore, we recover various results in the literature in which three-fund theorems have been derived under specific assumptions on distributions and preferences. Birge and Chavez-Bedoya (2016) and Birge and Chavez-Bedoya (2020) derive a three-fund theorem under expected utility theory with exponential utility, using t-skewed returns and GH returns, respectively. As yet another example, when preferences are according to cumulative prospect theory and returns follow a t-skewed distribution Kwak and Pirvu (2018) show three-fund separation. All these results now immediately follow from Proposition 2.4.4.

Remark 2.4.1 (Connection with Mencía and Sentana (2009).)**.** *In Proposition 2.4.3, we show that all optimal portfolios for a FSD-investor belong to set $\mathcal{B}^*$, the "mean-skewness-variance" efficiency frontier. Observe that a mean-variance-skewness frontier was also derived in Mencía and Sentana (2009) under the assumption that asset returns are distributed according to a location-scale mixture of normals, a case that is included*[4] *in our set-up. Mencía and Sentana (2009)*

[4] The location-scale mixture of normals (LSMN) considered in Mencía and Sentana (2009) can be seen as a special case of location-scale mixtures (Definition 2.4.1) for which $Z \sim N(0, 1)$.

obtain a three-fund separation for the portfolios maximizing "skewness" $\gamma_{\boldsymbol{\omega}}$, given $\mu_{\boldsymbol{\omega}}$ and $\sigma_{\boldsymbol{\omega}}$. Specifically, using our parametrization they solve a problem that can equivalently expressed as

$$\max_{\boldsymbol{\omega}} \ \gamma_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \sigma, \mu_{\boldsymbol{\omega}} = \mu. \tag{2.4.15}$$

In their Proposition 4, Mencía and Sentana (2009) call all the portfolios that solve (2.4.15) *mean-variance-skewness efficient portfolios, and show that these portfolios can be expressed as a linear combination of the risk-free asset and the two funds $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{1})$ and $\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$. Observe that any portfolio in $\mathcal{B}^*$ is also a solution of Problem* (2.4.15). *Specifically, in Appendix 2.7.2 we show that $\mathcal{B}^*$ is a strict subset of the mean-variance-skewness frontier derived in Mencía and Sentana (2009). Thus, assuming that the objective is FSD-consistent makes it possible to reduce further the set of all mean-variance-skewness efficient portfolios, as in Mencía and Sentana (2009). Furthermore, additional assumptions on $V(\cdot)$ may lead to sets of optimal portfolios that are even smaller. For instance, Birge and Chavez-Bedoya (2020) prove that the set of optimal portfolios for investors maximizing expected exponential utility (denoted as Q-KE frontier) can be described using only two parameters, instead of three, as it is the case for $\mathcal{B}^*$. Interestingly, the Q-KE frontier has a shape that resembles the mean-variance efficient frontier.*

Remark 2.4.2 (Three-fund theorem extension.)**.** *Some of the results obtained so far for FSD-consistent objective functions can be extended under the weaker assumption that $V(\cdot)$ is μ-increasing only. We say that an objective function $V(\cdot)$ is μ-increasing if given two portfolios $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ such that $\sigma_{\boldsymbol{\omega}} = \sigma'_{\boldsymbol{\omega}}$, $\gamma_{\boldsymbol{\omega}} = \gamma'_{\boldsymbol{\omega}}$ and $\mu_{\boldsymbol{\omega}} \geqslant \mu'_{\boldsymbol{\omega}}$ we have that $V(W_{\boldsymbol{\omega}}) \geqslant V(W_{\boldsymbol{\omega}'})$. A μ-increasing objective function reflects the investor's preferences towards the distribution with a higher location parameter (ceteris paribus). It is clear that if $V(\cdot)$ is consistent with FSD then it is also μ-increasing, but the opposite is not true in general. Under the same distributional assumption as in Proposition 2.4.4, the optimal portfolios of μ-increasing objective functions exhibit a three-fund separation in the sense that part of the initial wealth is invested in the risk-free asset and part in a linear combination of the two funds $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{1})$ and $\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$. This feature follows from the fact that an optimal portfolio for a μ-increasing investor needs to be a solution of Problem* (2.4.9). *Contrary to the case of FSD optimal portfolios, a portfolio that is optimal for a μ-increasing investors, belongs to set $\mathcal{B}_1^*$, defined in* (2.4.11), *but not necessarily to $\mathcal{B}^*$, the "mean-skewness-variance" frontier. To shed more light on this, let us consider the following class of objective functions:*

$$V_h(W_{\boldsymbol{\omega}}) = \mathbb{E}(W_{\boldsymbol{\omega}}) + \sum_{j=1}^{h} a_j \mathbb{E}[(W_{\boldsymbol{\omega}} - \mathbb{E}(W_{\boldsymbol{\omega}}))^j],$$

with $h \in \mathbb{N}$ and $a_j \in \mathbb{R}$. Functions in the form of V_h can describe preference towards the central moment of the distribution, (for example, $a_2 < 0$ reflects variance aversion and $a_3 > 0$ implies positive skewness seeking) and are consistent with a Taylor approximation of expected utility around $\mathbb{E}(W_{\boldsymbol{\omega}})$. Under the location-scale mixture distributional assumption, $\mathbb{E}(W_{\boldsymbol{\omega}})$ is an increasing function of $\mu_{\boldsymbol{\omega}}$, while $\mathbb{E}[(W_{\boldsymbol{\omega}} - \mathbb{E}(W_{\boldsymbol{\omega}}))^j]$ do not depend on the location parameter. Hence, $V_h(W_{\boldsymbol{\omega}})$ is always μ-increasing and three-fund separation holds. More generally, consider the case where $V(\cdot)$ can be expressed as the difference between a "mean" function $f_1(\mu_{\boldsymbol{\omega}})$ and a "risk" function $f_2(\sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}})$, i.e., $V(W_{\boldsymbol{\omega}}) = f_1(\mu_{\boldsymbol{\omega}}) - f_2(\sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}})$. This kind of performance decomposition *appeared in Section 3.2 of Birge and Chavez-Bedoya (2020) in the context of portfolio selection for an exponential utility maximizer. Again, even if $V(\cdot)$ is not consistent with FSD, as soon as $f_1(\mu_{\boldsymbol{\omega}})$ is a non-decreasing function we have that three-fund separation applies.*

2.4.2 Optimal Portfolios for SSD-investors

Recall that SSD-investors are also FSD-investors and Proposition 2.3.5 thus also applies to SSD-investors. In this section, we explore whether for SSD-investors a more specific characterization can be derived for their optimal portfolio. In this regard, it can be shown (see e.g., Appendix 2.7.3) that SSD-investors choose a portfolio that belongs to the set $\mathcal{C}^*$ of portfolios that solve the auxiliary problem

$$\min_{\boldsymbol{\omega}} \ \sigma_{\boldsymbol{\omega}} \text{ subject to } \mu_{\boldsymbol{\omega}} = \mu, \gamma_{\boldsymbol{\omega}} = \gamma. \tag{2.4.16}$$

Hence the optimal portfolios for SSD-investors are to be found in the intersection $\mathcal{C}^* \cap \mathcal{B}^*$, limiting the set of admissible portfolios and perhaps also leading to a more refined characterization for the optimal portfolio under SSD-consistent preferences. However, for every couple $(\mu_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}})$, the value of $\sigma_{\boldsymbol{\omega}}$ such that $(\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \in \mathcal{B}^*$ is uniquely determined (see Appendix 2.7.4), and thus we get that $\mathcal{B}^*$ is a subset of $\mathcal{C}^*$ and hence $\mathcal{B}^* = \mathcal{C}^* \cap \mathcal{B}^*$ and it does not seem useful to further characterize $\mathcal{C}^*$.

Nevertheless, the following result shows that under certain market conditions, $\mathcal{B}^*$ is actually too broad in that it contains portfolios that cannot be optimal for an SSD-investor.

Proposition 2.4.5. *If $k + \mathbb{E}(Y)g < 0$, there exist portfolios in $\mathcal{B}^*$ having an expected return that is lower than the return given by the risk-free investment. Specifically, if $k + \mathbb{E}(Y)g < 0$ then for $(\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \in \mathcal{B}^*$ it holds that*

$$\mathbb{E}(W_{\boldsymbol{\omega}}) < W_0(1+r) \iff \gamma_{\boldsymbol{\omega}} \in \left(\sigma_{\boldsymbol{\omega}}\sqrt{g}\frac{\sqrt{hg-k^2}}{\sqrt{(k+\mathbb{E}(Y)g)^2 + hg - k^2}}, \sigma_{\boldsymbol{\omega}}\sqrt{g} \right). \tag{2.4.17}$$

Proof. Assume $k + \mathbb{E}(Y)g < 0$. Proposition 2.4.3 shows that in $\mathcal{B}^*$ there exists a specific relationship between the parameter $\mu_{\boldsymbol{\omega}}$ and the parameters $(\sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}})$. Here we use it to characterize portfolios in $\mathcal{B}^*$ with an expected return that is lower than the return given by the risk-free investment.

$$\begin{aligned}
\mathbb{E}(W_{\boldsymbol{\omega}}) < W_0(1+r) &\iff \mu_{\boldsymbol{\omega}} + \mathbb{E}(Y)\gamma_{\boldsymbol{\omega}} < W_0(1+r) \\
&\iff W_0(1+r) + \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{hg-k^2}\sqrt{\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2}}{g} + \mathbb{E}(Y)\gamma_{\boldsymbol{\omega}} < W_0(1+r) \\
&\iff \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{hg-k^2}\sqrt{\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2}}{g} + \mathbb{E}(Y)\gamma_{\boldsymbol{\omega}} < 0 \\
&\iff \gamma_{\boldsymbol{\omega}}(k + \mathbb{E}(Y)g) < -\sqrt{hg-k^2}\sqrt{\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2}.
\end{aligned}$$

Since the right side of the last inequality is negative, it follows from the assumption $k+\mathbb{E}(Y)g < 0$

that if $\gamma_{\boldsymbol{\omega}} \leqslant 0$ then $\mathbb{E}(W_{\boldsymbol{\omega}}) \geqslant W_0(1+r)$. Next, we consider the case $\gamma_{\boldsymbol{\omega}} > 0$,

$$\begin{aligned}
\mathbb{E}(W_{\boldsymbol{\omega}}) < W_0(1+r) &\iff \gamma_{\boldsymbol{\omega}} > -\frac{\sqrt{hg-k^2}\sqrt{\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2}}{k+\mathbb{E}(Y)g} > 0 \\
&\iff \gamma_{\boldsymbol{\omega}}^2 > \frac{(hg-k^2)(\sigma_{\boldsymbol{\omega}}^2 g - \gamma_{\boldsymbol{\omega}}^2)}{(k+\mathbb{E}(Y)g)^2} \\
&\iff \gamma_{\boldsymbol{\omega}}^2 > \left(\frac{(k+\mathbb{E}(Y)g)^2}{(k+\mathbb{E}(Y)g)^2 + hg - k^2}\right)\sigma_{\boldsymbol{\omega}}^2 g \frac{(hg-k^2)}{(k+\mathbb{E}(Y)g)^2} \\
&\iff \gamma_{\boldsymbol{\omega}}^2 > \sigma_{\boldsymbol{\omega}}^2 g \frac{hg-k^2}{(k+\mathbb{E}(Y)g)^2 + hg - k^2}.
\end{aligned}$$

Therefore, for the portfolios $W_{\boldsymbol{\omega}}$ in $\mathcal{B}^*$,

$$\mathbb{E}(W_{\boldsymbol{\omega}}) < W_0(1+r) \iff \gamma_{\boldsymbol{\omega}} \in \left(\sigma_{\boldsymbol{\omega}}\sqrt{g}\frac{\sqrt{hg-k^2}}{\sqrt{(k+\mathbb{E}(Y)g)^2 + hg - k^2}}, \sigma_{\boldsymbol{\omega}}\sqrt{g}\right).$$

□

2.5 Application

Bellman (2015) explained that to optimize an n-dimensional function on a continuous domain by exhaustively searching a discrete grid (obtained by a crude discretization), one could easily end up with making trillions of evaluations of the function. This is what he called "curse of dimensionality." Specifically, in the context of portfolio optimization, the initial dimensionality of the problem (i.e., the number n of assets) is typically in the range 30-1000. Hence, if one considers a grid of spacing 1/100 on the unit cube in 30 dimensions, we already have 100^{30} evaluations to make, which is not feasible in practice and out-rules the use of exhaustive enumeration strategies. It is well-known that computational tractability is greatly enhanced if the optimization problem at hand is convex, as in this case local optima are global optima, a feature, which allows local search algorithms to guarantee optimal solutions. However, demonstrating convexity is not always straightforward nor always true or desirable. For instance, if one aims to maximize a lower quantile of terminal wealth then one is using a non-convex objective that is however FSD-consistent.

A main contribution of this analysis is to show that under the rather flexible assumption of a multivariate location-scale mixture distribution of the asset returns, *any* optimization problem that is FSD-consistent can be readily approached using exhaustive search in a two-dimensional grid. Moreover, as the optimization problem is essentially of a two-dimensional nature, explicit solutions for concave objectives might be in reach or at least they can be easily obtained numerically.

We illustrate both features by revisiting a portfolio optimization problem that was also considered in Birge and Chavez-Bedoya (2016). Specifically, we show that the framework developed in this chapter makes it indeed possible to transform their n-dimensional (concave) portfolio optimization problem into a concave two-dimensional problem, which is easier to deal with. Furthermore, we show that an exhaustive search can also provide the solution in a very fast way and is straightforward to implement.

2.5.1 Concave optimization and exhaustive search

Birge and Chavez-Bedoya (2016) assume that the vector $\mathbf{X}$ of asset returns follows a so-called Generalized Hyperbolic Skew-t distribution; that is, the variables Y and Z in the model (2.4.1) follow, respectively, an Inverse Gaussian distribution with parameter $\nu > 3$, and a Gaussian distribution. Hence, $W_{\boldsymbol{\omega}} \sim Skew{-}t(\nu, \mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}})$ in which $(\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}})$ are as given in (2.4.3). Furthermore, it is assumed that the investor preferences are described by an exponential utility, i.e. $U(x) = -e^{-ax}$, in which $a > 0$ is the risk-aversion coefficient. So either for their assumptions on preferences or on the assets returns distribution, their setting is included in ours. The moment generating function of $W_{\boldsymbol{\omega}}$ is well-known and given as

$$\mathbb{E}\left(e^{sW_{\boldsymbol{\omega}}}\right) = \frac{e^{\mu_{\boldsymbol{\omega}} s} 2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \left(-\nu(\sigma_{\boldsymbol{\omega}}^2 s^2 + 2s\gamma_{\boldsymbol{\omega}})\right)^{\frac{\nu}{4}} K_{\frac{\nu}{2}}\left(\sqrt{-\nu(\sigma_{\boldsymbol{\omega}}^2 s^2 + 2s\gamma_{\boldsymbol{\omega}})}\right), \tag{2.5.1}$$

which only exists for $s \in \mathbb{R}$ such that $2s\gamma_{\boldsymbol{\omega}} + \sigma_{\boldsymbol{\omega}}^2 s^2 \leqslant 0$. The expected exponential utility $\mathbb{E}\left(U(W_{\boldsymbol{\omega}})\right)$ is thus given as

$$\mathbb{E}\left(U(W_{\boldsymbol{\omega}})\right) = \frac{-e^{-\mu_{\boldsymbol{\omega}} a} 2^{1-\frac{\nu}{2}}}{\Gamma_{\boldsymbol{\omega}}(\frac{\nu}{2})} \left(-\nu(\sigma_{\boldsymbol{\omega}}^2 a^2 - 2a\gamma_{\boldsymbol{\omega}})\right)^{\frac{\nu}{4}} K_{\frac{\nu}{2}}\left(\sqrt{-\nu(\sigma_{\boldsymbol{\omega}}^2 a^2 - 2a\gamma_{\boldsymbol{\omega}})}\right), \tag{2.5.2}$$

in which a is such that $\sigma_{\boldsymbol{\omega}}^2 a^2 - 2a\gamma_{\boldsymbol{\omega}} \leqslant 0$.

To find the optimal portfolio Birge and Chavez-Bedoya (2016) show the concavity of the objective function w.r.t the weights ω_i. Sometimes, explicit solutions for the optimal vector of weights $\boldsymbol{\omega}^*$ can be obtained. In contrast, we directly find the vector $(\mu_{\boldsymbol{\omega}^*}, \sigma_{\boldsymbol{\omega}^*}, \gamma_{\boldsymbol{\omega}^*})$ in the set $\mathcal{B}^*$ that yields maximum expected utility and from this we infer the optimal vector of weights $\boldsymbol{\omega}^*$.

For the ease of exposition we further omit the subscripts $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^*$. Using the functional relationship between μ and (σ, γ) in $\mathcal{B}^*$ (see equation (2.4.7)), we obtain that the objective function (2.5.2) can be written as a function of (σ,γ) only. We denote it by $f(\sigma, \gamma)$, where

$$f(\sigma, \gamma) = \frac{-e^{-a\left(W_0(1+r)+\frac{k\gamma}{g}+\frac{\sqrt{hg-k^2}\sqrt{\sigma^2 g-\gamma^2}}{g}\right)} 2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \left(-\nu(\sigma^2 a^2 - 2a\gamma)\right)^{\frac{\nu}{4}} K_{\frac{\nu}{2}}\left(\sqrt{-\nu(\sigma^2 a^2 - 2a\gamma)}\right). \tag{2.5.3}$$

In Appendix 2.7.5, we show that this problem can be alternatively formulated as

$$\max_{(\beta,\gamma)\in D} q(\beta, \gamma), \tag{2.5.4}$$

where $q(\beta, \gamma)$ is two-dimensional concave function and the domain D is given as

$$D = \left\{ (\beta, \gamma) \;\middle|\; \gamma^2 - 2\frac{g}{a}\gamma + \beta \leqslant 0,\ \beta \geqslant 0 \right\}.$$

The definition and properties of $q(\beta, \gamma)$ are illustrated in Appendix 2.7.5. Once the optimal values β^* and γ^* are found, we obtain the optimal value $\sigma^* = \sqrt{\frac{\beta^* + \gamma^{*2}}{g}}$. The optimal values for μ^* and the weights ω_i^* as functions of σ^* and γ^* follow from equation (2.4.14) and Proposition 2.4.4, respectively. Hereafter, we illustrate that the solution obtained using the concave optimization corresponds closely to the one that we would obtain using exhaustive search.

Specifically, we use the market conditions considered in Section 4.3 of Vanduffel and Yao (2017), i.e., we consider the model $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5) \sim Skew\text{−}t(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ with parameters reported in Table 2.1. The initial wealth is $W_0 = 1$, the risk-free rate is set $r = 0.01/252$ and $\nu = 9.557292$. For the risk-aversion parameter a, we consider $a = 2$.

Table 2.1: Market parameters for five assets.

	μ_i	γ_i	σ_{1i}	σ_{2i}	σ_{3i}	σ_{4i}	σ_{5i}
X_1	-0.00863	0.01488	0.000569	0.000103	0.000116	0.000086	0.000107
X_2	-0.00181	0.00296	0.000103	0.000146	0.000061	0.000063	0.000071
X_3	-0.00122	0.00537	0.000116	0.000061	0.000182	0.000070	0.000079
X_4	0.00177	0.00035	0.000086	0.000063	0.000070	0.000118	0.000056
X_5	0.00122	0.00547	0.000107	0.000071	0.000079	0.000056	0.000147

In order to find the maximum of the objective function $f(\sigma, \gamma)$ given in equation (2.5.3) using an exhaustive search approach, we first describe the set of feasible solutions.

From equation (2.5.3), a couple (σ, γ) that belongs to the domain of $f(\sigma, \gamma)$ must satisfy $\gamma \geqslant \frac{a\sigma^2}{2}$. Furthermore, from the definition of $\mathcal{B}^*$ in (2.4.7), we have the additional constraint that $\gamma \in \left[\sigma\frac{k}{\sqrt{h}}, \sigma\sqrt{g}\right)$. Putting these conditions together, we obtain the set of feasible solutions S:

$$S = \left\{ (\sigma, \gamma) \mid \sigma \in \left(0, \frac{2\sqrt{g}}{a}\right),\ \gamma \in \left[\max\left\{\frac{a\sigma^2}{2}; \sigma\frac{k}{\sqrt{h}}\right\}, \sigma\sqrt{g}\right)\right\} \tag{2.5.5}$$

Observe that S is a bounded convex set in $\mathbb{R}^2$, which significantly eases the implementation of an exhaustive search.

In Table 2.2 we display the optimal values for μ^*, σ^* and γ^* obtained using exhaustive search and obtained using concave optimization. A two-dimensional grid is considered, with 2,000 values for σ and for each σ we obtain 2,000 values for γ, i.e., a total number of 4,000,000 points to be evaluated.

This example shows how the implementation of exhaustive search on a two-dimensional grid merely requires the derivation of the set of feasible points. Therefore, this approach can in principle be applied to any optimization problem within our framework, and this turns out to be particularly useful when the objective function is not concave.

Table 2.2: Optimal solutions under concave optimization and exhaustive search.

	$\mu_{\boldsymbol{\omega}}$	$\gamma_{\boldsymbol{\omega}}$	$\sigma_{\boldsymbol{\omega}}$
Concave optimization	1.0015	0.17275	0.29492
Exhaustive search	1.0015	0.17274	0.29494

2.5.2 Mean-variance approximations

Levy and Markowitz (1979) and Markowitz (2014) provide some theoretical support for the observation that the optimal portfolio of an investor who maximizes expected utility can always be approximated by a mean-variance efficient portfolio; see also Birge and Chavez-Bedoya (2016) for some numerical evidence. Note, however, that under the assumption of a location-scale mixture for the assets returns, the optimal portfolio is a combination of three funds and thus not two, like in the case of mean-variance efficient portfolios. Hence, it is not obvious that one always find a mean-variance efficient portfolio (two funds) that is very close to an EUT-optimal portfolio (three funds).

Here, we show that under the assumption of a location-scale mixture for the asset returns there may exist portfolios that are optimal for an exponential utility investor and that cannot be well approximated by a mean-variance efficient portfolio. To measure the distance $d\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right)$ between two portfolios $\boldsymbol{\omega}^1$ and $\boldsymbol{\omega}^2$, we use the Euclidean distance, i.e.,

$$d\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right) = \sqrt{\sum_{i=1}^{n}\left(\omega_i^1 - \omega_i^2\right)^2} \tag{2.5.6}$$

In Appendix 2.7.6, we explain how for a given portfolio $\boldsymbol{\omega}^*_{EU}(a)$ that is optimal for an expected utility maximizing investor with risk aversion coefficient a, one can determine the mean-variance efficient portfolio $\boldsymbol{\omega}^*_{MV}(a)$ that is closest, i.e., such that $d(\boldsymbol{\omega}^*_{EU}(a), \boldsymbol{\omega}^*_{MV}(a))$ becomes minimum among all mean-variance efficient portfolios $\boldsymbol{\omega}_{MV}$.

Figure 2.2 displays, the behavior of this minimum distance, first, for a range of values for the risk aversion parameter a, $d(\boldsymbol{\omega}^*_{EU}(a), \boldsymbol{\omega}^*_{MV}(a))^2$ in Panel A and then as a function of the parameter γ_2 of the second risky asset in Panel B in the case of $a = 1$. Unless otherwise specified, all market parameters are set as in Table 2.3.

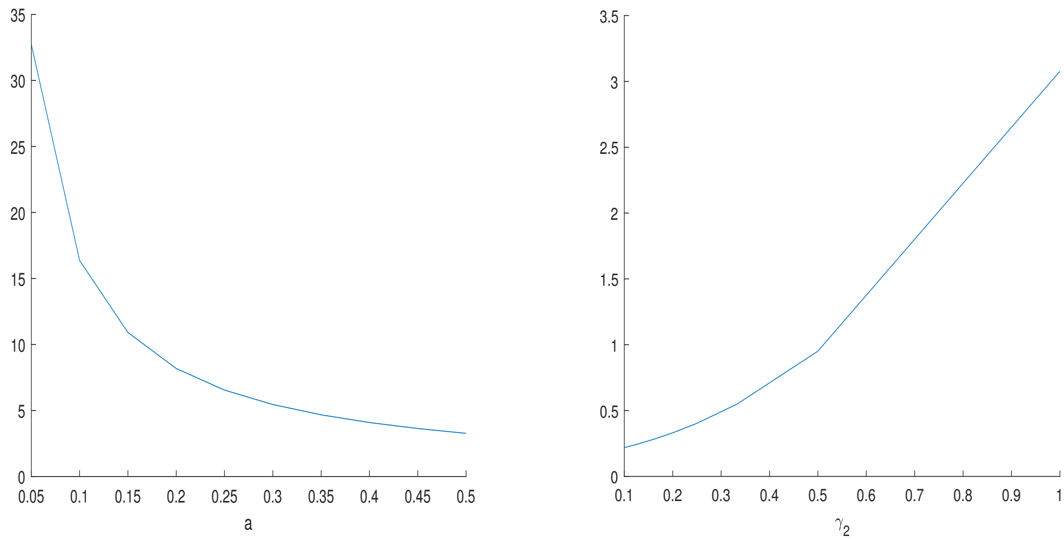
Table 2.3: Market parameters for the two-dimensional market.

	μ	γ	σ	ρ	ν
X_1	0.05	-0.3	0.3	0.3	5
X_2	0.08	0.7	0.5	0.3	5

Both panels in Figure 2.2 show that the minimum distance between the exponential utility maximizing portfolio and a mean-variance efficient portfolio can be significant. This situation occurs when the risk-aversion parameter a approaches 0 or when the parameter γ_2 of the second risk asset is big enough.

In Figures 2.3 and 2.4 we compare in more detail the composition of the EUT-optimal port-

Figure 2.2: Minimum distance between the optimal exponential utility maximizing portfolio and the closest mean-variance efficient portfolio as a function of the risk-aversion parameter a in Panel A and as a function of γ_2 in Panel B. In Panel A, for each value of a, we find the portfolio $\boldsymbol{\omega}^*_{EU}(a)$ that is optimal for the exponential utility investor and determine the mean-variance portfolio $\boldsymbol{\omega}^*_{MV}(a)$ that has minimum Euclidean distance with $\boldsymbol{\omega}^*_{EU}(a)$. In Panel B, for each value of the parameter γ_2, we find the portfolio $\boldsymbol{\omega}^*_{EU}(\gamma_2)$ that is optimal for the exponential utility investor and determine the mean-variance portfolio $\boldsymbol{\omega}^*_{MV}(\gamma_2)$ that has minimum Euclidean distance with $\boldsymbol{\omega}^*_{EU}(\gamma_2)$. The remaining market parameters are fixed according to Table 2.3. In both panels, we plot the squared minimum distances.

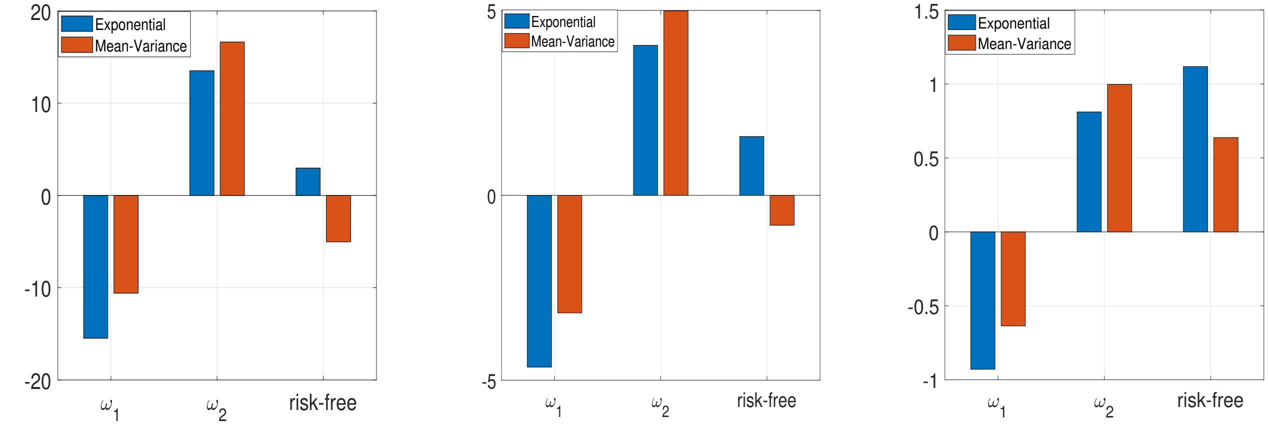


Panel A Panel B

folio with the composition of the closest mean-variance efficient portfolio. On the one hand, the portfolios exhibit similarities in that the decision of going long or short in an asset appears to be the same in both cases. On the other hand, the amounts invested in the various assets are clearly different and they are particularly different when the parameter a approaches zero or when γ_2 is increasing. This feature is consistent with the observations made in Figure 2.2. Note also that in all considered cases the exponential utility maximizer appears to invest more in the risk-free asset than the corresponding mean-variance maximizer.

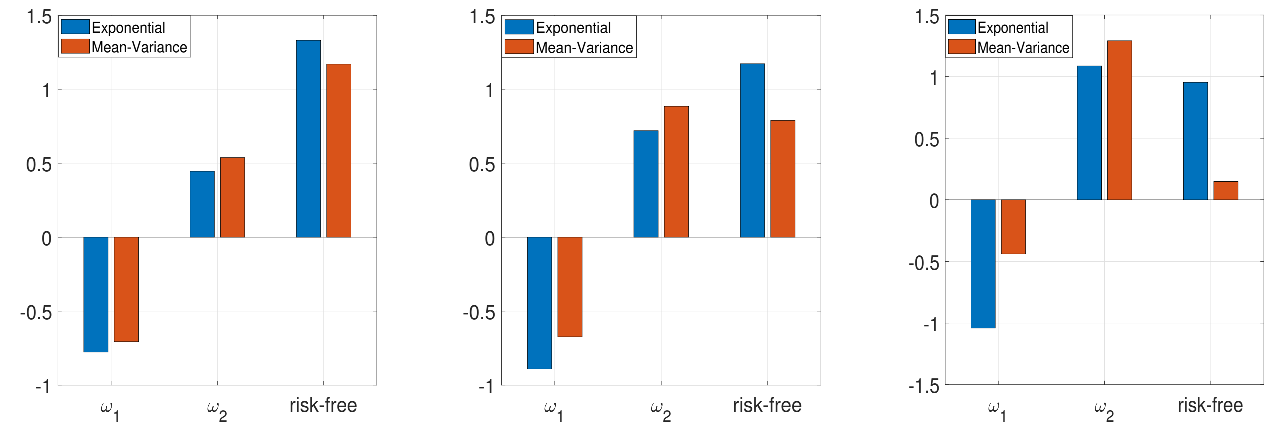
All in all, the examples provided in this numerical study indicate that optimal portfolios for expected utility investors cannot always be closely approximated by a mean-variance portfolio.

Figure 2.3: Comparing the portfolios as the parameter a takes three values $a = 0.3$, $a = 1$ and $a = 5$. Each optimal exponential utility portfolio is compared with the mean-variance portfolio that minimizes the Euclidean distance (2.5.6). The market parameters are given in Table 2.3. The respective weights in the risk-free asset and in the two risky assets are displayed as bars.



a=0.3 a=1 a=5

Figure 2.4: Comparing the portfolios as γ_2 takes three values $\gamma_2 = 0.3$, $\gamma_2 = 0.6$ and $\gamma_2 = 1$. Each optimal exponential utility portfolio is compared with the mean-variance portfolio that minimizes the Euclidean distance (2.5.6). The other market parameters are given in Table 2.3 and the risk aversion parameter is set to $a = 5$. The respective weights in the risk-free asset and in the two risky assets are displayed as bars.



$\gamma_2 = 0.3$ $\gamma_2 = 0.6$ $\gamma_2 = 1$

2.6 Final remarks

Under very weak conditions on the investor's preferences we provide conditions on the multivariate distribution of asset returns that lead to two-fund or three-fund separation. Specifically, a location-scale property of the multivariate distribution (elliptical distributions) leads to a two-fund separation for all investors with law-invariant and increasing preferences whereas a weaker conditional location-scale property implies three-fund separation. The latter condition is for instance satisfied by a multivariate hyperbolic distribution and is known to be fairly realistic for modeling real-world asset returns. Thus, the assumptions made on the assets'returns distribution are the key element to switch from a two-fund theorem to a three-fund theorem. The specificity of the objective function, which reflects the specific assumptions on the investor's preferences, is only used to select the optimal weights that are allocated to each of the two, respectively three funds.

Using this characterization of the optimal portfolio, it is then possible to significantly reduce the complexity of finding an optimal portfolio, as only a one-dimensional optimization in the case of two-fund separation (respectively a two-dimensional optimization in the case of three-fund separation) is needed even in a market setting in which there are thousands of assets.

Using our theoretical approach and general characterization of optimal portfolio under very general preferences, we are able to show that two-fund and three-fund theorems that have appeared in the literature and that were derived under specific assumptions on the investor's preferences and on the market setting are special cases of our general characterization results. Finally, we provide some evidence that the optimal portfolio of an expected utility maximizer cannot always be approximated by a mean-variance efficient portfolio even though this was claimed in the literature (e.g., Levy and Markowitz (1979) and Markowitz (2014)).

2.7 Appendix

2.7.1 Well-posedness of Problem (2.4.9)

The condition $\gamma_{\boldsymbol{\omega}} \in (-\sqrt{g}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}})$ is related to the well-posedness of the maximization problem in (25). Using the value g defined in (2.4.8), this condition can be equivalently expressed as follows:

$$\begin{aligned}\gamma_{\boldsymbol{\omega}} \in (-\sqrt{g}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}}) &\iff \sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}} > 0 \\ &\iff \left(\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\right)\left(\boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}\right) > \left(\boldsymbol{\omega}^T \boldsymbol{\gamma}\right)^2 .\end{aligned}$$

Since the matrix $\boldsymbol{\Sigma}$ is positive definite, using the extended Cauchy-Schwarz inequality, it can be shown that

$$\left(\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\right)\left(\boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}\right) \geqslant \left(\boldsymbol{\omega}^T \boldsymbol{\gamma}\right)^2 \tag{2.7.1}$$

holds true for all $\boldsymbol{\omega}$. Therefore, once we fix $\sigma_{\boldsymbol{\omega}} = \sigma$, $\gamma_{\boldsymbol{\omega}}$ must belong to the interval $[-\sqrt{g}\sigma, \sqrt{g}\sigma]$ otherwise the problem in (25) is ill-posed in that there are no portfolios $\boldsymbol{\omega}$ satisfying both constraints $\sigma_{\boldsymbol{\omega}} = \sigma$ and $\gamma_{\boldsymbol{\omega}} = \gamma$. With the condition $\gamma_{\boldsymbol{\omega}} \in (-\sqrt{g}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}})$, we are excluding the extreme cases $\gamma_{\boldsymbol{\omega}} = -\sqrt{g}\sigma_{\boldsymbol{\omega}}$ and $\gamma_{\boldsymbol{\omega}} = \sqrt{g}\sigma_{\boldsymbol{\omega}}$. Let us consider the case $\gamma_{\boldsymbol{\omega}} = \sqrt{g}\sigma_{\boldsymbol{\omega}}$. Note that

$$\begin{aligned}\gamma_{\boldsymbol{\omega}} = \sqrt{g}\sigma_{\boldsymbol{\omega}} &\implies \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} = \frac{\gamma^2_{\boldsymbol{\omega}}}{g} \\ &\implies \left(\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\right)\left(\boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}\right) = \left(\boldsymbol{\omega}^T \boldsymbol{\gamma}\right).^2\end{aligned}$$

Thus, the portfolios that satisfy $\gamma_{\boldsymbol{\omega}} = \sqrt{g}\sigma_{\boldsymbol{\omega}}$ must satisfy also $\left(\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\right)\left(\boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}\right) = \left(\boldsymbol{\omega}^T \boldsymbol{\gamma}\right)^2$. This last condition holds true if and only if there exists $c \in \mathbb{R}$ such that, $\boldsymbol{\omega}^T = c\boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1}$. Therefore, once we fix $\sigma_{\boldsymbol{\omega}} = \sigma$, there exists only one portfolio $\boldsymbol{\omega}$ such that the condition $\gamma_{\boldsymbol{\omega}} = \sqrt{g}\sigma_{\boldsymbol{\omega}}$ is satisfied and this portfolio is $\boldsymbol{\omega}^T = \frac{\sigma}{\sqrt{g}}\boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1}$. In a nutshell, the condition $\gamma_{\boldsymbol{\omega}} \in (-\sqrt{g}\sigma_{\boldsymbol{\omega}}, \sqrt{g}\sigma_{\boldsymbol{\omega}})$ is necessary for Problem (25) to be a true optimization problem.

2.7.2 Comparison with Mencía and Sentana (2009)

In this Appendix we show that set $\mathcal{B}^*$, derived in Proposition 2.4.3, is a strict subset of the mean-variance-skewness frontier derived in Mencía and Sentana (2009). First, we show that any portfolio in $\mathcal{B}^*$ is a solution of Problem (2.4.15). If $(\overline{\mu}, \overline{\sigma}, \overline{\gamma}) \in \mathcal{B}^*$, we know that $(\overline{\mu}, \overline{\sigma}, \overline{\gamma})$ must be a solution of Problem (2.4.9), namely

$$\overline{\mu} = \max_{\boldsymbol{\omega}} \ \mu_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \overline{\sigma}, \gamma_{\boldsymbol{\omega}} = \overline{\gamma}, \tag{2.7.2}$$

Additionally, in $\mathcal{B}^*$ the partial derivative of $\mu_{\boldsymbol{\omega}}$ with respect to $\gamma_{\boldsymbol{\omega}}$ is strictly negative and therefore

$$\text{for all } \gamma^* > \overline{\gamma},\ \overline{\mu} > \max_{\boldsymbol{\omega}}\ \mu_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \overline{\sigma}, \gamma_{\boldsymbol{\omega}} = \gamma^*. \tag{2.7.3}$$

We want to show that $(\overline{\mu}, \overline{\sigma}, \overline{\gamma})$ solve a problem in the form of (2.4.15). It is clear that

$$\overline{\gamma} \leqslant \max_{\boldsymbol{\omega}}\ \gamma_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \overline{\sigma}, \mu_{\boldsymbol{\omega}} = \overline{\mu}. \tag{2.7.4}$$

Now, assume

$$\overline{\gamma} < \gamma' = \max_{\boldsymbol{\omega}}\ \gamma_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \overline{\sigma}, \mu_{\boldsymbol{\omega}} = \overline{\mu}. \tag{2.7.5}$$

This implies that there exists a portfolio such that $\sigma_{\boldsymbol{\omega}} = \overline{\sigma}, \mu_{\boldsymbol{\omega}} = \overline{\mu}$ and $\gamma_{\boldsymbol{\omega}} = \gamma'$, with $\gamma' > \overline{\gamma}$. Thus, it must be true that

$$\exists \gamma^* > \overline{\gamma} \text{ such that } \overline{\mu} \leqslant \max_{\boldsymbol{\omega}}\ \mu_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \overline{\sigma}, \gamma_{\boldsymbol{\omega}} = \gamma^*. \tag{2.7.6}$$

Note that (2.7.6) is a contradiction of (2.7.3). Hence, assumption (2.7.5) cannot be true and we have that

$$\overline{\gamma} = \max_{\boldsymbol{\omega}}\ \gamma_{\boldsymbol{\omega}} \text{ subject to } \sigma_{\boldsymbol{\omega}} = \overline{\sigma}, \mu_{\boldsymbol{\omega}} = \overline{\mu}. \tag{2.7.7}$$

holds true for all portfolios $(\overline{\mu}, \overline{\sigma}, \overline{\gamma}) \in \mathcal{B}^*$. Therefore, $\mathcal{B}^*$ is a subset of the mean-variance efficient frontier derived in Mencía and Sentana (2009). To conclude the proof, it is sufficient to show that the mean-variance-skewness frontier of Mencía and Sentana (2009) contains some portfolios that are not in $\mathcal{B}^*$. In the proof of Proposition 2.4.3, we showed that not all solutions of Problem (2.4.9) (set $\mathcal{B}_1$ in equation (2.4.11)) are also solutions of Problem (2.4.15). Since problems (2.4.9) and (2.4.15) are specular[5], with a similar argument one can easily check that not all the solutions of Problem (2.4.15) are also solution of Problem (2.4.9) and therefore cannot be in $\mathcal{B}^*$. The conclusion is that $\mathcal{B}^*$ is a strict subset of the mean-variance-skewness frontier derived in Mencía and Sentana (2009).

2.7.3 SSD-investors optimal portfolios are in $\mathcal{C}^*$

Recall that any objective function $V(\cdot)$ that satisfies Assumption 2.2.2 is also consistent with convex order, i.e., given two portfolios W_1 and W_2 such that $W_1 \prec_{cx} W_2$, then $V(W_1) \geqslant V(W_2)$. Hence, to prove that all SSD-optimal portfolios are in $\mathcal{C}^*$ it is sufficient to prove that given two portfolios with location-scale mixture property, same parameter $\mu_{\boldsymbol{\omega}}$ and $\gamma_{\boldsymbol{\omega}}$ but different $\sigma_{\boldsymbol{\omega}}$, then the one with a lower $\sigma_{\boldsymbol{\omega}}$ is also lower in convex order, as stated in the next proposition.

Proposition 2.7.1. *Let Z be a random variable with $\mathbb{E}(Z) = 0$ and $std(Z) = 1$, and let $Y \geqslant 0$ a.s. be a random variable that is independent of Z. Let W_1 and W_2 be two random variables such that $W_1 \overset{d}{=} \mu + Y\gamma + \sqrt{Y}\sigma_1 Z$ and $W_2 \overset{d}{=} \mu + Y\gamma + \sqrt{Y}\sigma_2 Z$ with $\mu \in \mathbb{R}$, $\gamma \in \mathbb{R}$ and $0 < \sigma_1 < \sigma_2$.*

[5] Problems (2.4.9) and (2.4.15) are specular in the sense that we can go from one problem to another simply by switching $\gamma_{\boldsymbol{\omega}}$ with $\mu_{\boldsymbol{\omega}}$, or equivalently $\boldsymbol{\gamma}$ with $\boldsymbol{\mu} - r\mathbf{1}$.

Then, for all convex functions f,

$$\mathbb{E}(f(W_1)) \leqslant \mathbb{E}(f(W_2)) \tag{2.7.8}$$

Proof. First, observe that $\mathbb{E}(W_1) = \mathbb{E}(W_2)$. For $i = 1, 2$ and for all $y \geqslant 0$, let us denote with W_{iy} the random variable $W_i | Y = y$. The expected value and the standard deviation of W_{iy}, will be denoted with $m_y = \mu + y\gamma$ and $\delta_{iy} = \sqrt{y}\sigma_i$, respectively. Under our hypothesis on the distribution of W_1 and W_2, we can write

$$\begin{aligned} W_{iy} &\stackrel{d}{=} \mu + y\gamma + \sqrt{y}\sigma_i Z \\ &\stackrel{d}{=} m_y + \delta_{iy} Z \end{aligned}$$

It is clear that for all convex functions f,

$$\mathbb{E}(f(W_1)) = \mathbb{E}(\mathbb{E}(f(W_1|Y=y))) \leqslant \mathbb{E}(\mathbb{E}(f(W_2|Y=y))) = \mathbb{E}(f(W_2))$$

□

2.7.4 Uniqueness of $\sigma^2_{\boldsymbol{\omega}}$ in $\mathcal{B}^*$

Proof. From the definition of the set $\mathcal{B}^*$ (Proposition 2.4.3) we can deduce that for any couple $(\mu_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}})$, there exists a unique value of $\sigma^2_{\boldsymbol{\omega}}$ such that $(\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \in \mathcal{B}^*$. To see this point, we can simply invert the functional relationship given in equation (2.4.7).

$$\begin{aligned} (\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \in \mathcal{B}^* &\implies \mu_{\boldsymbol{\omega}} = W_0(1+r) + \frac{k\gamma_{\boldsymbol{\omega}}}{g} + \frac{\sqrt{hg-k^2}\sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}}}{g} \\ &\implies \frac{(\mu_{\boldsymbol{\omega}} - W_0(1+r))g - k\gamma_{\boldsymbol{\omega}}}{\sqrt{hg-k^2}} = \sqrt{\sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}}} \end{aligned}$$

Thus

$$\begin{aligned} (\mu_{\boldsymbol{\omega}}, \sigma_{\boldsymbol{\omega}}, \gamma_{\boldsymbol{\omega}}) \in \mathcal{B}^* &\implies \frac{((\mu_{\boldsymbol{\omega}} - W_0(1+r))g - k\gamma_{\boldsymbol{\omega}})^2}{hg-k^2} = \sigma^2_{\boldsymbol{\omega}} g - \gamma^2_{\boldsymbol{\omega}} \\ &\implies \frac{\left((\mu_{\boldsymbol{\omega}} - W_0(1+r))\sqrt{g} - \frac{k\gamma_{\boldsymbol{\omega}}}{\sqrt{g}}\right)^2}{hg-k^2} + \frac{\gamma^2_{\boldsymbol{\omega}}}{g} = \sigma^2_{\boldsymbol{\omega}} \\ &\implies \frac{\left((W_0(1+r) - \mu_{\boldsymbol{\omega}})\sqrt{g} + \frac{k\gamma_{\boldsymbol{\omega}}}{\sqrt{g}}\right)^2}{hg-k^2} + \frac{\gamma^2_{\boldsymbol{\omega}}}{g} = \sigma^2_{\boldsymbol{\omega}}. \end{aligned}$$

□

2.7.5 Proof of Concavity for Problem (2.5.4)

Proof. Considering $f(\sigma,\gamma)$ in (2.5.3), we switch the parametrization from (σ,γ) to (β,γ), with $\beta = \sigma^2 g - \gamma^2$ and obtain the following portfolio optimization problem :

$$\max_{(\beta,\gamma)\in D} f(\beta,\gamma), \tag{2.7.9}$$

in which $f(\beta,\gamma)$ is given by

$$\frac{-e^{-a\left(W_0(1+r)+\frac{k\gamma}{g}+\frac{\sqrt{hg-k^2}\sqrt{\beta}}{g}\right)}2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\left(-\nu\left(\frac{\beta+\gamma^2}{g}a^2-2a\gamma\right)\right)^{\frac{\nu}{4}}K_{\frac{\nu}{2}}\left(\sqrt{-\nu\left(\frac{\beta+\gamma^2}{g}a^2-2a\gamma\right)}\right),$$

and where the domain D is given as

$$D=\left\{(\beta,\gamma)\;\middle|\;\gamma^2-2\frac{g}{a}\gamma+\beta\leqslant 0,\ \beta\geqslant 0\right\}.$$

To find a solution to Problem (2.7.9), we will in a first step rewrite $f(\beta,\gamma)$ as an easier-to-deal-with function with the same maximum. To this end, consider the function

$$\theta(y)=\frac{-\ln(-y)+\ln(2^{1-\frac{\nu}{2}})-\ln(\Gamma(\frac{\nu}{2}))}{a}-W_0(1+r).$$

Since $\theta(y)$ is increasing for $y<0$, $q(\beta,\gamma):=\theta(f(\beta,\gamma))$ has the same maximum and domain as $f(\beta,\gamma)$. Furthermore, $q(\beta,\gamma)$ writes as

$$q(\beta,\gamma)=\theta(f(\beta,\gamma))=\frac{k\gamma}{g}+\frac{\sqrt{hg-k^2}\sqrt{\beta}}{g}-\frac{g(\beta,\gamma)}{a},$$

in which

$$g(\beta,\gamma)=\ln\left(\left(-\nu\left(\frac{\beta+\gamma^2}{g}a^2-2a\gamma\right)\right)^{\frac{\nu}{4}}K_{\frac{\nu}{2}}\left(\sqrt{-\nu\left(\frac{\beta+\gamma^2}{g}a^2-2a\gamma\right)}\right)\right).$$

In a second step we prove that $q(\beta,\gamma)$ is concave w.r.t. the variables (β,γ). We start by proving that $-g(\beta,\gamma)$ is concave. In this regard, note that $-g(\beta,\gamma)=r(A(\beta,\gamma))$, where

$$r(y)=-\ln\left(y^{\frac{\nu}{4}}K_{\frac{\nu}{2}}(\sqrt{y})\right),\qquad A(\beta,\gamma)=-\nu\left(\frac{\beta+\gamma^2}{g}a^2-2a\gamma\right).$$

In their Appendix C1, Birge and Chavez-Bedoya (2016) prove that $r(y)$ is an increasing and concave function. Furthermore, as $A(\beta,\gamma)$ is concave as well, $-g(\beta,\gamma)$ is concave. Finally, $q(\beta,\gamma)$ can be seen as a sum of concave functions, so it is concave itself. □

2.7.6 Minimum distance to mean-variance efficient portfolios

Proof. In our numerical study, each portfolio $\boldsymbol{\omega}_{EU}^*$ optimal for an exponential utility investor is compared with the mean-variance portfolio that minimizes the Euclidean distance (2.5.6) with $\boldsymbol{\omega}_{EU}^*$. In this appendix, we illustrate how, given a generic n-dimensional vector $\boldsymbol{x}$ interpretable as a benchmark portfolio, it is possible to determine the mean-variance efficient portfolio that has minimum Euclidean distance with $\boldsymbol{x}$.

Considering n risky assets whose returns have a positive definite covariance matrix $\boldsymbol{\Delta}$ and an expected value $\boldsymbol{m}$, all mean-variance efficient portfolios can be written as

$$\boldsymbol{\omega}_\delta^* = \frac{\delta}{\sqrt{h}}\boldsymbol{\Delta}^{-1}(\boldsymbol{m} - r\mathbf{1}).$$

Therefore, $\boldsymbol{\omega}_\delta^* = \delta\boldsymbol{\omega}_1^*, \forall\delta > 0$. Our aim is to solve the following problem:

$$\min_\delta \sum_{i=1}^n (x_i - \delta\omega_{1i}^*)^2$$

Let $d(\delta)$ be the function we aim to minimize, i.e. $d(\delta) = \sum_{i=1}^n (x_i - \delta\omega_{1i}^*)^2$. The first and the second-order derivatives of $d(\delta)$ are

$$d'(\delta) = \sum_{i=1}^n -2\omega_{1i}^* (x_i - \delta\omega_{1i}^*), \quad d''(\delta) = 2\sum_{i=1}^n \omega_{1i}^{*2}$$

Assuming $\boldsymbol{\omega}_1^* \neq \mathbf{0}$, $d(\delta)$ is strictly convex. To find the minimum we look for the value of δ such that $d'(\delta) = 0$, namely

$$d'(\delta) = 0 \iff \sum_{i=1}^n -2\omega_{1i}^* (x_i - \delta\omega_{1i}^*) = 0 \iff \sum_{i=1}^n \omega_{1i}^{*2}\delta = \sum_{i=1}^n \omega_i^* x_i \iff \delta = \frac{\sum_{i=1}^n \omega_{1i}^* x_i}{\sum_{i=1}^n \omega_{1i}^{*2}}$$

Hence, $\frac{\sum_{i=1}^n \omega_{1i}^* x_i}{\sum_{i=1}^n \omega_{1i}^{*2}}\boldsymbol{\omega}_1^*$ is the mean-variance efficient portfolio with minimum Euclidean distance with the vector $\boldsymbol{x}$. □

Chapter 3

The impact of correlation on (Range)Value-at-Risk

3.1 Introduction

Insurance operations are based on the existence of risk diversification among risks and insurers' capital requirements should therefore reflect the dependence and diversification effects among the different risks they face. Regulatory capital frameworks, such as Basel III and Solvency II, acknowledge diversification in that they make it possible for a company to employ correlations for determining their capital. In a first step, the stand-alone capital requirements for the various risks are determined and in a second step a correlation matrix is used to determine the aggregate capital requirement hereby using the so-called square-root formula (Christiansen et al. (2012)).

Embrechts et al. (2002) clarify the role of copulas as a proper concept for reflecting dependence among risks. Moreover, these authors warn that knowledge of correlations can give a false feeling of security in that they may not well reflect the true dependence, and thus lead to an underestimation of the true risk. This is because in general for a given set of correlations, several copulas that preserve the correlations will exist and each of these copulas will give rise to one particular loss distribution. Chernih et al. (2010) illustrate this feature in the context of credit portfolio risk assessment. They build a portfolio model that uses exactly the same input parameters as industry standard (Moody's KMV) and show that the portfolio Value-at-Risk (VaR) - which translates into the capital requirement - can be orders of magnitude higher than what is computed under the industry standard. Such examples clearly demonstrate that model error is a real concern and that it is useful to determine the worst-case model and best-case model, i.e., the models that yields the highest resp. lowest possible value for the risk measure at hand given the information that is available.

The problem of finding VaR bounds, i.e., the maximum and minimum VaR, when the marginal distributions are fixed but no dependence information is assumed, has attracted a considerable interest in the actuarial literature. In the case of two risks, this problem has been completely solved in Makarov (1981) and Rüschendorf (1982). In arbitrary dimensions, it is more complicated and a general solution is generally not available. Nevertheless, important results have been obtained in this direction. For identically distributed risks (homogeneous case), some explicit formulas for the sharp VaR bounds have been proposed for instance in Wang and Wang (2011), Puccetti and Rüschendorf (2013b) and Wang et al. (2013). In the inhomogeneous case, Puccetti and Rüschendorf (2012) and Embrechts et al. (2013) developed the Rearrangement Algorithm[1] (RA), which turns out to be a highly efficient numerical procedure (Hofert et al. (2017), Bernard and McLeish (2016)).

In any case, the bounds that are obtained in absence of using dependence information (unconstrained bounds) are typically very wide, and this has motivated further research on VaR bounds when some partial knowledge on the dependence structure (copula) is assumed (constrained bounds). Some first results in this direction were obtained in Embrechts et al. (2003) and Embrechts and Puccetti (2006), where copula bounds are assumed to be known. Many authors noticed that even when some knowledge on the copula is assumed, the risk of underestimating the VaR at a high confidence level remains significant and attempted to quantify the level of underestimation as a function of dependence information. Among others, problems of this kind were studied for instance in Bignozzi et al. (2015), where the authors showed that adding a positive dependence restriction leaves the VaR upper bound considerably high. Bernard and Vanduffel (2015) reached similar conclusions, assuming that the copula of interest is known only on a subset of its domain.

In the banking and insurance industry, the dependence between two or more risks is often described via the Pearson correlation. Therefore, it is of practical interest to study how bounds on VaR and its generalization Range Value-at-Risk (RVaR) are affected when such dependence information is available. With the same goal in mind, Bernard et al. (2017) obtained some (non sharp) bounds on VaR using the variance of the sum as source of dependence information. Kaas et al. (2009) have studied a similar problem as the one we consider in this paper and in Section 3.2.1 we compare their results with ours. Similarly to Kaas et al. (2009), our study should be considered as a pedagogic warning that the knowledge of one or more dependence measures does not automatically translate into a lower worst-case VaR, with respect to the case when no dependence information is available.

Our contributions can be summarized as follows. In Section 3.2, we study best-possible bounds on VaR and Tail-Value-at-Risk (TVaR) for a sum of two risks in the presence of information about a measure of association (dependence) with focus on three well-known dependence measures: Spearman's rho, Kendall's tau and Pearson correlation. We show that when the dependence constraint takes value in a certain interval, which we can specify, then the constrained upper bounds coincide with the unconstrained ones. A similar result holds for the case of lower bounds. For probability levels used in risk management, however, the interval is wide for the case of upper bounds and narrow for the case of lower bounds. Hence, using correlation information has typically no effect on the highest possible VaR whereas it can affect the lowest possible VaR.

In Section 3.3, we study the more general problem of finding the best-possible upper and lower bounds for the Range-Value-at-Risk[2] (RVaR) of the sum of $n \in \mathbb{N}$ risks, when the knowledge of

[1] The RA is not only a reference tool for assessing risk bounds numerically, further developments and generalizations of this algorithm have also been successfully applied in operations research to various allocation and synchronization problems (Boudt et al. (2018), Cornilly et al. (2022))

[2] This risk measure is a generalization of TVaR and includes VaR and TVaR as special cases.

the marginal distributions is assumed and some information on the average correlation is given. We derive bounds that are in general not best-possible but provide a condition under which they are best-possible. We also explicitly derive the best-possible constrained RVaR bounds for the case of a sum of $n \geqslant 3$ uniformly distributed risks. As far as we know, this result is the first that provides an explicit best-possible bound on RVaR for a general sum of $n \geqslant 3$ risks (uniformly distributed) under a correlation constraint. Indeed, even in the unconstrained setting, obtaining best possible bounds on (R)VaR is an open problem in general: explicit bounds are only available for small portfolios ($n = 2$), some homogeneous portfolios, and for portfolios that are asymptotically large ($n \to +\infty$); see also Puccetti and Rüschendorf (2013a) and Puccetti et al. (2013). Finally, we discuss under which circumstances an inequality constraint on correlation does or does not contain enough dependence information to affect the RVaR bounds.

Let us give the definitions of the risk measures considered in the present chapter. We denote by $\mathrm{VaR}_q(X)$ or by $F_X^{-1}(\cdot)$ the left inverse of the distribution function of a random variable (rv) X, i.e.,

$$\mathrm{VaR}_q(X) = \inf\{x \in \mathbb{R} \mid F_X(x) \geqslant q\}, \qquad q \in (0,1)\,.$$

We denote by $\mathrm{VaR}_q^+(X)$ the right inverse, i.e.,

$$\mathrm{VaR}_q^+(X) = \sup\{x \in \mathbb{R} \mid F_X(x) \leqslant q\}.$$

The Tail-Value-at-Risk (TVaR) at probability level $q \in (0,1)$ is defined as

$$\mathrm{TVaR}_q(X) = \frac{1}{1-q}\int_q^1 \mathrm{VaR}_\gamma(X)\, d\gamma.$$

Finally, the Range-Value-at-Risk (RVaR) of X, introduced in Cont et al. (2010) is formally defined as

$$\mathrm{RVaR}_{q,q'}(X) = \frac{1}{q'-q}\int_q^{q'} \mathrm{VaR}_\gamma(X)d\gamma,\, 0 < q < q' < 1.$$

RVaR is a risk measure that encompasses VaR and TVaR as limiting cases. Specifically,

$$\mathrm{VaR}_q(X) = \lim_{r \nearrow q} \mathrm{RVaR}_{r,q}(X) \quad \mathrm{VaR}_q^+(X) = \lim_{r \searrow q} \mathrm{RVaR}_{q,r}(X) \quad \text{and} \quad \mathrm{TVaR}_q(X) = \lim_{q' \nearrow 1} \mathrm{RVaR}_{q,q'}(X)\,.$$

In this Chapter, we study risk bounds for sums $S = X_1 + ... + X_n$ assuming that in addition to knowledge of the dfs F_i of the risks X_i, $i = 1, ..., n$, also some information on their dependence (copula) is available. In this regard, we will assume that all dfs are continuous, as this allows to exploit the one-to-one relationship between copulas and the joint distribution of $X_1, ..., X_n$. In what follows, we tacitly assume that all quantities we define exist.

3.2 Risk bounds with two risks

In this section, we study VaR and TVaR bounds for a sum $S = X_1 + X_2$, $X_1 \sim F_1$ and $X_2 \sim F_2$ under a dependence constraint. Specifically, we consider the problems

$$\begin{aligned} \overline{\varrho}^d := \sup \quad & \varrho(X_1 + X_2) \\ \text{subject to} \quad & X_j \sim F_j,\ j = 1, 2 \\ & \delta(X_1, X_2) = d, \end{aligned} \tag{3.2.1}$$

and

$$\begin{aligned} \underline{\varrho}^d := \inf \quad & \varrho(X_1 + X_2) \\ \text{subject to} \quad & X_j \sim F_j,\ j = 1, 2 \\ & \delta(X_1, X_2) = d. \end{aligned} \tag{3.2.2}$$

Here, $\varrho(\cdot)$ coincides with either VaR_q^+, VaR_q or with TVaR_q, $0 < q < 1$, and $\delta(X_1, X_2)$ is some measure of association (dependence), i.e., a mapping that assigns a value to (X_1, X_2) with the aim to reflect the strength of dependence between X_1 and X_2. In this regard, Sklar's Theorem states that the joint distribution of (X_1, X_2) is completely specified by its continuous marginal distributions F_1 and F_2 and by the unique copula C such that $(X_1, X_2) =^d (F_1^{-1}(U), F_2^{-1}(V))$, $U, V \sim \mathcal{U}(0, 1)$ and $P(U \leqslant u, V \leqslant v) = C(u, v)$ for all $(u, v) \in [0, 1]^2$. Thus, in what follows we may write $\delta(C, F_1, F_2)$ instead of $\delta(X_1, X_2)$. Note also that problems (3.2.1) and (3.2.2) amount to constrained optimization problems over a set of copulas: Given the marginal distributions F_1 and F_2 of X_1 and X_2, respectively, we aim to determine the best possible bounds for $\varrho(X_1 + X_2)$ by letting vary all copulas C under the constraint that the value of the dependence measure $\delta(C, F_1, F_2)$ is equal to d. If we omit the dependence constraint in problems (3.2.1) or (3.2.2) (but retain the condition that the marginal distribution functions F_1 and F_2 are known), then we label these optimization problems as "unconstrained optimization problems" and denote the unconstrained bounds by $\overline{\varrho}$ and $\underline{\varrho}$, respectively.

As for the measure of dependence δ, we make the following assumption.

Assumption 3.2.1. *For any given dfs F_1 and F_2 and any given copulas C_1 and C_2 it holds that:*

- *Let C_1 be pointwise lower than C_2, that is $C_1(u, v) \leqslant C_2(u, v)$, $\forall (u, v) \in [0, 1]^2$. Then, $\delta(C_1, F_1, F_2) \leqslant \delta(C_2, F_1, F_2)$.*
- *Let $C_\alpha = \alpha C_1 + (1 - \alpha) C_2, \alpha \in [0, 1]$. Then, the mapping $\alpha \in [0, 1] \mapsto \delta(C_\alpha, F_1, F_2)$ is continuous.*

The first condition in Assumption 3.2.1 states that δ is consistent with pointwise (or concordance) copula ordering. This condition is natural and satisfied by most measures of association. In fact, it is typically considered as a desirable axiom that a measure of dependence should satisfy, see also Nelsen (2010) for more details. This condition also implies that for any copula C, $\delta(C^a, F_1, F_2) \leqslant \delta(C, F_1, F_2) \leqslant \delta(C^c, F_1, F_2)$ where C^a and C^c denote respectively the antimonotonic and the comonotonic copula, $C^a(u, v) = \max(u + v - 1, 0)$, $C^c(u, v) = \min(u, v)$,

$(u, v) \in [0,1]^2$. The second condition in Assumption 3.2.1 is of a more technical nature. It is satisfied by the following three popular measures of dependence: Pearson correlation "corr", Spearman's rho "ρ" and Kendall's tau "τ". These are defined as

- Pearson correlation, $\delta(\cdot,\cdot) = \text{corr}(\cdot,\cdot)$: Let X_1, X_2 be two random variables. Then,

$$\text{corr}(X_1, X_2) := \frac{\mathbb{E}((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)))}{\text{std}(X_1)\text{std}(X_2)}. \tag{3.2.3}$$

- Spearman's rho, $\delta(\cdot,\cdot) = \rho(\cdot,\cdot)$: Let X_1, X_2 be two random variables with dfs F_1 and F_2, respectively with a copula C. Then,

$$\rho(X_1, X_2) = \text{corr}(F_1(X_1), F_2(X_2)). \tag{3.2.4}$$

  Furthermore, the Spearman's rho admits the following copula based representation,

$$\rho(X_1, X_2) = 12 \int_0^1 \int_0^1 C(u, v)dudv - 3. \tag{3.2.5}$$

- Kendall's tau, $\delta(\cdot,\cdot) = \tau(\cdot,\cdot)$: Let (X_1, X_2) and (X_1', X_2') be two independent and identically distributed random pairs with a copula C. Then,

$$\tau(X_1, X_2) := P[(X_1 - X_1')(X_2 - X_2') > 0] - P[(X_1 - X_1')(X_2 - X_2') < 0]. \tag{3.2.6}$$

  Kendall's tau thus reflects the probability of concordance between the variables X_1 and X_2 minus their probability of discordance. Furthermore, the Kendall's tau admits the following representation,

$$\tau(X_1, X_2) := 4 \int_0^1 \int_0^1 C(u, v)dC(u, v) - 1. \tag{3.2.7}$$

In particular, $\rho(X_1, X_2)$ and $\tau(X_1, X_2)$ only depend on the copula C of (X_1, X_2) and we may thus also write $\rho(C)$ instead of $\rho(X_1, X_2)$ and likewise $\tau(C)$ instead of $\tau(X_1, X_2)$. Throughout this first section, we establish that when the risk is measured with VaR or TVaR there is a large range of values for d for which the dependence constraint does not affect the maximum possible risk, i.e., the constrained bound $\overline{\varrho}^d$ coincide with the unconstrained bound $\overline{\varrho}$ in most cases of practical interest.

3.2.1 Upper VaR bound

Let $\varrho(\cdot) = \text{VaR}_q^+$, $q \in (0, 1)$. We specifically denote the solution $\overline{\varrho}^d$ to the constrained upper bound Problem (3.2.1) by $\overline{\text{VaR}}_q^d$ whereas the solution $\overline{\varrho}$ for the unconstrained case is denoted as $\overline{\text{VaR}}_q$. Makarov (1981) and Rüschendorf (1982) solved the problem

$$\begin{array}{ll} \sup & P(X_1 + X_2 > t) \\ \text{subject to} & X_j \sim F_j,\ j = 1, 2, \end{array} \tag{3.2.8}$$

and from their analysis one obtains the following formula for $\overline{\text{VaR}}_q$:

$$\overline{\text{VaR}}_q = \inf_{u \in (q,1)} \{F_1^{-1}(u) + F_2^{-1}(1+q-u)\}. \tag{3.2.9}$$

Formula (3.2.9) shows that for the computation of $\overline{\text{VaR}}_q$ one only needs to deal with the restrictions of the quantile functions F_1^{-1} and F_2^{-1} to the upper part $[q, 1]$ of their domain. It is then also clear that for given dfs F_1 and F_2, there exist many distributions for (X_1, X_2) (and thus many copulas) such that the upper bound in (3.2.9) is achieved. In fact, the formula suggests that we only need to construct risks $X_1 \sim F_1$ and $X_2 \sim F_2$ such that when X_1 is big (taking values higher than its VaR at level q) also X_2 is big (taking values higher than its VaR at level q) and, moreover, that the risks are anti-monotonic in this scenario. To make this intuition clear, we define for each $q \in (0, 1)$, the set $\mathscr{C}(q)$ as the set of all copulas that are anti-monotonic on $[q, 1]^2$, i.e.,

$$\mathscr{C}(q) = \left\{\text{copula } C \mid \text{support of } C \text{ on } [q,1]^2 \text{ is } \{(u,v) \mid v = 1+q-u, u \in [q,1]\}\right\}, \tag{3.2.10}$$

and we formulate the following proposition; a formal proof of which can be found in Embrechts et al. (2013).

Proposition 3.2.1. *Let $q \in (0, 1)$ and let $X_1 \sim F_1$ and $X_2 \sim F_2$ having copula $C \in \mathscr{C}(q)$. It holds that*

$$\mathit{VaR}_q^+(X_1 + X_2) = \overline{\mathit{VaR}}_q.$$

The unconstrained upper bound $\overline{\text{VaR}}_q$ is thus attained whenever $X_1 \sim F_1$, $X_2 \sim F_2$ have a copula $C \in \mathscr{C}(q)$. We show hereafter that there exists an interval $[\delta_{min}, \delta_{max}]$ such that when the dependence constraint $d \in [\delta_{min}, \delta_{max}]$ the solution $\overline{\text{VaR}}_q^d$ of the constrained problem in (3.2.1), coincides with $\overline{\text{VaR}}_q$, i.e., the solution to the optimization problem in the absence of a dependence constraint. In this regard, two copulas in $\mathscr{C}(q)$ are of particular interest. We denote them as C_{min}^q and C_{max}^q and provide their expressions in equations (3.2.11) and (3.2.12), respectively.

$$C_{min}^q = \begin{cases} \max(0, u+v-q), \ \forall (u,v) \in [0,q]^2 \\ \min(u,v), \ \forall (u,v) \in [0,q] \times [q,1] \ \mathit{or} \ \in [q,1] \times [0,q] \\ \max(q, u+v-1), \ \forall (u,v) \in [q,1]^2, \end{cases} \tag{3.2.11}$$

and

$$C_{max}^q = \begin{cases} \min(u,v), \ \forall (u,v) \in [0,1]^2 \setminus [q,1]^2 \\ \max(q, u+v-1), \ \forall (u,v) \in [q,1]^2. \end{cases} \tag{3.2.12}$$

The support of C_{min}^q is thus the set containing all pairs $(u, v) \in [0, 1]^2$ such that

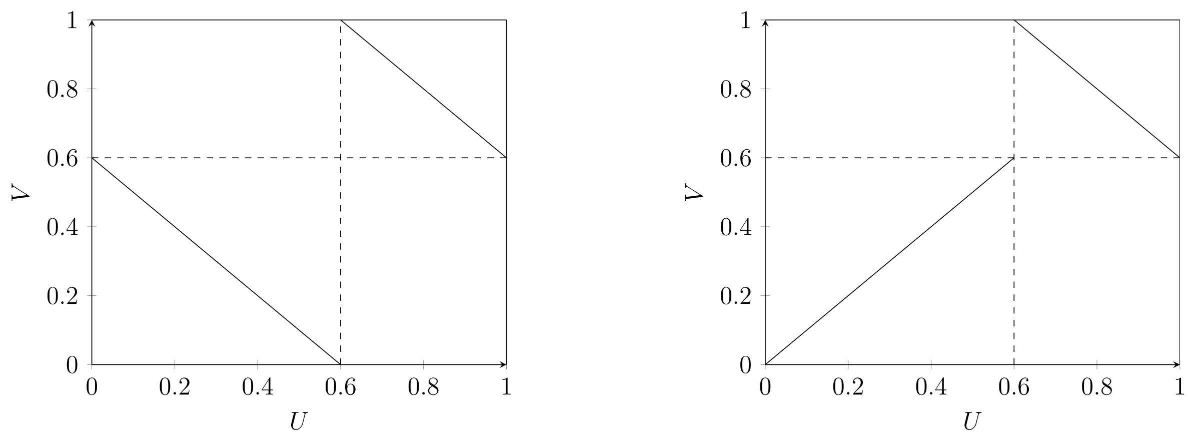
$$\begin{cases} v = q-u, \ \forall (u,v) \in [0,1]^2 \setminus [q,1]^2 \\ v = 1+q-u, \ \forall (u,v) \in [q,1]^2, \end{cases} \tag{3.2.13}$$

and the support of C^q_{max} is the set containing all pairs $(u, v) \in [0,1]^2$ such that

$$\begin{cases} v = u, \ \forall (u,v) \in [0,1]^2 \setminus [q,1]^2 \\ v = 1 + q - u, \ \forall (u,v) \in [q,1]^2. \end{cases} \tag{3.2.14}$$

In Figure 3.1, we display the supports of C^q_{min} and C^q_{max}. These copulas both reflect an anti-monotonic dependence on $[q,1]^2$ whereas on $[0,q]^2$ the dependence is either of a comonotonic nature (case of C^q_{max}) or is anti-monotonic again (case of C^q_{min}). The next lemma will be useful

Figure 3.1: Supports of C^q_{min} (left panel) and C^q_{max} (right panel) for $q = 0.6$.



for our analysis.

Lemma 3.2.2. *Let $q \in (0,1)$ and C be a bivariate copula. Then,*

$$C \in \mathscr{C}(q) \iff C^q_{min}(u,v) \leqslant C(u,v) \leqslant C^q_{max}(u,v), \tag{3.2.15}$$

for all $(u,v) \in [0,1]^2$.

Lemma 3.2.2 states C^q_{min} and C^q_{max} are the pointwise lower and resp. upper copula bounds for all copulas in $\mathscr{C}(q)$. Therefore, if δ is a dependence measure that satisfies assumption 3.2.1 and C is a copula in the set $\mathscr{C}(q)$, then $\delta(C^q_{min}, F_1, F_2) \leqslant \delta(C, F_1, F_2) \leqslant \delta(C^q_{max}, F_1, F_2)$ holds true for any pair of marginal distributions F_1 and F_2.

In the following theorem, we clarify the role of C^q_{min} and C^q_{max} in the computation of $\overline{\text{VaR}}^d_q$.

Theorem 3.2.3 (Constrained bounds that coincide with unconstrained ones)**.** *Let $q \in (0,1)$, $X_1 \sim F_1$ and $X_2 \sim F_2$, let δ be a measure of dependence that satisfies assumption 3.2.1. Let $\delta_{min} = \delta(C^q_{min}, F_1, F_2)$ and $\delta_{max} = \delta(C^q_{max}, F_1, F_2)$. Then, for every $d \in [\delta_{min}, \delta_{max}]$ it holds that*

$$\overline{VaR}^d_q = \overline{VaR}_q. \tag{3.2.16}$$

Moreover, $\overline{VaR}^d_q$ is attainable i.e., for every $d \in [\delta_{min}, \delta_{max}]$, there exists a copula C such that

$\delta(C, F_1, F_2) = d$, *and for* $X_1 \sim F_1$, $X_2 \sim F_2$ *having copula* C *it holds that*

$$VaR_q^+(X_1 + X_2) = \overline{VaR}_q^d. \tag{3.2.17}$$

The proof of Theorem 3.2.3 is given in Appendix 3.5.2. From Theorem 3.2.3, we obtain that the interval $[\delta_{min}, \delta_{max}]$ describes a range of values for the dependence constraint δ such that the dependence information coming from the constraint does not make it possible to reduce the unconstrained upper bound $\overline{\text{VaR}}_q$. In this regard, note that the values of δ_{min} and δ_{max} will depend on $q \in (0, 1)$: when q increases, the upper interval $[q, 1]$ shrinks implying that C^q_{min} and C^q_{max} tend more and more to the anti-monotonic resp. comonotonic copula C^a and C^c. As the dependence measures we consider are consistent with pointwise ordering on copulas (Assumption 3.2.1), the interval $[\delta_{min}, \delta_{max}]$ will become wider.

One could wonder what happens if one formulates problem (3.2.1) using VaR as the risk measure at hand instead of VaR^+. Under some mild additional assumptions, we show in the next proposition that also in this case the worst-case (unconstrained) VaR cannot be readily improved by the knowledge of the dependence measure.

Proposition 3.2.4. *Let* $q \in (0, 1)$, $X_1 \sim F_1$ *and* $X_2 \sim F_2$, *let* δ *be a measure of dependence that satisfies assumption 3.2.1. Assume that* F_1^{-1} *and* F_2^{-1} *are continuous on* $(0, 1)$ *and that* δ *is such that the mappings* $\delta_{min} : q \in (0, 1) \mapsto \delta(C^q_{min}, F_1, F_2)$ *and* $\delta_{max} : q \in (0, 1) \mapsto \delta(C^q_{max}, F_1, F_2)$ *are continuous. Then, for every* $d \in (\delta_{min}, \delta_{max})$ *it holds that* $\overline{VaR}_q$ *is the solution to*

$$\begin{array}{ll} \textit{sup} & VaR_q(X_1 + X_2) \\ \textit{subject to} & X_j \sim F_j,\ j = 1, 2 \\ & \delta(X_1, X_2) = d \end{array} \tag{3.2.18}$$

Under the assumptions of Proposition 3.2.4, the upper bound on VaR_q and VaR_q^+ for a sum of two risks X_1 and X_2 with given marginal dfs F_1 and F_2 and a given value for their dependence $\delta(X_1, X_2)$ coincides with the unconstrained upper bound, with the subtle difference that unlike in the case of VaR_q^+, the upper bound on VaR_q is not attainable.

Constraint on Spearman's Rho or Kendall's Tau.

In this section, we apply Theorem 3.2.3 to the case in which the measure of dependence δ considered is specifically given as either Spearman's rho (i.e., $\delta(\cdot, \cdot) = \rho(\cdot, \cdot)$) or Kendall's tau (i.e., $\delta(\cdot, \cdot) = \tau(\cdot, \cdot)$). We first formulate the following lemma.

Lemma 3.2.5 (Expressions for (ρ_{min}, ρ_{max}) and (τ_{min}, τ_{max}))**.** *Let* $q \in (0, 1)$. *It holds that*

$$\rho_{min} = \rho(C^q_{min}) = -6q\,(q - 1) - 1, \quad \rho_{min} = \rho(C^q_{max}) = 1 - 2(1 - q)^3, \tag{3.2.19}$$

and

$$\tau_{min} = \tau(C^q_{min}) = -4q\,(q - 1) - 1, \quad \tau_{max} = \tau(C^q_{max}) = -2\,(q - 1)^2 + 1. \tag{3.2.20}$$

The proof of Lemma 3.2.5 is relegated to Appendix 3.5.4. By combining Theorem 3.2.3 and Lemma 3.2.5 we obtain the following result.

Theorem 3.2.6 (Constrained bounds that coincide with unconstrained ones - case of $\delta(\cdot,\cdot)=\rho(\cdot,\cdot)$ or $\delta(\cdot,\cdot)=\tau(\cdot,\cdot)$)**.** *Let $q\in(0,1)$, $X_1\sim F_1$ and $X_2\sim F_2$.*

If $\delta(\cdot,\cdot)=\rho(\cdot,\cdot)$ and $-6q\,(q-1)-1\leqslant d\leqslant 1-2(1-q)^3$, then

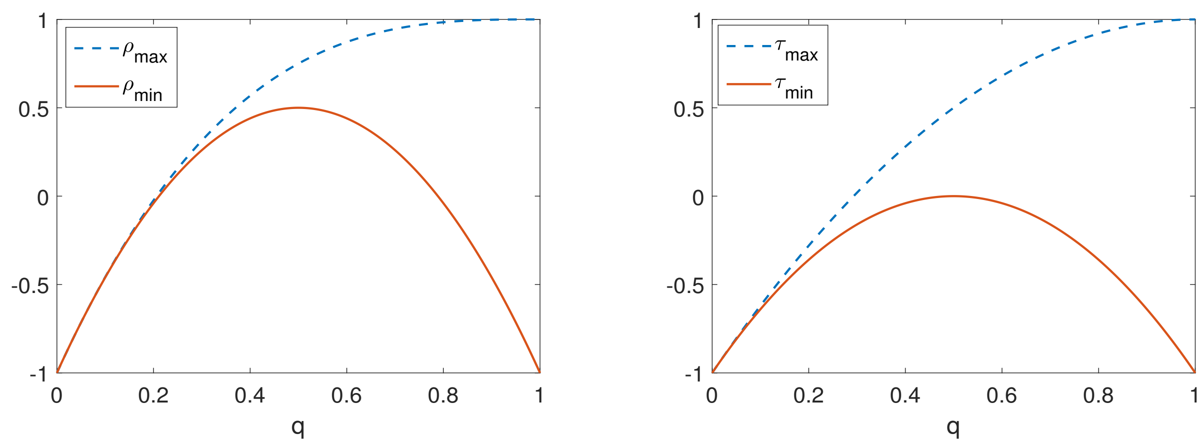
$$\overline{VaR}_q^d=\overline{VaR}_q. \tag{3.2.21}$$

If $\delta(\cdot,\cdot)=\tau(\cdot,\cdot)$, and $-4q\,(q-1)-1\leqslant d\leqslant -2\,(q-1)^2+1$, then

$$\overline{VaR}_q^d=\overline{VaR}_q, \tag{3.2.22}$$

and the bounds are attainable.

Figure 3.2: We display as a function of q the ranges $[\rho_{min},\rho_{max}]$ and $[\tau_{min},\tau_{max}]$ given in Lemma 3.2.5. When the dependence measure ρ resp. τ takes value in $[\rho_{min},\rho_{max}]$ resp. $[\tau_{min},\tau_{max}]$, the constrained upper bound $\overline{\text{VaR}}_q^d$ does not improve on the unconstrained upper bound $\overline{\text{VaR}}_q$.



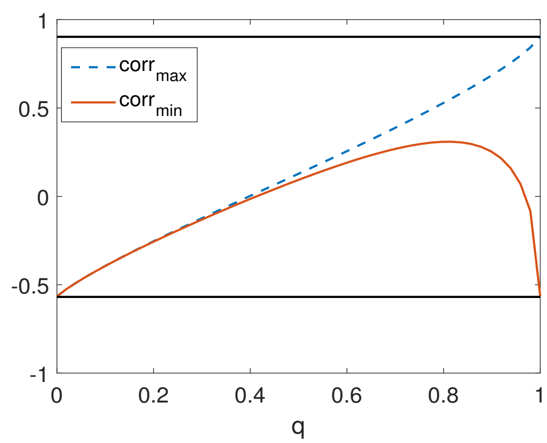
Panel A: $[\rho_{min},\rho_{max}]$ w.r.t. q Panel B: $[\tau_{min},\tau_{max}]$ w.r.t. q

Figure 3.2 displays the intervals $[\rho_{min},\rho_{max}]$ and $[\tau_{min},\tau_{max}]$ for which the constrained upper bound $\overline{\text{VaR}}_q^d$ ($d\in[\rho_{min},\rho_{max}]$ resp. $d\in[\tau_{min},\tau_{max}]$) does not improve the unconstrained upper bound $\overline{\text{VaR}}_q$. We observe that for the levels of q usually considered in risk management ($q\geqslant 0.95$), $[\rho_{min},\rho_{max}]$ and $[\tau_{min},\tau_{max}]$ almost cover the interval $[-1,1]$. For instance, when $q=0.995$, we obtain that $[\rho_{min},\rho_{max}]$=$[-0.970,1.000]$ and $[\tau_{min},\tau_{max}]$=$[-0.980,1.000]$, and when $q=0.95$, we obtain that $[\rho_{min},\rho_{max}]$=$[-0.715,1.000]$ and $[\tau_{min},\tau_{max}]$=$[-0.810,0.995]$. Thus, when $q\geqslant 0.95$, it is, practically speaking, rather unlikely that dependence information (coming from Spearman's rho or Kendall's tau) is such that the unconstrained VaR upper bound can be reduced.

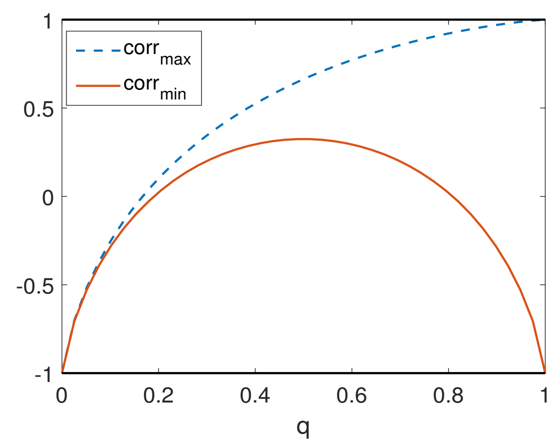
Constraint on Pearson Correlation

Theorem 3.2.6 can be extended to include the case of the Pearson correlation as source of dependence information. An issue, however, is that, as compared to Spearman's rho and Kendall's tau, the Pearson correlation also depends on the marginal distributions F_1 and F_2 of the risks X_1 and X_2, implying that an explicit formula for $\text{corr}_{min} = \text{corr}(C^q_{min}, F_1, F_2)$ and $\text{corr}_{max} = \text{corr}(C^q_{max}, F_1, F_2)$ is not always available. Nevertheless, corr_{min} and corr_{max} can always be obtained via simulations. We illustrate this feature in Figure 3.3. In Panel A we consider $F_1 \sim Gamma(2, 3)$ and $F_2 \sim Lognormal(2, 1)$, whereas in Panel B we assume $F_1 \sim Normal(2, 3)$ and $F_2 \sim Normal(2, 4)$. Note that as compared to the case of Spearman's rho and Kendall's tau, the absolute bounds on the Pearson correlation are in general no longer equal to -1 resp. $+1$, but depend on the marginal distributions F_1 and F_2 at hand: see Panel A in particular, where the absolute bounds are equal to -0.565 resp. 0.905, that are the bounds obtained by computing the correlation between X_1 and X_2 when the copula C is the comonotonic copula C^c (for the maximum correlation) and the anti-monotonic copula C^a (for the minimum correlation). We obtain the same conclusion as in the previous section: when $q \geqslant 0.95$ and certainly when $q \geqslant 0.995$, the knowledge of the Pearson correlation among two risks typically does not make it possible to reduce the unconstrained VaR upper bound $\overline{\text{VaR}}_q$.

Figure 3.3: We display as a function of q the range $[\text{corr}_{min}, \text{corr}_{max}]$. In Panel A we consider $F_1 \sim Gamma(2, 3), F_2 \sim Lognormal(2, 1)$ and in Panel B we assume $F_1 \sim Normal(2, 3), F_2 \sim Normal(2, 4)$. The solid black lines depict the absolute bounds on the Pearson correlation for the marginal distributions at hand. When the Pearson correlation $\text{corr}(\cdot, \cdot)$ takes value in $[\text{corr}_{min}, \text{corr}_{max}]$, the constrained upper bound $\overline{\text{VaR}}^d_q$ does not improve on the unconstrained upper bound $\overline{\text{VaR}}_q$.



Panel A: $[\text{corr}_{min}, \text{corr}_{max}]$ w.r.t. q



Panel B: $[\text{corr}_{min}, \text{corr}_{max}]$ w.r.t. q

Comparison With Kaas et al. (2009)

Our main problem (3.2.1) was (among other results) also studied in Kaas et al. (2009). However, their approach is very different from ours from a methodological point of view. In this section we aim to point out these methodological differences and the additional insights that we offer with respect to their analysis.

We first start with a brief description of their methodology in the case where Spearman's rho is used as measure of dependence. Their problem then reads as

$$\begin{aligned} &\sup && \mathrm{VaR}_q(X_1+X_2) \\ &\text{subject to} && X_j \sim F_j, \\ & && \rho(C)=d, \end{aligned} \tag{3.2.23}$$

where $d \in (0,1)$ and we optimize over the set of copulas having a Spearman's rho equal to d.

To deal with this problem, the general idea in Kaas et al. (2009) is that in a first step the information coming from a measure of association may allow to improve the Fréchet-Hoeffding bounds C^a and C^c on copulas (see Nelsen et al. (2004) and Nelsen and Úbeda Flores (2005) for details). In a second step they use the improved copula bounds to improve the unconstrained VaR bounds.

Specifically, in a first step they derive for each $d \in (0,1)$ a copula $\underline{C}_{\rho,d}$ that is pointwise lower than any copula having a Spearman's rho equal to d, i.e. solving

$$\underline{C}_{\rho,d} = \inf\{C \mid C \text{ is a copula}, \rho(C)=d\}.$$

For the analytical expression of $\underline{C}_{\rho,d}$, see equation (7) in Kaas et al. (2009). However, observe that the Spearman's rho of $\underline{C}_{\rho,d}$ is not equal to d in general. Second, they apply Theorem 3.1 in Embrechts and Puccetti (2006) and Theorem 5 in Embrechts et al. (2005) to obtain the solution to the following problem

$$\begin{aligned} &\sup && \mathrm{VaR}_q(X_1+X_2) \\ &\text{subject to} && X_j \sim F_j, \\ & && C \geqslant C_L. \end{aligned} \tag{3.2.24}$$

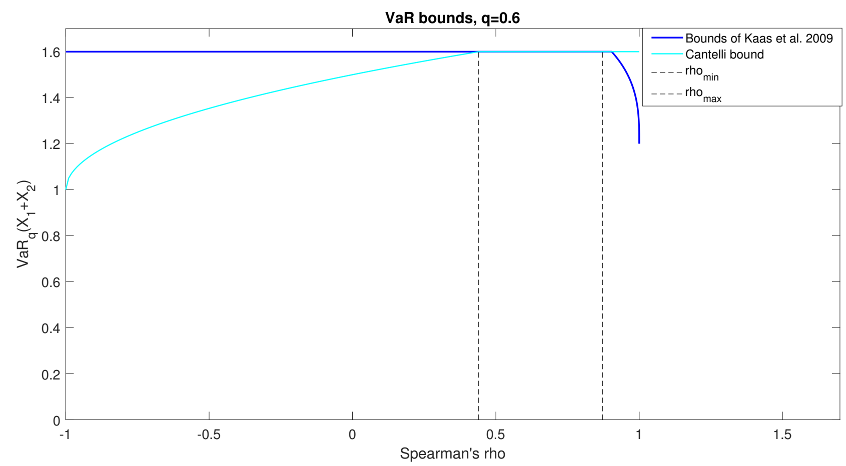
in which they take $C_L = \underline{C}_{\rho,d}$. Finally, they propose the solution of this problem as a bound for VaR of a sum of two risks with given marginal distributions and when their Spearman's correlation is given.

Observe now that if a copula satisfies the constraint of problem (3.2.23), by construction it satisfies also the constraint in problem (3.2.24) with $C_L = \underline{C}_{\rho,d}$, but not vice versa (take for instance the comonotonic copula, it satisfies the condition in (3.2.24) but not the one in (3.2.23)). This of course implies that the supremum in (3.2.24) is in general higher than the supremum in (3.2.23). Therefore, the constrained VaR upper bound proposed in Kaas et al. (2009) may not be a solution for problem (3.2.23). This point is important in that Kaas et al. (2009) did not explicitly discuss whether their VaR bounds are best-possible or not. Let us now illustrate that their bounds are indeed not best-possible in general. Let $\widetilde{\mathrm{VaR}}_q^d$ denote the solution of (3.2.23) and consider the case where $X_i \sim \mathcal{U}[0,1]$, for $i=1,2$. Note that for marginal distributions that are standard uniforms, Spearman's rho and the Pearson correlation coincide. In particular, if $X_i \sim \mathcal{U}[0,1]$, for $i=1,2$ a constraint $\rho(C)=d$ on Spearman's rho is equivalent to a constraint on the variance of X_1+X_2 in that $\mathrm{var}(X_1+X_2)=\frac{1+\rho(C)}{6}$. Thus, when $\rho(C)=d$, $\widetilde{\mathrm{VaR}}_q^d$ has to satisfy the following inequality

$$\widetilde{\text{VaR}}_q^d \leqslant \min\left\{ 1 + \sqrt{\frac{1+d}{6}}\sqrt{\frac{q}{1-q}}\ ,\ \overline{\text{VaR}}_q \right\}. \tag{3.2.25}$$

The bound in (3.2.25) is built using the best-possible unconstrained upper bound $\overline{\text{VaR}}_q$ and the so-called Cantelli moment bound, that provides the bound on VaR for the sum $X_1 + X_2$ when their mean and variance is known; see also Barrieu and Scandolo (2015) and Bernard et al. (2017) for more details. Figure 3.4 then shows that whenever $d < \rho_{min}$ the bound proposed by Kaas et al. (2009) is actually bigger than the bound in (3.2.25), i..e., bigger than $\widetilde{\text{VaR}}_q^d$ and thus is not best-possible in general. As we have established that $\overline{\text{VaR}}_q$ is the best possible constrained bound when $d \in [\rho_{min}, \rho_{max}]$ and as the optimal solution of their problem (3.2.24) is necessarily smaller than $\overline{\text{VaR}}_q$, their bound thus matches $\overline{\text{VaR}}_q$ whenever $d \in [\rho_{min}, \rho_{max}]$.

Figure 3.4: Upper bounds on $\text{VaR}_q(S)$ when $S = X_1 + X_2$ with $X_i \sim \mathcal{U}[0,1]$, $q = 0.6$ and $\rho(C) = d$, $0 < d < 1$. The dark blue line represents the bound proposed by Kaas et al. (2009) as a function of d The light-blue line represents the upper bound defined in (3.2.25), also expressed as a function of d. The dashed vertical lines represent ρ_{min} and ρ_{max}.



As compared to Kaas et al. (2009), our analysis provides some additional insights. First, under a mild assumption on the dependence measure $\delta(\cdot,\cdot)$ we show how to obtain the interval $(\delta_{min}, \delta_{max})$ having the property that if the dependence measure $\delta(\cdot,\cdot)$ takes value in this interval, the constrained VaR bound coincides with the unconstrained VaR upper bound. Moreover, while this interval can always be determined using simulation, we establish this interval explicitly for the case of Spearman's ρ and Kendall's τ. By contrast, the results provided in Kaas et al. (2009) do not allow to determine such interval. We believe that the definition and the study of the interval $(\delta_{min}, \delta_{max})$ provides a substantial contribution to the debate regarding the impact of a measure of dependence constraint on VaR bounds, in that we show that for the probability levels usually considered in practice and required by solvency regulations such as Solvency II and Basel III, i.e., $q \geqslant 0.95$, the interval $(\delta_{min}, \delta_{max})$ covers the range of values that the measure of dependence δ may assume in practice, leaving outside only values of δ that are typically not observed in real-world data. This feature is also illustrated in Figure 2, Figure 3 and Table 1. This result can be then used to effectively argue that the knowledge of a dependence measure does not guarantee a reduction for the worst-case VaR scenario. Second, our analysis also deals with the case of the Pearson correlation as source of dependence information, which is by far the most popular measure of dependence in the financial and insurance industry and which was not considered in Kaas et al. (2009). In addition, we show that if one reformulates (3.2.23) in terms of VaR^+ instead of VaR, the problem becomes more tractable in that it is possible to obtain bounds that are not only best-possible but also attainable (see Theorem 3.2.3).

Remark 3.2.1. *Kaas et al. (2009) left an open question regarding VaR bounds for a sum of two risks when both Kendall's tau and Spearman's rho are given. Solving an optimization problem with more than one constraint, i.e., involving various measures of association, can be a challenging task. Nonetheless, we show that Theorem 3.2.3 and Lemma 3.2.5 make it possible to solve the problem*

$$\begin{array}{ll} \textit{sup} & \textit{VaR}_q^+(X_1+X_2) \\ \textit{subject to} & \begin{cases} X_j \sim F_j, j=1,2 \\ \rho(X_1,X_2) \leqslant \rho* \\ \tau(X_1,X_2) \leqslant \tau^* \\ corr(X_1,X_2) \leqslant c^*. \end{cases} \end{array} \tag{3.2.26}$$

In Problem (3.2.26)*, we thus look for the maximum* $\textit{VaR}_q^+(S)$ *given that the random vector* (X_1, X_2) *simultaneously satisfies three constraints, one for each measure of association considered in our analysis. Of course, having three constraints is more restrictive than having only one of them and one could expect that this would automatically translate in a reduction of the maximum* $\textit{VaR}_q^+(S)$*. Nevertheless, it is clear that if* $\rho^* \geqslant \rho_{min}$*,* $\tau^* \geqslant \tau_{min}$ *and* $c^* \geqslant corr_{min}$*, then the constraints will be satisfied for the random pair* (X_1, X_2) *with copula* C^q_{min}*,* $X_1 \sim F_1$ *and* $X_2 \sim F_2$*. As* $\textit{VaR}_q(X_1+X_2) = \overline{\textit{VaR}}_q$ *we can thus conclude that the solution to Problem* (3.2.26) *is again* $\overline{\textit{VaR}}_q$*.*

We consider the case in which the risks have a Gamma resp. Lognormal distribution, i.e., $X_1 \sim Gamma(2,3)$ *and* $X_2 \sim Lognormal(2,1)$*. Table 3.1 shows the values of* δ_{min} *(Panel A) and* δ_{max} *(Panel B) for various probability levels and for the three measures of dependence we studied (Pearson correlation, Kendall's tau, Spearman's rho). We underline that for* $q = 99.5\%$*, which is the probability level required by Solvency II for capital calculations, the maximum VaR is impacted by the information on the dependence measure (and decreases) only if* $\rho^* \in [-1, -0.970[$*,* $\tau^* \in [-1, -0.980[$ *or* $c^* \in [-0.565, -0.321[$*.*

What if $d \notin [\delta_{min}, \delta_{max}]$?

So far, our analysis has focused on the definition and the study of the interval $[\delta_{min}, \delta_{max}]$. We showed that, for the probability levels usually considered in risk management, i.e., $q \geqslant 0.95$, this interval allows us to obtain the solution of problem (3.2.1) for most reasonable values that the dependence measure δ can take (see Table 3.1, Figures 3.2 and 3.3). Nonetheless, a complete solution of problem (3.2.1) indeed requires to study the VaR^+ upper bound also for $d \notin [\delta_{min}, \delta_{max}]$. This problem seems mathematically very challenging, as illustrated in the following example. Fix $q \in (0,1)$, and consider problem (3.2.1) for $X_1, X_2 \sim \mathcal{U}[0,1]$, $\delta(\cdot,\cdot) = \rho(\cdot,\cdot)$, and $d = \frac{3}{2}q(1-q)+1$.

Table 3.1: Values of δ_{min} (Panel A) and of δ_{max} (Panel B).

q	ρ_{min}	τ_{min}	corr_{min}
60.0%	0.440	-0.040	0.191
75.0%	0.125	-0.250	0.296
95.0%	-0.715	-0.81	0.117
97.5%	-0.854	-0.90	-0.036
99.0%	-0.941	-0.960	-0.216
99.5%	-0.970	-0.980	-0.321

Panel A

q	ρ_{max}	τ_{max}	corr_{max}
60.0%	0.872	0.680	0.256
75.0%	0.969	0.875	0.458
95.0%	0.999	0.995	0.779
97.5%	0.999	0.998	0.832
99.0%	0.999	0.999	0.871
99.5%	0.999	0.999	0.886

Panel B

Note that $\frac{3}{2}q(1-q)+1 < \rho_{min}$ holds for any $q \in (0,1)$. Let now us denote with C_q the copula having the following support:

$$\begin{cases} v = l_q - u, \ \forall u \in \left[0, \frac{q}{2}\right] \cup \left[\frac{1}{2}, \frac{1+q}{2}\right] \\ v = u_q - u, \ \forall u \in \left[\frac{q}{2}, \frac{1}{2}\right] \cup \left[\frac{1+q}{2}, 1\right] \end{cases} \tag{3.2.27}$$

with $l_q = \frac{1}{2} + \frac{q}{2}$, $u_q = 1 + \frac{q}{2}$. We start by showing that the equations in (3.2.27) describe the support of a copula. In order to do so, we only need to show that the four segments described in (3.2.27) do not overlap. This can be readily done by observing that when we fix $q \in (0,1)$, the following conditions hold:

1. $l_q \leqslant 1 \leqslant u_q$,
2. $u_q - \frac{q}{2} = 1$,
3. $l_q - \frac{q}{2} = u_q - \frac{1+q}{2}$,
4. $l_q - \frac{1}{2} = u_q - 1$,
5. $l_q - \frac{1+q}{2} = 0$.

We conclude that (3.2.27) describes the support of a shuffle of Min copula, as illustrated in Figure 3.5. If $(X_1, X_2) \sim C_q$, we obtain that $X_1 + X_2 =^d S_q$ with

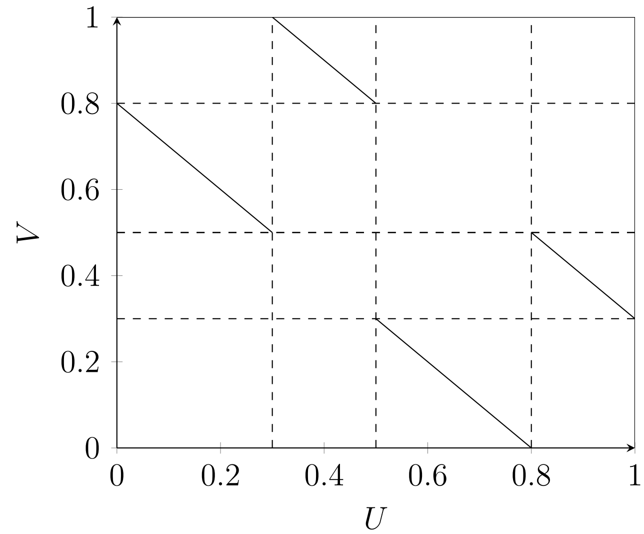
$$S_q = \begin{cases} l_q \text{ with probability } q, \\ u_q \text{ with probability } 1-q. \end{cases} \tag{3.2.28}$$

Thus, $S = X_1 + X_2$ has a two-point distribution, taking value equal to l_q with probability q and to u_q with probability $1-q$. After simple calculations we obtain that if $(X_1, X_2) \sim C_q$, than

$$\text{VaR}_q^+(S) = u_q = \widetilde{\text{VaR}}_q^d < \overline{\text{VaR}}_q^d, \tag{3.2.29}$$

where $\widetilde{\text{VaR}}_q^d$ is the upper bound given in (3.2.25), i.e., matching the Cantelli bound in this case. Moreover, we observe that $\rho(C_q) = \frac{3}{2}q(1-q)+1 = d$ and we have thus shown that the upper bound given in (3.2.25) can be attainable and thus best possible for some specific combinations of $q \in (0,1)$ and $d < \delta_{min}$. That is, we obtain for a given $q \in (0,1)$ a best-possible bound for problem (3.2.1) for the choice $d = \frac{3}{2}q(1-q)+1 \notin [\delta_{min}, \delta_{max}]$. Unfortunately, it seems that a shuffle of Min of the form C_q does not make it possible to solve problem (3.2.1) when $d \neq \frac{3}{2}q(1-q)+1$. Indeed, when $d \neq \frac{3}{2}q(1-q)+1$, a numerical analysis conducted using the RA algorithm (Puccetti and Rüschendorf (2012)) suggests that the bound in (3.2.25) is no longer best-possible, and thus that shuffles of Min in the form of C_q do not always lead to the best possible bound. This example highlights that already in the simple case of uniform marginal distributions, if $d \notin [\delta_{min}, \delta_{max}]$, the copula structure that solves problem (3.2.1) changes according to the specific combination of q and d considered, indicating that the problem appears to be highly challenging.

Figure 3.5: The support of C_q, described in (3.2.27), for $q = 0.6$. The support consists in the four black segments.



3.2.2 Lower VaR bound

In Section 3.2.1 we have shown that for probability levels considered in risk management (i.e., $q = 0.95$ or higher), the availablity of correlation information typically does not make it possible to reduce the unconstrained VaR bounds. Here we study whether this form of dependence information impacts the VaR lower bound. We will show that dependence information typically makes it possible to improve either the unconstrained lower bound (when the probability level q is high) or the unconstrained upper bound (when the probability level q is low). Hence, correlation information can reduce the dependence uncertainty spread (i.e., the difference between the best- and the worst-case value of a risk measure, computed over all dependence structures compatible with the available information). However, the improvement typically comes from the improvement of the unconstrained lower bound (as in practice the probability level q is high).

Let $\varrho(\cdot) = \mathrm{VaR}_q$ for $q \in (0,1)$ and denote the solution $\underline{\varrho}^d$ to the constrained lower bound Problem (3.2.2) by $\underline{\mathrm{VaR}}_q^d$. If we omit the dependence constraint we denote the solution as $\underline{\mathrm{VaR}}_q$, i.e.,

$$\begin{aligned} \underline{\mathrm{VaR}}_q = \inf \quad & \mathrm{VaR}_q(X_1 + X_2) \\ \text{subject to} \quad & X_j \sim F_j,\ j = 1, 2. \end{aligned} \tag{3.2.30}$$

From the analysis in Makarov (1981) and Rüschendorf (1982) it follows that

$$\underline{\mathrm{VaR}}_q = \sup_{u \in [0,q]} \{F_1^{-1}(u) + F_2^{-1}(q-u)\} \tag{3.2.31}$$

The next lemma relates the constrained lower bound Problem (3.2.2) with the constrained upper bound Problem (3.2.1). To establish this connection, it is useful to represent risks X_1 and X_2 as $X_1 = F_{X_1}^{-1}(U)$ and $X_2 = F_{X_2}^{-1}(V)$, respectively, where $(U, V) \sim C$ and C is a copula.

Lemma 3.2.7. *Let $\mathscr{C}$ be any arbitrary set of copulas. Then*

$$\min_{(U,V)\sim C \in \mathscr{C}} \mathit{VaR}_q \left(F_{X_1}^{-1}(U) + F_{X_2}^{-1}(V)\right) = \max_{(U,V)\sim C \in \mathscr{C}} \mathit{VaR}^+_{1-q} \left(F_{-X_1}^{-1}(1-U) + F_{-X_2}^{-1}(1-V)\right). \tag{3.2.32}$$

Proof. From lemma 1 in Dhaene et al. (2006),

$$\mathrm{VaR}_q \left(F_{X_1}^{-1}(U) + F_{X_2}^{-1}(V)\right) = -\mathrm{VaR}^+_{1-q} \left(-F_{X_1}^{-1}(U) - F_{X_2}^{-1}(V)\right) = -\mathrm{VaR}^+_{1-q} \left(F_{-X_1}^{-1}(1-U) + F_{-X_2}^{-1}(1-V)\right),$$

from which the stated assertion follows. □

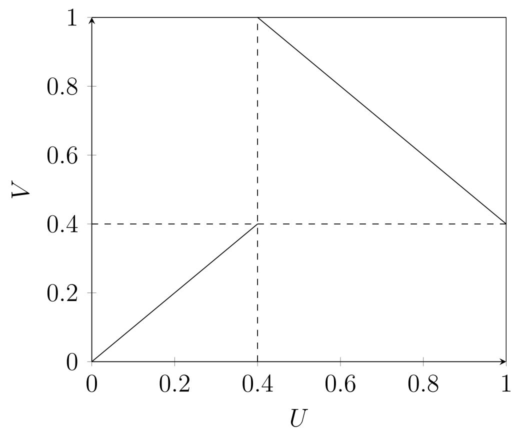
In what follows, if $(U, V) \sim C$ then we denote by C^* the joint distribution function of $(1 - U, 1 - V)$ and we say that C^* is the associated copula of C. Specifically, $C^*(u, v) = u + v - 1 + C(1 - u, 1 - v), (u, v) \in [0, 1]^2$. In this regard, note that when (X_1, X_2) has copula C then $(-X_1, -X_2)$ has copula C^*, see Lemma 2.3.1 in May and Scherer (2014) for a proof. The following properties are easy to verify. If the copulas C_1, C_2 provide pointwise bounds for C, then the same holds for the associated copulas, i.e., $C_1 \leqslant C \leqslant C_2$ implies $C_1^* \leqslant C^* \leqslant C_2^*$. Furthermore $(C^*)^* = C$. Finally, note that the support of $(1-U, 1-V) \sim C^*$ follows in a straightforward way from the joint distribution C of (U, V).

From Lemma 3.2.7 and its proof, $\mathrm{VaR}_q(X_1+X_2), X_1 \sim F_1, X_2 \sim F_2$ takes minimum value for (X_1, X_2) having copula C if and only if $\mathrm{VaR}^+_{1-q}(-X_1-X_2)$ takes maximum value for $(-X_1, -X_2)$ having copula C^*, which holds whenever $C^* \in \mathscr{C}(1-q)$ (Proposition 3.2.1). Hence, from Proposition 3.2.1 and Lemma 3.2.2 it follows that as long as (X_1, X_2) is such that their associated copula C^* satisfies $C_{min}^{1-q} \leqslant C^* \leqslant C_{max}^{1-q}$ we find that $\mathrm{VaR}_q(X_1 + X_2)$ is minimum. Let us now observe that

$$C_{min}^{1-q} \leqslant C^* \leqslant C_{max}^{1-q} \iff (C_{min}^{1-q})^* \leqslant (C^*)^* \leqslant (C_{max}^{1-q})^* \iff C_{min}^{q} \leqslant C \leqslant C_{max}^{q,L},$$

where in the last step we used that the associated copula of C_{min}^{1-q} is simply given by C_{min}^{q} and where we denoted the associated copula for C_{max}^{1-q} by $C_{max}^{q,L}$. In Figure 3.6 we illustrate how the support of $C_{max}^{q,L}$ (right panel) follows in a straightforward way from the support of C_{max}^{1-q}.

Figure 3.6: Support of C^{1-q}_{max} (left panel) and its associated copula $C^{q,L}_{max}$ (right panel) for $q = 0.6$.



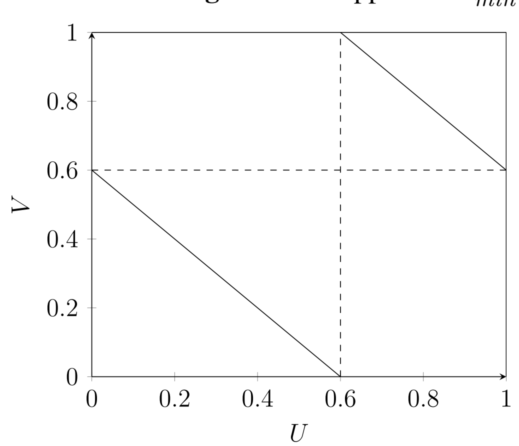
The support of $C^{q,L}_{max}$ is specifically given as the set $(u, v) \in [0, 1]^2$ such that

$$\begin{cases} v = q - u, & \text{for } u \in [0, q] \\ v = u, & \text{for } u \in [q, 1]. \end{cases} \tag{3.2.33}$$

and we readily obtain the following explicit expression for this copula:

$$C^{q,L}_{max}(u, v) = \begin{cases} \max(0, u + v - q), \ \forall (u, v) \in [0, q]^2 \\ \min(u, v), \ \forall (u, v) \in [0, 1]^2 \setminus [0, q]^2 \end{cases} \tag{3.2.34}$$

Figure 3.7: Support of C^q_{min} (left panel) and $C^{q,L}_{max}$ (right panel) for $q = 0.6$.



The copulas C^q_{min} and $C^{q,L}_{max}$ exhibit anti-monotonicity on $[0, q]$ and are extreme in that using a similar argument as in proof of Lemma 3.2.2 one can show all other copulas C that exhibit anti-monotonicity on $[0, q]$ satisfy $C^q_{min}(u, v) \leqslant C(u, v) \leqslant C^{q,L}_{max}(u, v), (u, v) \in [0, 1]^2$. We denote these copulas by $\mathscr{C}_L(q)$, i.e.,

$$\mathscr{C}_L(q) = \left\{\text{copula } C \mid \text{support of } C \text{ on } [0,q]^2 \text{ is } \{(u,v) \mid v = q-u, u \in [0,q]\}\right\}. \qquad (3.2.35)$$

The following proposition is now proven.

Proposition 3.2.8. *Let $q \in (0,1)$ and let $X_1 \sim F_1$ and $X_2 \sim F_2$ have a copula $C \in \mathscr{C}_L(q)$. Then, it holds that*

$$VaR_q(X_1+X_2) = \underline{VaR}_q.$$

We can now also formulate the following theorem.

Theorem 3.2.9. *Let $q \in (0,1)$, $X_1 \sim F_1$ and $X_2 \sim F_2$, let δ be a measure of dependence that satisfies assumption 3.2.1. Let also $\delta_{min} = \delta(C^q_{min}, F_1, F_2)$ and $\delta^L_{max} = \delta(C^{q,L}_{max}, F_1, F_2)$. For every $d \in [\delta_{min}, \delta^L_{max}]$ it holds that*

$$\underline{VaR}^d_q = \underline{VaR}_q. \qquad (3.2.36)$$

Moreover, $\underline{VaR}^d_q$ is attainable, i.e., for every $d \in [\delta_{min}, \delta^L_{max}]$, there exists a copula C such that $\delta(C, F_1, F_2) = d$, and for $X_1 \sim F_1$, $X_2 \sim F_2$ having copula C it holds that

$$VaR_q(X_1+X_2) = \underline{VaR}_q. \qquad (3.2.37)$$

The proof of Theorem 3.2.9 is similar to the proof of Theorem 3.2.3 and thus omitted. Using Theorem 3.2.9 one can thus always obtain an interval for $[\delta_{min}, \delta^L_{max}]$ such that for any $d \in [\delta_{min}, \delta^L_{max}]$, it holds that $\underline{\text{VaR}}^d_q = \underline{\text{VaR}}_q$. In the next lemma, we provide these boundary values for δ_{min} and δ^L_{max} in explicit form for the case of Spearman's rho and Kendall's tau.

Lemma 3.2.10 (Expressions for $(\rho_{min}, \rho^L_{max})$ and $(\tau_{min}, \tau^L_{max})$)**.** *Let $q \in (0,1)$. It holds that*

$$\rho_{min} = \rho(C^q_{min}) = -6q\,(q-1) - 1, \quad \rho^L_{max} = \rho(C^{q,L}_{max}) = 1 - 2q^3. \qquad (3.2.38)$$

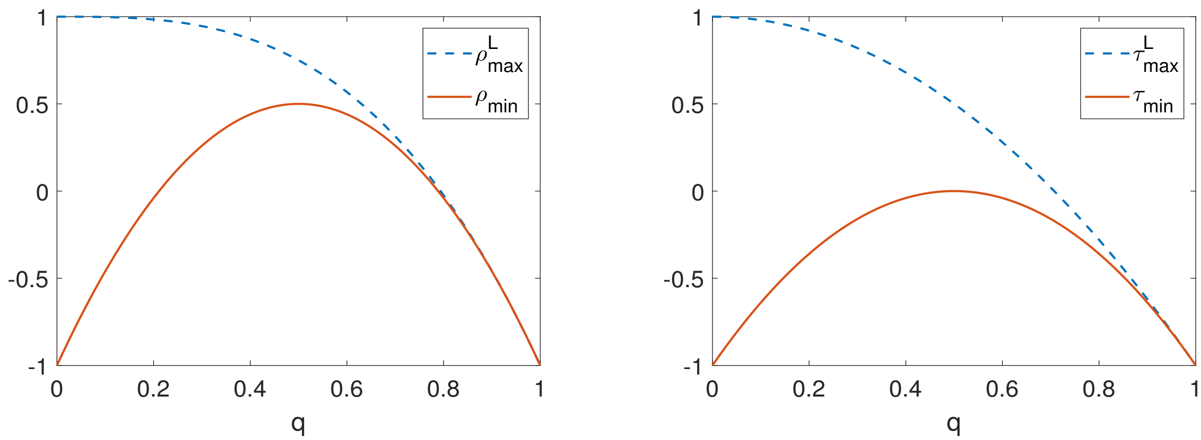
and

$$\tau_{min} = \tau(C^q_{min}) = -4q\,(q-1) - 1, \quad \tau^L_{max} = \tau(C^{q,L}_{max}) = 1 - 2q^2. \qquad (3.2.39)$$

As compared to the case of upper VaR$^+$, for high probability levels typically used in risk management, the interval $[\delta_{min}, \delta^L_{max}]$ is rather small as shown in Figures 3.8 and 3.9. Indeed, when q is high, C^q_{min} and $C^{q,L}_{max}$ are not very different (they only differ on $[q,1]$), which implies that δ_{min} and δ^L_{max} are similar. By contrast, when q is low the interval becomes wider. In summary, the information coming from the correlation constraint makes it possible to improve either the unconstrained VaR upper bound or the unconstrained VaR lower bound and the latter is the typical situation in risk management applications (as the probability level q used is usually high).

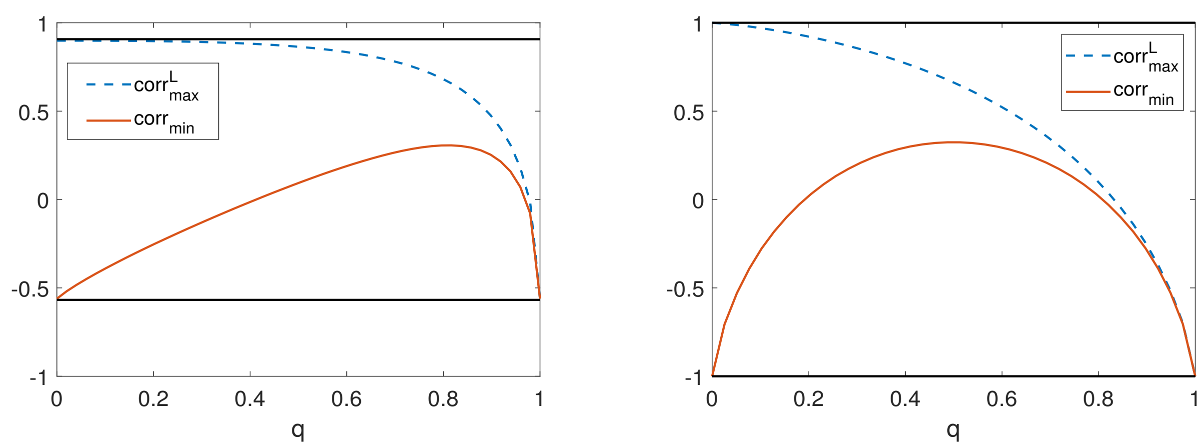
We point out that Lemma 3.2.8 only gives a sufficient condition for a copula of a pair (X_1, X_2) (with given marginal distributions) such that the VaR of their sum is equal to the unconstrained VaR lower bound : A full characterization of the copulas leading to the VaR lower bound is still an open problem (the same assertion can be made for the VaR upper bound). This implies that there

Figure 3.8: We display as a function of q the ranges $[\rho_{min}, \rho^L_{max}]$ and $[\tau_{min}, \tau^L_{max}]$. When the dependence measure takes value in $[\rho_{min}, \rho^L_{max}]$ resp. $[\tau_{min}, \tau^L_{max}]$, the constrained lower bound $\underline{\text{VaR}}^d_q$ does not improve on the unconstrained lower bound $\underline{\text{VaR}}_q$.



Panel A: $[\rho_{min}, \rho^L_{max}]$ w.r.t. q Panel B: $[\tau_{min}, \tau^L_{max}]$ w.r.t. q

Figure 3.9: Illustration of Theorem 3.2.9 for the Pearson correlation. We display the range of values $[\text{corr}_{min}, \text{corr}^L_{max}]$ as a function of q, which leads to no improvement in the lower VaR bound. Panel A corresponds to $X_1 \sim Gamma(2,3)$ and $X_2 \sim Lognormal(2,1)$ and Panel B corresponds to two normal distributions $N(3,2)$ and $N(2,4)$. In both Panels, the upper and lower black lines describe the highest and the lowest levels of correlation attainable with the given marginal distributions, respectively.



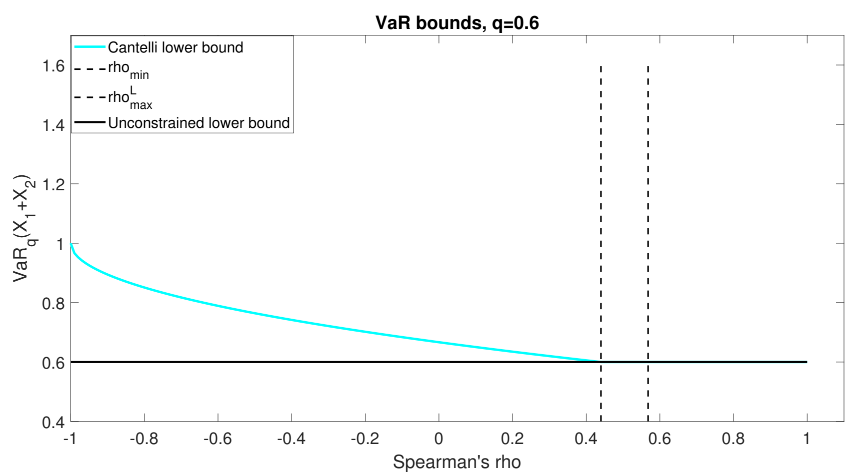
Panel A Panel B

could exist copulas for (X_1, X_2) such that for the dependence constraint δ at hand, $\delta(X_1, X_2) \notin [\delta_{min}, \delta^L_{max}]$ but still leading to a VaR for the sum that coincides with the unconstrained VaR lower bound. Nonetheless, we want to give an example of marginal distributions for which we know this cannot hold. To this end, consider the case in which $X_j \sim \mathcal{U}(0,1)$ for $j = 1, 2$ and let $\rho(X_1, X_2) = d$, $d \in [0, 1]$. As pointed out in section 3.2.1, it holds that

$$\underline{\text{VaR}}^d_q \geqslant \max\left\{1 - \sqrt{\frac{1+d}{6}}\sqrt{\frac{1-q}{q}}, \underline{\text{VaR}}_q\right\}. \tag{3.2.40}$$

where the first quantity on the right hand side corresponds to the Cantelli moment lower bound for VaR and note that $\underline{\text{VaR}}_q = q$.

Figure 3.10: Cantelli lower bound on $\text{VaR}_q(S)$ when $S = X_1 + X_2$ with $X_i \sim \mathcal{U}(0,1)$ and $q = 0.6$. The light-blue line represents the Cantelli bound defined in (3.2.40), also expressed in terms of Spearman's rho. The horizontal black line describes the unconstrained VaR lower bound. The dashed vertical lines represent ρ_{min} and ρ^L_{max}.



We plot this bound (right hand side of (3.2.40)) in Figure 3.10. Figure 3.10 shows that when $\rho < \rho^L_{max}$ the VaR of the sum $X_1 + X_2$ will be strictly greater than the unconstrained VaR lower bound. However, it could still be possible that there exist copulas with $\rho(X_1, X_2) > \rho^L_{max}$ such that the VaR of the sum $X_1 + X_2$ coincides with the VaR unconstrained lower bound. Proposition 3.2.11 shows however that this is not possible in the case of standard uniformly distributed risks.

Proposition 3.2.11. *Let $q \in (0, 1)$ and let $X_i \sim \mathcal{U}(0, 1)$ for $i = 1, 2$ with a copula C. Then,*

$$VaR_q(X_1 + X_2) = \underline{VaR}_q \quad \Longleftrightarrow \quad C \in \mathscr{C}^L(q). \tag{3.2.41}$$

The proof of Proposition 3.2.11 is given in Appendix 3.5.5.

3.2.3 Upper TVaR bound

In this section we extend the results obtained for the VaR upper bound to the case of TVaR. Hence, let $\varrho(\cdot) = \text{TVaR}_q$ for $q \in (0,1)$ and denote the solution $\overline{\varrho}^d$ to the constrained upper bound problem (3.2.1) by $\overline{\text{TVaR}}_q^d$ whereas the solution $\overline{\varrho}$ for the unconstrained case is denoted as $\overline{\text{TVaR}}_q$.

It is well-known that TVaR is a subadditive risk measure, i.e., for every random pair (X_1, X_2) it holds that

$$\text{TVaR}_q(X_1 + X_2) \leqslant \text{TVaR}_q(X_1) + \text{TVaR}_q(X_2), \tag{3.2.42}$$

and moreover that equality is obtained if X_1 and X_2 are comonotonic. Hence, we obtain the following formula for $\overline{\text{TVaR}}_q$ as solution to the unconstrained version of Problem (3.2.1),

$$\overline{\text{TVaR}}_q = \text{TVaR}_q(X_1) + \text{TVaR}_q(X_2)$$

and $\overline{\text{TVaR}}_q$ is attained if $X_1 \sim F_1$ and $X_2 \sim F_2$ are comonotonic. Nonetheless, comonotonicity is a sufficient but not necessary condition to maximize the TVaR. Indeed, Wang and Zikitis (2020b) obtained a complete characterization of the dependence structures between $X_1 \sim F_1$ and $X_2 \sim F_2$ such that $\text{TVaR}_q(X_1 + X_2) = \overline{\text{TVaR}}_q$. In this regard, the notion of q-concentrated random pairs (X_1, X_2) turns out to be useful (Wang and Zikitis (2020b)).

Definition 3.2.12. *A random pair (X_1, X_2) is said to be q-concentrated if and only if*

$$A^q_{X_1} = A^q_{X_2} \ a.s. \tag{3.2.43}$$

where $A^q_{X_i} = \{\omega \in \Omega \mid X_i(\omega) > \text{VaR}_q(X_i)\}, i = 1, 2$.

Thus, q-concentrated random pairs (X_1, X_2) are such that if one of the two random variables takes a high value (i.e., higher than its VaR at level q), then also the other random variable takes a high value (i.e., also higher than its VaR at level q). Wang and Zikitis (2020b) described this condition by saying that they share a q-tail event. In what follows, we use an equivalent formulation of q-concentration, expressed in terms of copulas.

Lemma 3.2.13. *Let $q \in (0,1)$, and (X_1, X_2) be a random vector with a copula C.*

$$(X_1, X_2) \text{ is } q\text{-concentrated} \iff C(q,q) = q. \tag{3.2.44}$$

Proof. This is a special case of Theorem 3 in Wang and Zikitis (2020b). □

Lemma 3.2.14. *Let $q \in (0,1)$, and (X_1, X_2) is a random vector having a copula C. Then,*

$$\text{TVaR}_q(X_1 + X_2) = \text{TVaR}_q(X_1) + \text{TVaR}_q(X_2) \iff C(q,q) = q. \tag{3.2.45}$$

We denote $\mathscr{C}_{\text{TVaR}}(q)$, $q \in (0,1)$, as the set of copulas maximizing the $\text{TVaR}_q(X_1 + X_2)$ under the constraint $X_1 \sim F_1$ and $X_2 \sim F_2$. Lemma 3.2.13 implies that

$$\mathscr{C}_{\text{TVaR}}(q) = \{\text{copula } C \mid C(q,q) = q\}. \tag{3.2.46}$$

Remark 3.2.2. *The condition $C(q,q) = q$ can be alternatively formulated using the concept of C-volume*[3]*, denoted by V_C. Specifically, the condition $C(q,q) = q$ is equivalent to each of the three following properties expressed in terms of V_C.*

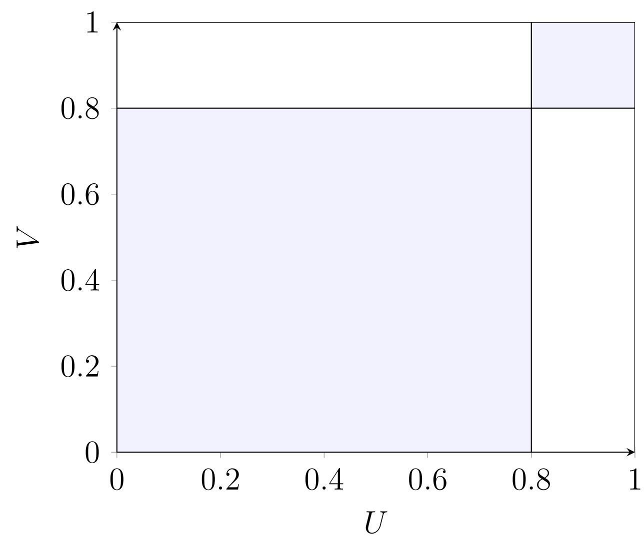
$$V_C([0,q] \times [q,1]) = 0,\ V_C([q,1] \times [0,q]) = 0 \text{ or } V_C([q,1] \times [q,1]) = 1 - q. \tag{3.2.47}$$

These equivalences are straightforward to prove. For instance, observe that $C(q,q) + V_C([0,q] \times [q,1]) = C(q,1) = q$ holds true for every copula. Consequently, $V_C([0,q] \times [q,1]) = 0 \iff C(q,q) = q$. A graphical illustration is provided in Figure 3.11.

Lemma 3.2.15. *Let $q \in (0,1)$. The set $\mathscr{C}_{TVaR}(q)$ strictly contains $\mathscr{C}(q)$ and is convex.*

The proof of Lemma 3.2.15 is given in Appendix 3.5.6. From Lemma 3.2.15, $\mathscr{C}(q)$ can thus be seen as the subset of $\mathscr{C}_{TVaR}(q)$ containing those copulas that exhibit an anti-monotonic dependence in the upper q-quadrant $[q,1]^2$ of their support. We are now ready for our main result concerning the solution to Problem (3.2.1) for $\varrho(\cdot) = \text{TVaR}_q$. Recall that C^c denotes the comonotonic copula and that the definition of C^q_{min} is given in (3.2.11).

Figure 3.11: This figure depicts the structure of copulas $C \in \mathscr{C}_{TVaR}(q)$. We consider the case $q = 0.8$, but other cases are similar. As stated in (3.2.47), the C-volumes of the two white rectangles $[0, 0.8] \times [0.8, 1]$ and $[0.8, 1] \times [0, 0.8]$ must be 0. Therefore, a copula is in $\mathscr{C}_{TVaR}(0.8)$ if and only if its support is contained in the two light-blue squares.



Theorem 3.2.16. *Given $q \in (0,1)$, $X_1 \sim F_1$ and $X_2 \sim F_2$, let δ be a measure of dependence that satisfies assumption 3.2.1. Let also $\delta_{min} = \delta(C^q_{min}, F_1, F_2)$ and $\delta^c = \delta(C^c, F_1, F_2)$. For every $d \in [\delta_{min}, \delta^c]$ it holds that*

$$\overline{TVaR}^d_q = \overline{TVaR}_q. \tag{3.2.48}$$

[3]The C-volume of a rectangle $[x_1, x_2] \times [y_1, y_2]$ with $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$ is defined as $V_C([x_1, x_2] \times [y_1, y_2]) = C(x_2, y_2) - C(x_2, y_1) - C(x_1, y_2) + C(x_1, y_1)$. See Definition 2.1.1 in Nelsen (2010) for further details.

Moreover, there exists a copula C such that $\delta(C, F_1, F_2) = d$ and $TVaR_q(X_1 + X_2) = \overline{TVaR}_q$ if and only if $d \in [\delta_{min}, \delta^c]$.

The proof of Theorem 3.2.16 is given in Appendix 3.5.7. The role of Theorem 3.2.16 in the analysis of the constrained TVaR problem is similar to the role of Theorem 3.2.3 for dealing with the constrained VaR problem. The interval $[\delta_{min}, \delta^c]$ describes a range of values for d such that the dependence information coming from the constraint does not make it possible to reduce the unconstrained upper bound $\overline{\text{TVaR}}_q$. A slight difference between Theorem 3.2.16 and Theorem 3.2.3 is that in the TVaR case the unconstrained upper bound can be achieved if and only if $d \in [\delta_{min}, \delta^c]$, while we have only a sufficient condition for the VaR$^+$ upper bound achievement. This is due to the fact that a complete characterization of the dependence structure maximizing TVaR$_q(X_1+X_2)$ under the constraint $X_1 \sim F_1$ and $X_2 \sim F_2$ is available in the literature (Wang and Zikitis (2020b)), while this is not the case for VaR$_q^+(X_1 + X_2)$.

It is interesting to notice that for each given $q \in (0, 1)$, the interval $[\delta_{min}, \delta^c]$ is bigger than $[\delta_{min}, \delta_{max}]$. This is a direct consequence of the well-known fact that every copula is pointwise dominated by the comonotonic copula. As all our measures of dependence that we consider are consistent with pointwise ordering on copulas (see Assumption 3.2.1), it thus follows that $\delta_{max} < \delta^c$ and moreover that $\rho^c = \tau^c = 1$, $q \in (0, 1)$. We also point out that even though the length of the interval $[\delta_{min}, \delta_{max}]$ tends to zero as q approaches zero (see also Figure 3.2), the interval $[\delta_{min}, \delta^c]$ can be wide even in this case. Take for instance the case of Spearman's rho: ρ_{min} (equation (3.2.19)) goes to -1 when q approaches 0 but ρ^c is always equal to 1.

It is not clear how to generalize the results for the constrained upper TVaR to the case of the constrained lower TVaR and we leave this problem for further research.

3.3 RVaR risk bounds with n risks

In this section, we study the RVaR of a portfolio of n risks $X_i \sim F_i$ under correlation information. Specifically, we use a weighted average of pairwise Pearson correlations corr(X_i, X_j) as the source of dependence information.

Definition 3.3.1. *Given a random vector $\mathbf{X} = (X_1, ..., X_n)$, the average correlation of $\mathbf{X}$, acorr $(\mathbf{X})$, is defined as*

$$acorr\,(\mathbf{X}) = \frac{\sum_{i\neq j} corr(X_i, X_j) std(X_i) std(X_j)}{\sum_{i\neq j}^{n} std(X_i) std(X_j)}. \tag{3.3.1}$$

Note that for $n = 2$, the average correlation acorr(X_1, X_2) coincides with the pairwise Pearson correlation corr(X_1, X_2). Furthermore, as the marginal distributions F_i of the X_i are given, the variance of the sum, var$(X_1 + ... + X_n)$, and the average correlation acorr$(\mathbf{X})$ are directly related:

$$\text{acorr}(\mathbf{X}) = \frac{var(S) - \sum_{i=1}^{n} var(X_i)}{\sum_{i\neq j}^{n} \text{std}(X_i)\text{std}(X_j)}. \tag{3.3.2}$$

Knowledge of the average correlation among the X_i (equivalently, knowledge of the variance of the portfolio sum $X_1 + ... + X_n$ is a fairly reasonable assumption that can be made. By contrast, knowledge of all pairwise correlations among the risks is quite ambitious. For instance, in a credit risk context, knowledge of pairwise correlations is essentially equivalent to knowledge of all single and pairwise default probabilities among obligors, which is difficult to achieve, as (joint) default events are scarce events. The average correlation acorr($\mathbf{X}$) reaches its highest value when the variance of $X_1 + ... + X_n$ is maximum, that is when the risks X_i are comonotonic. As for the lowest possible value of acorr($\mathbf{X}$), it is clear that acorr($\mathbf{X}$) is always higher than $\frac{-\sum_{i=1}^{n} var(X_i)}{\sum_{i\neq j}^{n} \text{std}(X_i)\text{std}(X_j)}$, as this corresponds to the case in which var$(X_1 + ... + X_n) = 0$. In general, it is however not always possible to construct a dependence among risks $X_i \sim F_i$ such that their sum has zero variance (consider for example the case in which the marginal distributions F_i are bounded to the left but unbounded to the right). In fact, the problem of finding the minimum variance of a sum for given marginal distributions of its components is generally unresolved[4] The constrained risk bounds problems we study can be formulated as follows.

$$\begin{aligned} \overline{\text{RVaR}}^{d}_{q,q'} := \sup \quad & \text{RVaR}_{q,q'}(S) \\ \text{subject to} \quad & X_j \sim F_j, \\ & \text{acorr}\,(X_1, ..., X_n) \leqslant d, \end{aligned} \tag{3.3.3}$$

$$\begin{aligned} \underline{\text{RVaR}}^{d}_{q,q'} := \inf \quad & \text{RVaR}_{q,q'}(S) \\ \text{subject to} \quad & X_j \sim F_j, \\ & \text{acorr}\,(X_1, ..., X_n) \leqslant d, \end{aligned} \tag{3.3.4}$$

When we omit the dependence constraint, we specifically denote the best-possible bounds by $\underline{\text{RVaR}}_{q,q'}$ and $\overline{\text{RVaR}}_{q,q'}$, respectively. In what follows, we tacitly assume that these problems are well-posed in that the constraint (i.e., the value for d) on the average correlation constraint is high enough.

Given the one-to-one relationship between acorr$(X_1, ..., X_n)$ and var$(X_1 + ... + X_n)$, problems (3.3.3) and (3.3.4) amount to finding RVaR bounds for portfolio sums under knowledge of their variance as well as the marginal distributions of its components and we thus merely generalize the problems on VaR bounds studied by Bernard et al. (2017) to the case of RVaR. Bounds for RVaR for portfolios under knowledge of the components' marginal distributions have not yet received a lot of attention in the literature and formulas for $\underline{\text{RVaR}}^{d}_{q,q'}$ and $\overline{\text{RVaR}}^{d}_{q,q'}$ are missing.[5] We first study the unconstrained case and the study of the constrained bounds follows.

In our analysis we will use the Left Tail-Value-at-Risk (LTVaR) at probability level $q \in (0, 1)$, defined by

$$\text{LTVaR}_q(X) = \frac{1}{q} \int_0^q \text{VaR}_\gamma(X)\, d\gamma,$$

and note that $\mathbb{E}(X) = q\text{LTVaR}_q(X) + (1 - q)\text{TVaR}_q(X)$.

[4]See e.g., Puccetti and Wang (2015b) and the references therein for a detailed study. A possible approach to estimate the minimum achievable variance is the RA algorithm (Puccetti and Rüschendorf (2012)).

[5]Embrechts et al. (2018) obtained some inequality results that yield upper bounds for RVaR of the sum under sole knwowledge of marginal distributions of the sum's components; see also Remark 3.3.1.

3.3.1 Unconstrained RVaR bounds

In order to study the unconstrained bounds $\underline{\text{RVaR}}_{q,q'}$ and $\overline{\text{RVaR}}_{q,q'}$, we first recall the concept of (tail) mixability (see Wang and Wang (2011), Wang et al. (2013) and Embrechts et al. (2013)) which is known to play a central role in many risk aggregation problems. Consider n random variables $X_1, ..., X_n$ and $S = \sum_{i=1}^n X_i$. Given $c \in \mathbb{R}$ and $q \in (0,1)$, we say that the $X_1, ..., X_n$ are q-upper tail mixing with mixing constant c, if

$$P(S = c \,|S \geqslant \text{VaR}_q^+(S)) = 1.$$

Analogously, we say that the $X_1, ..., X_n$ are q-lower tail mixing with mixing constant c, if

$$P\left(S = c \,|S \leqslant \text{VaR}_q(S)\right) = 1.$$

In what follows, for the ease of presentation we will also say that in these instances S is mixing in the upper q-part of its distribution resp. lower q-part of its distribution with mixing constant c. Observe that given a set of random variables $X_1, ..., X_n$, the existence and the analytical expression of a copula among the X_i leading to the mixability of their sum depend on the marginal distributions F_i of the X_i at hand. For instance, when one deals with two risks that are uniformly distributed on the interval $[0,1]$, their sum is mixing in the upper q-part of its distribution for any copula in the set $\mathscr{C}(q)$ (defined in (3.2.10)). Such property however cannot hold for instance if the two random risks are lognormally distributed.

In the special case in which RVaR reduces to VaR, formulas for $\underline{\text{RVaR}}_{q,q'}$ and $\overline{\text{RVaR}}_{q,q'}$ exist in the homogeneous setting (i.e., all F_i are identical); see Embrechts and Puccetti (2006), Wang et al. (2013) and Puccetti and Rüschendorf (2013b) whereas in the heterogeneous set-up algorithms such as the Rearrangement Algorithm (RA) can be used to obtain approximations of good quality. Furthermore, in the special case of TVaR bounds, it is well-known that the upper bound writes as the sum of the TVaRs of the components and is attained under a comonotonic dependence among the risks.

Proposition 3.3.2 (RVaR bounds with given marginals)**.** *Let* $0 < q < q' < 1$. *It holds that*

$$\underline{\mathit{RVaR}}_{q,q'} \geqslant A(q') := \sum_{i=1}^n \mathit{LTVaR}_{q'}(X_i), \tag{3.3.5}$$

$$\overline{\mathit{RVaR}}_{q,q'} \leqslant B(q) := \sum_{i=1}^n \mathit{TVaR}_q(X_i). \tag{3.3.6}$$

Furthermore, the lower bound $A(q')$ is attained if and only if there exists a copula such that S is mixing in the lower q'-part of its distribution with mixing constant $A(q')$. The upper bound $B(q)$ is attained if and only if there exists a copula such that S is mixing in the upper q-part of its distribution with mixing constant $B(q)$.

Proof. For any set of random variables $X_1, ..., X_n$ with sum $S = X_1 + ... + X_n$ it holds that

$$\text{VaR}_q^+(S) \leqslant \text{RVaR}_{q,q'}(S) \leqslant \text{TVaR}_q(S) \leqslant \sum_{i=1}^{n} \text{TVaR}_q(X_i) = B(q), \tag{3.3.7}$$

and in a similar way

$$A(q') = \sum_{i=1}^{n} \text{LTVaR}_{q'}(X_i) \leqslant \text{LTVaR}_{q'}(S) \leqslant \text{RVaR}_{q,q'}(S) \leqslant \text{VaR}_{q'}(S). \tag{3.3.8}$$

From Theorems 1 and 2 in Bernard et al. (2017), we know that the mixability of S in the upper q-part of its distribution with mixing constant $B(q)$ is equivalent to $\text{VaR}_q^+(S) = B(q)$ and that when S is mixing in the lower q'-part of its distribution, with mixing constant $A(q')$ then $\text{VaR}_{q'}(S) = A(q')$. Hence, from (3.3.7) and (3.3.8),

$$\text{VaR}_q^+(S) = B(q) \implies \text{RVaR}_{q,q'}(S) = B(q), \tag{3.3.9}$$

$$\text{VaR}_{q'}(S) = A(q') \implies \text{RVaR}_{q,q'}(S) = A(q'). \tag{3.3.10}$$

We only need to show the opposite implications, i.e., that $\text{RVaR}_{q,q'}(S) = B(q) \implies \text{VaR}_q^+(S) = B(q)$ and $\text{RVaR}_{q,q'}(S) = A(q') \implies \text{VaR}_{q'}(S) = A(q')$. To this end, we proceed by contradiction. Let us thus assume that S is such that $\text{RVaR}_{q,q'}(S) = B(q)$ and $\text{VaR}_q^+(S) < B(q)$. These two conditions together imply $\text{VaR}_{q'}^+(S) > B(q), q' < q^* < 1$. It is easy to see that $\text{RVaR}_{q,q^*}(S)$ can be expressed as follows:

$$\text{RVaR}_{q,q^*}(S) = \frac{q' - q}{q^* - q}\text{RVaR}_{q,q'}(S) + \frac{q^* - q'}{q^* - q}\text{RVaR}_{q',q^*}(S). \tag{3.3.11}$$

Thus, $\text{RVaR}_{q,q^*}(S)$ can be seen as a mixture of $\text{RVaR}_{q,q'}(S)$ and $\text{RVaR}_{q',q^*}(S)$, with weights $\frac{q'-q}{q^*-q} > 0$ and $\frac{q^*-q'}{q^*-q} > 0$ that satisfy $\frac{q'-q}{q^*-q} + \frac{q^*-q'}{q^*-q} = 1$. Note $\text{VaR}_{q'}^+(S) > B(q) \implies \text{RVaR}_{q',q^*}(S) > B(q)$. This last inequality together with the assumption $\text{RVaR}_{q,q'}(S) = B(q)$ and the relation (3.3.11) implies $\text{RVaR}_{q,q^*}(S) > B(q) = \sum_{i=1}^{n} \text{TVaR}_q(X_i)$, which is a contradiction of (3.3.7).

Hence, we conclude that S cannot be such that $\text{RVaR}_{q,q'}(S) = B(q)$ and $\text{VaR}_q^+(S) < B(q)$ holds simultaneously. Since it must hold that either $\text{VaR}_q^+(S) < B(q)$ or $\text{VaR}_q^+(S) = B(q)$, we deduce that

$$\text{RVaR}_{q,q'}(S) = B(q) \implies \text{VaR}_q^+(S) = B(q). \tag{3.3.12}$$

Putting together (3.3.9) and (3.3.12), we obtain that $\text{RVaR}_{q,q'}(S) = B(q) \iff \text{VaR}_q^+(S) = B(q) \iff S$ is mixing in the upper q-part of its distribution with mixing constant $B(q)$. The proof for the lower bound $A(q')$ is similar and thus we omit it. □

From Proposition 3.3.2 and its proof, we obtain that if there exists a copula such that S is mixing in the q-upper part of its distributions the worst-case VaR_q^+, TVaR_q and $\text{RVaR}_{q,q'}$ coincide, i.e.,

$\overline{\text{VaR}}_q^+ = \overline{\text{RVaR}}_{q,q'} = \overline{\text{TVaR}}_q$ and similar for lower bounds. As pointed out in Bernard et al. (2017), tail mixability is a strong assumption when one is dealing with a small number of risks. Nonetheless, when the portfolio size is large enough, there is numerical evidence that tail mixability can approximately be obtained.

Remark 3.3.1 (Comparison with Embrechts et al. (2018))**.** *Embrechts et al. (2018) proved that RVaR satisfies a special form of subadditivity. With our parametrization, Theorem 1 in their paper becomes*

$$RVaR_{q,q'}(S) \leqslant \sum_{i=1}^{n} RVaR_{q(n),q'(n)}(X_i) \tag{3.3.13}$$

where $q(n) = 1 - \frac{1-q'}{n} - q' + q$ *and* $q'(n) = 1 - \frac{1-q'}{n}$*. For all* $n \in \mathbb{N}$*,* $q(n) \geqslant q$ *and* $q'(n) \geqslant q'$*. Embrechts et al. (2018) did not discuss whether the bound* (3.3.13) *was best possible. Both bounds* $\sum_{i=1}^{n} RVaR_{q(n),q'(n)}(X_i)$ *and* $\sum_{i=1}^{n} TVaR_q(X_i)$ *depend solely on the marginal distributions and it is not clear which one is higher in general. However, the bound derived in Proposition 3.3.2, for large portfolios, is lower than the one proposed in Embrechts et al. (2018), regardless of the marginal distributions considered. Indeed, when* $n \longrightarrow +\infty$*, then* $q'(n) \longrightarrow 1$ *and* $q(n) \longrightarrow 1 - q' + q > q$*. Hence, for large portfolios the upper bound in Proposition 3.3.2 offers an improvement with respect to the one proposed in Theorem 1 of Embrechts et al. (2018). In particular, our bound* $B(q)$ *for* $\overline{RVaR}_{q,q'}$ *is asymptotically smaller than the one derived in Embrechts et al. (2018) as*

$$\lim_{n\to+\infty} \frac{\sum_{i=1}^{n} RVaR_{q(n),q'(n)}(X_i)}{\sum_{i=1}^{n} TVaR_q(X_i)} = \lim_{n\to+\infty} \frac{\sum_{i=1}^{n} TVaR_{1-q'+q}(X_i)}{\sum_{i=1}^{n} TVaR_q(X_i)} \geqslant 1.$$

3.3.2 Constrained RVaR bounds

We focus now on upper and lower bounds on RVaR for a portfolio when the marginal distributions F_i of the components X_i are known and their average correlation acorr$(X_1, ..., X_n)$ is lower than a certain value. To this end, it is useful to recall RVaR bounds under the sole knowledge of the portfolio's mean and variance.

Hence, we denote by $D_p(a, b)$, $p \in (0, 1)$, a two-point distribution having mass points values equal to $a, b \in \mathbb{R}$, with probability p and $1 - p$, respectively. Furthermore, for each $\mu \in \mathbb{R}$ and $s^2 \geqslant 0$, let $V(\mu, s^2)$ be the set of random variables having mean μ and variance s^2, i.e.,

$$V(\mu, s^2) = \left\{X : \mathbb{E}(X) = \mu,\ var(X) = s^2\right\}. \tag{3.3.14}$$

Lemma 3.3.3 (Moment bounds)**.** *Let* $X \in V(\mu, s^2)$ *with* $\mu \in \mathbb{R}$ *and* $s^2 \geqslant 0$*, and let* $0 < q < q' < 1$*. Then,*

$$\mu - s\sqrt{\frac{1-q'}{q'}} \leqslant RVaR_{q,q'}(X) \leqslant \mu + s\sqrt{\frac{q}{1-q}}. \tag{3.3.15}$$

Proof. See Li et al. (2018) for the upper bound and Bernard et al. (2020b) for the lower bound. □

The stated lower bound in Lemma 3.3.3 is attained for $X \in V(\mu, s^2)$ that is distributed with $D_{q'}\left(\mu - s\sqrt{\frac{1-q'}{q'}}, \mu + s\sqrt{\frac{q'}{1-q'}}\right)$ and the upper bound is attained when $X \in V(\mu, s^2)$ has distribution $D_q\left(\mu - s\sqrt{\frac{1-q}{q}}, \mu + s\sqrt{\frac{q}{1-q}}\right)$.

Let $X_i \sim F_i$, $S = X_1 + ... + X_n$, $\mu = \mathbb{E}(S)$. For any $p \in (0,1)$ and $d \in \mathbb{R}$, we define the quantities

$$\begin{aligned} l(p) &:= \max\left(\mu - \sqrt{d\sum_{i\neq j}^{n} \text{std}(X_i)\text{std}(X_j) + \sum_{i=1}^{n} var(X_i)}\sqrt{\frac{1-p}{p}}, A(p)\right), \\ u(p) &:= \min\left(\mu + \sqrt{d\sum_{i\neq j}^{n} \text{std}(X_i)\text{std}(X_j) + \sum_{i=1}^{n} var(X_i)}\sqrt{\frac{p}{1-p}}, B(p)\right), \\ c(p) &:= \frac{p\left(A(p)-\mu\right)^2 + (1-p)\left(B(p)-\mu\right)^2 - \sum_{i=1}^{n} var(X_i)}{\sum_{i\neq j}^{n} \text{std}(X_i)\text{std}(X_j)}. \end{aligned} \tag{3.3.16}$$

and note that the expressions for $B(p)$ and $A(p)$ were given in (3.3.5) and (3.3.6). When $p = 1$, we set $l(1) = \mu$, $u(1) = \mu$.

Proposition 3.3.4 (RVaR bounds with given marginals and the average correlation constraint)**.** *Let* $0 < q < q' < 1$. *It holds that*

$$\underline{RVaR}^d_{q,q'} \geqslant l(q'). \tag{3.3.17}$$

$$\overline{RVaR}^d_{q,q'} \leqslant u(q). \tag{3.3.18}$$

Furthermore, the equality $\underline{RVaR}^d_{q,q'} = l(q')$ *holds if there exists a copula such that* $acorr(\mathbf{X}) \leqslant d$ *and* S *is mixing in the lower* q'*-part of its distribution with mixing constant* $l(q')$. *The equality* $\overline{RVaR}^d_{q,q'} = u(q)$ *holds if there exists a copula such that* $acorr(\mathbf{X}) \leqslant d$ *and* S *is mixing in the upper* q*-part of its distribution with mixing constant* $u(q)$.

The proof is given in Appendix 3.5.8.

Remark 3.3.2. *Proposition 3.3.4 makes it possible to identify the levels of average correlation that allow reducing the bounds* $A(q')$ *and/or* $B(q)$ *for RVaR. These threshold values are* $c(q')$ *and* $c(q)$*, respectively. When* $d \geqslant \max(c(q), c(q'))$ *the availability of average correlation does no longer make it possible to improve the bounds* $A(q')$ *and* $B(q)$. *In particular, if* $d \geqslant c(q)$ *then* $u(q) = B(q)$ *and if* $d \geqslant c(q')$*, then* $l(q') = A(q')$. *A numerical illustration regarding the behaviour of the threshold values corresponding to various probability levels is given in Figure 3.12.*

We can now derive bounds for VaR and TVaR as limiting case of the RVaR bounds proposed in Proposition 3.3.4.

Proposition 3.3.5. *Let* $0 < q < q' < 1$, $q \in (0,1)$. *It holds that*

$$l(q) \leqslant \underline{VaR}^d_q \leqslant \overline{VaR}^d_q \leqslant u(q). \tag{3.3.19}$$

Moreover,

$$\mu = l(1) \leqslant \underline{\mathit{TVaR}}_q^d \leqslant \overline{\mathit{TVaR}}_q^d \leqslant u(q). \tag{3.3.20}$$

Proof. Let $q \in (0,1)$, $X_i \sim F_i$ $(i = 1, 2, ..., n)$ satisfy $\text{acorr}(X_1, X_2, ..., X_n) \leqslant d$, $\mu = \sum_{i=1}^n \mathbb{E}(X_i)$ and $S = \sum_{i=1}^n X_i$. Then, using Proposition 3.3.4 and the inequalities in (3.3.7) and (3.3.8), we obtain

$$l(q) \leqslant \lim_{r \nearrow q} \underline{\text{RVaR}}_{r,q}^d \leqslant \underline{\text{VaR}}_q^d \leqslant \text{VaR}_q(S) \leqslant \text{VaR}_q^+(S) \leqslant \overline{\text{VaR}}_q^d \leqslant \lim_{r \searrow q} \overline{\text{RVaR}}_{q,r}^d \leqslant u(q),$$

and

$$\mu \leqslant \lim_{r \nearrow 1} \underline{\text{RVaR}}_{q,r}^d \leqslant \underline{\text{TVaR}}_q^d \leqslant \text{TVaR}_q(S) \leqslant \overline{\text{TVaR}}_q^d \leqslant \lim_{r \nearrow 1} \overline{\text{RVaR}}_{q,r}^d \leqslant u(q).$$

□

The bounds in Proposition 3.3.4 are not best-possible in general, but we provide numerical evidence that they can be a good approximation for the solution to problems (3.3.3) and (3.3.4)

Numerical study

The goal of this numerical study is twofold. First, in Tables 3.2 and 3.3 we give numerical evidence that the bounds derived in Proposition 3.3.4 can be a very good approximation for the best-possible bounds $\underline{\text{RVaR}}_{q,q'}^d$ and $\overline{\text{RVaR}}_{q,q'}^d$. Second, in Figure 3.12 and the subsequent discussion we show how our results can be used to identify under which circumstances the dependence constraint $\text{acorr}(\mathbf{X}) \leqslant d$ has an impact on unconstrained RVaR bounds. Specifically, our analysis indicates that for q and q' close to 1 the average correlation constraint does not contain enough dependence information to impact the unconstrained RVaR bounds.

The numerical procedure adopted to obtain the RVaR bounds numerical approximations displayed in Table 3.2 and Table 3.3 builds on the idea, proposed in Bernard et al. (2018a), to use the RA to infer the dependence structure amongst $(X_1, X_2, ..., X_n)$ that makes the distribution function of $S = \sum_{i=1}^n X_i$ as close as possible to a certain target distribution (see also Bernard et al. (2017)). In our illustration of this numerical procedure, we focus on the RVaR upper bound as the lower bound numerical approximation can be computed in similar manner. Let us denote with Z a random variable having the two-point distribution $D_q(l(q), u(q))$ In order to obtain the numerical approximation of the solution of problem (3.3.3), we first apply the RA to the matrix obtained with the sample values of the random vector $(X_1, X_2, ..., X_n, -Z)$, and we then compute the $\text{RVaR}_{q,q'}(S)$ and $\text{acorr}(\mathbf{X})$ using the dependence structure obtained for $(X_1, X_2, ..., X_n)$ as output of the RA. If $\text{acorr}(\mathbf{X}) \leqslant d$, the numerical solution of problem (3.3.3) is found. If $\text{acorr}(\mathbf{X}) > d$, we iteratively repeat the procedure using $Z_\epsilon \sim D_q(l_\epsilon(q), u_\epsilon(q))$ with a gradually increasing $0 < \epsilon < d$ and

$$l_\epsilon(q) = \max\left(\mu - \sqrt{(d-\epsilon)\sum_{i \neq j}^n \text{std}(X_i)\text{std}(X_j) + \sum_{i=1}^n var(X_i)}\sqrt{\frac{1-q}{q}}, B(q)\right),$$

$$u_\epsilon(q) = \min\left(\mu + \sqrt{(d-\epsilon)\sum_{i\neq j}^{n}\text{std}(X_i)\text{std}(X_j) + \sum_{i=1}^{n} var(X_i)}\sqrt{\frac{q}{1-q}}, A(q)\right),$$

until the constraint on average correlation is met.

The rationale behind this procedure is the following: if the distribution of Z is admissible for the risks $(X_1, X_2, ..., X_n)$, meaning there exists a copula such that $S =^d Z$ with $S = \sum_{i=1}^n X_i$, then we know from Proposition 3.3.4 the upper-bound in (3.3.17) is best-possible. The numerical method described above aims to find the dependence structure such that distribution of S is as close as possible to the distribution of Z. Note that even when $D_q(l(q), u(q))$ is not admissible for S, the numerical procedure output gives a feasible distribution for S that can be used to check whether or not the numerical bounds are close to those derived in Proposition 3.3.4.

Table 3.2: Lognormal distribution. Bounds for $\text{RVaR}_{q,q'}(S)$, $S = \sum_{i=1}^n X_i$, $q = 0.95$, $q' = 0.98$, $X_i \sim Lognormal(\mu, \sigma)$ with $\mu = 2.5$ and $\sigma = 0.23$ for $i = 1, 2, ..., n$. In each table, the bounds obtained using the RA are reported under the column "RA", while the bounds computed as from Proposition 3.3.4, resp. Proposition 3.3.2 are reported under the column "Proposition 3.3.4", resp. "Proposition 3.3.2". In sub-tables 3.2a, 3.2b, 3.2c and 3.2d, we use values of d such that $d \leqslant \min(c(q), c(q'))$ for which we know $u(q) < B(q)$ and $l(q') > A(q')$. In sub-table 3.2e we assess whether the unconstrained bounds given in Proposition 3.3.2 can be attained. The number of discretization points is set equal to 10.000.

d	RA	Proposition 3.3.4
-0.4889	(37.48;40.79)	(37.42;40.81)
-0.4410	(37.29;45.04)	(37.28;45.08)
-0.3555	(37.15;49.28)	(37.14;49.34)
-0.2324	(37.01;53.54)	(37.00;53.61)

(a) Constrained RVaR bounds for $n = 3$.

d	RA	Proposition 3.3.4
-0.1029	(124.7;136.0)	(124.7;136.0)
-0.0674	(124.3;150.1)	(124.3;150.3)
-0.0041	(123.8;164.3)	(123.8;164.5)
0.0871	(123.3;178.6)	(123.3;178.7)

(b) Constrained RVaR bounds for $n = 10$.

d	RA	Proposition 3.3.4
-0.0128	(623.6;679.9)	(623.6;680.1)
0.0197	(621.3;750.6)	(621.3;751.3)
0.0779	(619.0;821.4)	(619.0;822.4)
0.1616	(616.7;892.3)	(616.6;893.5)

(c) Constrained RVaR bounds for $n = 50$.

d	RA	Proposition 3.3.4
-0.0026	(1247.0;1360.0)	(1247.0;1360.0)
0.0296	(1243.0;1502.0)	(1243.0;1503.0)
0.0871	(1238.0;1644.0)	(1238.0;1645.0)
0.1701	(1233.0;1786.0)	(1233.0;1787.0)

(d) Constrained RVaR bounds for $n = 100$.

	RA	Proposition 3.3.2
n=3	(36.99;58.60)	(36.99;58.92)
n=10	(123.3;196.4)	(123.3;196.4)
n=50	(616.5;982.0)	(616.5;982.0)
n=100	(1233;1964)	(1233;1964)

(e) Unconstrained RVaR bounds.

Table 3.3: Standard uniform distribution. Bounds for $\text{RVaR}_{q,q'}(S)$, $S = \sum_{i=1}^{n} X_i$, $X_i \sim \mathcal{U}[0,1]$ for $i = 1, 2, ..., n$ for $q = 0.90$ and $q' = 0.95$. In each table, the bounds obtained using the RA are reported under the column "RA", while the bounds computed as from Proposition 3.3.4, resp. Proposition 3.3.2 are reported under the column "Proposition 3.3.4", resp. "Proposition 3.3.2". In sub-tables 3.3a, 3.3b, 3.3c and 3.3d, we use values of d such that $d \leqslant \min(c(q), c(q'))$ for which we know $u(q) < B(q)$ and $l(q') > A(q')$. In sub-table 3.3e we assess whether the unconstrained bounds given in Proposition 3.3.2 can be attained. The number of discretization points is set equal to 10.000.

d	RA	Proposition 3.3.4
-0.4915	(1.485;1.696)	(1.485;1.696)
-0.4548	(1.466;1.951)	(1.466;1.951)
-0.3892	(1.446;2.205)	(1.446;2.206)
-0.2947	(1.427;2.460)	(1.427;2.461)

(a) Constrained RVaR bounds for $n = 3$.

d	RA	Proposition 3.3.4
-0.1048	(4.950;5.654)	(4.950;5.654)
-0.0776	(4.885;6.503)	(4.885;6.504)
-0.0290	(4.820;7.353)	(4.820;7.354)
0.0409	(4.755;8.203)	(4.755;8.203)

(b) Constrained RVaR bounds for $n = 10$.

d	RA	Proposition 3.3.4
-0.014	(24.75;28.27)	(24.75;28.27)
0.010	(24.43;32.52)	(24.43;32.52)
0.056	(24.10;36.77)	(24.10;36.77)
0.119	(23.78;41.02)	(23.78;41.02)

(c) Constrained RVaR bounds for $n = 50$.

d	RA	Proposition 3.3.4
0.003	(49.25;59.80)	(49.25;59.81)
0.030	(48.68;67.21)	(48.68;67.22)
0.072	(48.12;74.62)	(48.12;74.62)
0.128	(47.55;82.03)	(47.55;82.03)

(d) Constrained RVaR bounds for $n = 100$.

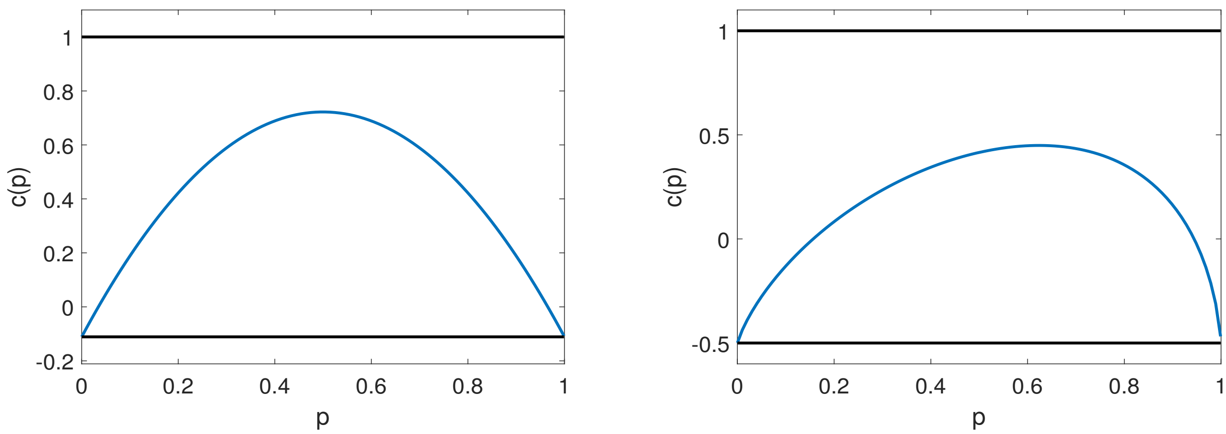
	RA	Proposition 3.3.2
n=3	(1.425; 2.849)	(1.425; 2.850)
n=10	(4.750; 9.499)	(4.750; 9.500)
n=50	(23.75; 47.50)	(23.75; 47.50)
n=100	(47.50; 95.00)	(47.50; 95.00)

(e) Unconstrained RVaR bounds.

In all cases reported in Tables 3.2 and 3.3, the bounds from Propositions 3.3.2 and 3.3.4 seem to be a good approximation of the bounds obtained numerically. For the sum of standard uniform distributions, this encouraging result is confirmed from a theoretical point of view in section 3.3.3, where we prove that in this case the bounds from Proposition 3.3.4 are in fact best-possible.

Figure 3.12 displays the quantity $c(p)$ defined in (3.3.16) for several values of $p \in (0, 1)$. As pointed in Remark 3.3.2, $c(p)$ allows us to identify which values of d are such that the constraint $\text{acorr}(\mathbf{X}) \leqslant d$ makes it possible that the RVaR bounds from Proposition 3.3.4 improve those obtained using solely the marginal distributions. Clearly, the analytical expression of $c(p)$ depends on the marginal distributions so each case needs its own evaluation. Nonetheless, we observe that in both cases considered in Figure 3.12, the function $c(p)$ exhibits a similar shape. We underline that in both cases reported in Figure 3.12, for p close to 1, $c(p)$ moves toward the average correlation

Figure 3.12: Illustration of the quantity $c(p)$ defined in (3.3.16) for $p \in (0,1)$. Panel A displays the case $X_i \sim \mathcal{U}[0,1]$ for $i = 1,2,...,n$ and $n = 10$. Panel B displays the case $X_i \sim Lognormal(\mu, \sigma)$, $\mu = 2.5$, $\sigma = 0.23$ for $i = 1,2,...,n$ and $n = 3$. In each panel, the two horizontal black lines describe the maximum and minimum level of acorr$(X_1, X_2, ..., X_n)$ for the given marginal distribution functions.



lower bound. This can be explained as follows. When the marginal distributions are bounded from the right, we have that $\lim_{p\nearrow 1} B(p) < +\infty$, and thus $\lim_{p\nearrow 1}(1-p)(B(p)-\mu)^2 = 0$. Moreover, $\lim_{p\nearrow 1}\ p(A(p)-\mu))^2 = 0$ always holds. From the expression of $c(p)$ given in (3.3.16), we deduce $\lim_{p\nearrow 1}\ c(p) = \frac{-\sum_{i=1}^{n} var(X_i)}{\sum_{i\neq j}^{n} \text{std}(X_i)\text{std}(X_j)}$. Note that this limiting value of $c(q)$ coincides with the average correlation lower bound. Any continuous distribution can be approximated with a discrete one with any degree of precision, but once the discretization is set, the approximate distributions we use to compute $c(p)$ and the RVaR bounds are bounded from the right, and thus the limits derived above are valid.

Therefore, regardless of the marginal distributions, for q and q' close to 1 one can reasonably expect that the inferred value for d will be higher than $c(q)$ and $c(q')$, and this implies that the bounds in Proposition 3.3.4 will coincide with $A(q')$ and $B(q)$, i.e., the unconstrained bounds. As for the case of the VaR upper bound, these conclusions are in line with the results obtained in Section 3.2.1, where we showed that for a high probability level, knowledge of the dependence constraint typically has no impact on the VaR upper bound. By contrast, in Section 3.2.2 we reached different conclusions for the VaR lower bound. This mismatch is determined by the fact that in case $n = 2$, we study the VaR lower bound with an equality constraint, while in the general case $n \geqslant 2$ we consider the RVaR lower bound under an inequality constraint, as stated in problem (3.3.4). From our results in Section 3.2.2, it is clear that if one re-formulates the constraint lower VaR problem using an inequality constraint in the form $\delta(C, F_1, F_2) \leqslant d$, then also the VaR lower bound for $n = 2$ can hardly be affected by the correlation constraint for q close to 1.

3.3.3 Standard uniform distributions

The bounds derived in Proposition 3.3.4 can offer a good approximation of the best-possible bounds, but they remain an approximation in general. To show that they are not best-possible in general, consider first the case in which $u(q)$ reduces to $B(q)$ (the case of unconstrained bounds). The upper bound $B(q)$ is then best-possible and even attainable when there exists a dependence among $Y_i \sim F_{i,q}$ such that $Y_1 + ... + Y_n$ is mixing with mixing constant $B(q)$, as in this case $\text{RVaR}_{q,q'}(X_1 + ... + X_n) = B(q)$ where $X_i = I_{U_i<q}F_{i,0}^{-1}(U_i) + I_{U_i \geqslant q}Y_i$. Here, U_i is standard uniformly distributed for $i = 1, 2, ..., n$, $F_{i,0}^{-1}$ denotes the restriction of the F_i^{-1} to $(0, q)$ and $F_{i,q}^{-1}$ is their restriction to $(q, 1)$. Puccetti and Wang (2015a) discuss extensively situations in which such dependence among $Y_i \sim F_{i,q}$ exists, the most relevant case is when the $F_{i,q}$ have decreasing density on a bounded support (e.g., Beta distributions, Uniform distributions). Importantly, mixing among the Y_i cannot be obtained when the F_i are unbounded to the right (e.g., normal or lognormal distributions). A similar reasoning also applies to the case of lower bounds.

The situation in which $u(q)$ does not reduce to $B(q)$ (case of constrained bounds) is interesting for practical purposes but to the best of our knowledge there are no results yet in the literature that demonstrate this bound is best-possible in some cases of interest. In the following theorem, we provide the first contribution to this kind of research. Motivated by the good numerical results displayed in Table 3.3, we completely solve the upper (lower) bound problem on RVaR for sums of uniformly distributed risks under a correlation constraint.

The proof of this result builds on a remarkable result of Mao et al. (2019) who proved that if $n \geqslant 3$ and $X_i \sim \mathcal{U}[0, 1]$ for $i = 1, 2, ..., n$, any distribution G lower than $U[0, n]$ in convex order belongs to the set of possible distributions for the sum $S = \sum_{i=1}^n X_i$.

When considering the sum of uniform distributions, the expressions of $l(p)$, $u(p)$ and $c(p)$, given in (3.3.16), become

$$\begin{aligned} l(p) &= \max\left(\frac{n}{2} - \sqrt{\frac{dn(n-1)+n}{12}}\sqrt{\frac{1-p}{p}}, \frac{n}{2}p\right), \\ u(p) &= \min\left(\frac{n}{2} + \sqrt{\frac{dn(n-1)+n}{12}}\sqrt{\frac{p}{1-p}}, \frac{n}{2}(1+p)\right), \\ c(p) &= \frac{3np(1-p)-1}{n-1}. \end{aligned} \tag{3.3.21}$$

Observe that $l(q')$ and $u(q)$ coincide with the unconstrained lower bound $A(q') = \frac{n}{2}q'$ and $B(q) = \frac{n}{2}(1+q)$, respectively, if and only if $d \geqslant c(q')$, resp. $d \geqslant c(q)$.

Theorem 3.3.6 (Best-possible RVaR bounds for sums of $n \geqslant 3$ uniformly distributed random variables under a correlation constraint)**.** *Let $X_i \sim \mathcal{U}[0, 1]$ for $i = 1, 2, ..., n$, acorr$(X_1, ..., X_n) \leqslant d$, and let $S = \sum_{i=1}^n X_i$. Then,*

$$\underline{RVaR}^d_{q,q'} = l(q')$$

and

$$\overline{RVaR}^d_{q,q'} = u(q).$$

Proof. We prove the result for case of the upper bound $u(q)$. The case of the lower bound is similar. Since the bounds $u(q)$ has two different analytical expressions according to the value of the constraint d we will consider these two cases separately.

In the first case: $d \geqslant c(q)$, we define a random variable S_1 having a two-point distribution described in (3.3.22):

$$S_1 = \begin{cases} \frac{n}{2}q & \text{with probability } q, \\ \frac{n}{2}(1+q) & \text{with probability } 1-q. \end{cases} \tag{3.3.22}$$

Furthermore, we have that for $p \in (q, 1)$

$$\text{TVaR}_p(S_1) = \frac{n}{2}(1+q) \leqslant \frac{n}{2}(1+p) = \text{TVaR}_p(U[0,n]).$$

and for $p \in (0, q]$

$$\text{TVaR}_p(S_1) = (1-p)\left((q-p)\frac{n}{2}q + (1-q)\,\frac{n}{2}(1+q)\right) = \\ \frac{n}{2}\frac{1-qp}{1-p} \leqslant \frac{n}{2}(1+p) = \text{TVaR}_p(U[0,n])$$

In other words, the condition $\text{TVaR}_p(S_1) \leqslant \text{TVaR}_p(U[0,n])$ holds for all $p \in (0,1)$, and this condition is equivalent to $S_1 \leqslant_{cx} U[0,n]$ (see Theorem 3.A.5 in Shaked and Shanthikumar (2007), for instance). Thanks to Theorem 4.3 in Mao et al. (2019), for $n \geqslant 3$, we deduce that the distribution of S_1 belongs to the set of possible distributions for the sum of n random variables uniformly distributed on the interval $[0,1]$.

In the second case: $d < c(q)$, we define a random variable S_2 having two mass points,

$$S_2 = \begin{cases} \frac{n}{2} - \sqrt{\frac{dn(n-1)+n}{12}}\sqrt{\frac{1-q}{q}} & \text{with probability } q, \\ \frac{n}{2} + \sqrt{\frac{dn(n-1)+n}{12}}\sqrt{\frac{q}{1-q}} & \text{with probability } 1-q. \end{cases} \tag{3.3.23}$$

The quantile function of S_2 up-crosses the quantile function of S_1, which implies $S_2 \leqslant_{cx} S_1 \leqslant_{cx} U[0,n]$. Hence, for $n \geqslant 3$, the distribution of S_2 is attainable by the sum of n uniformly distributed random variables, for the same argument explained above. □

It would be of great interest to extend the result stated in Theorem 3.3.6 to a more general set of marginal distributions and to also include the case $n = 2$. In this regard, we point out that our proof builds on a result of Mao et al. (2019) and these authors clarified that it seems difficult to extend their results to other than standard uniform distributions and to the case $n = 2$.

3.4 Final remarks

The recent literature related to risk aggregation problems under partial dependence uncertainty has focused on identifying the sources of dependence information that can be easily inferred from the available data but that are also able to reduce risks bounds. The contributions of our analysis to this complex question are the following. For the sum of two risks, we show that VaR and TVaR worst-case values are typically not affected by additional dependence information when the latter is described by Spearman's rho, Kendall's tau or Pearson correlation. By contrast, we show that such dependence information may impact the VaR lower bound value. For the sum of three or more risks, we derive explicit RVaR bounds in the presence of an average correlation constraint and we show that theses bounds are best-possible for the sum of $n \geqslant 3$ risks having standard uniform distributions. We give evidence that when the RVaR is assessed at high confidence levels, an inequality constraint on average correlation can hardly improve the dependence uncertainty spread between the worst- and the best-case scenarios. The VaR and TVaR bounds are derived as limiting cases. Overall, our results make clear that the sources of dependence information we consider here, despite being quite popular in the banking and insurance industry, do not guarantee a reduction of the worst-case values of tail risk measures.

3.5 Appendix

3.5.1 Proof of Lemma 3.2.2

Proof. Let $q \in (0,1)$. First, we prove that if $C \in \mathscr{C}(q)$, then the inequality in (3.2.15) holds. Note that for all copulas $C \in \mathscr{C}(q)$ the following observations hold:

- if $(u,v) \in [q,1]^2$, then $C(u,v) = C^q_{min}(u,v) = C^q_{max}(u,v) = \max(q, u+v-1)$.
- $C(q,q) = q$
- $\forall (u,v) \in [0,q] \times [1,q]$, $C(u,v) = u$
- $\forall (u,v) \in [q,1] \times [0,q]$, $C(u,v) = v$

From the observations above, the inequalities in (3.2.15) need to be proven only in $[0,q[^2$. We start by showing that in $[0,q[^2$ it is not possible to find a copula $C \in \mathscr{C}(q)$ that is pointwise lower than C^q_{min}. First, consider $(u,v) \in [0,q[^2$ such that $u+v-q \leqslant 0$. Since $C^q_{min}(u,v) = 0$ in such point, a copula C cannot be strictly lower than C^q_{min}. Second, consider (u,v) such that $u+v-q > 0$. By definition, any copula must satisfy the following two-increasing property (Nelsen (2010))

$$C(q,q) + C(u,v) - C(u,q) - C(q,v) \geqslant 0, \tag{3.5.1}$$

but for the point (u,v) we are considering and $C \in \mathscr{C}(q)$, the inequality (3.5.1) can be expressed as

$$q + C(u,v) - u - v \geqslant 0. \tag{3.5.2}$$

If $C(u,v) < C^q_{min}(u,v)$, then

$$q + C(u,v) - u - v < q + C^q_{min}(u,v) - u - v = 0, \tag{3.5.3}$$

which violates the assumption that C is a copula. The first inequality in (3.2.15) is completely proven.

As for the second inequality, observe that in $[0,q[^2$, C^q_{max} coincides with the Fréchet-Hoeffding upper bound. Hence, in $[0,q[^2$, C^q_{max} is pointwise higher or equal than any other copula.

Note the implication $C^q_{min}(u,v) \leqslant C(u,v) \leqslant C^q_{max}(u,v) \implies C \in \mathscr{C}(q)$ simply follows from the fact that C^q_{min} and C^q_{max} coincide on $[q,1]^2$, and thus also such C must exhibit an anti-monotonic dependence structure on $[q,1]^2$. □

3.5.2 Proof of Theorem 3.2.3

Proof. First note that thanks to Lemma 3.2.2, the interval $[\delta_{min}, \delta_{max}]$ is well defined ($\delta_{min} \leqslant \delta_{max}$ always holds). To prove the statement, let C_α be a mixture of the two copulas C^q_{min} and C^q_{max}, i.e., $C_\alpha = (1-\alpha)C^q_{min} + \alpha C^q_{max}$ with $\alpha \in [0,1]$. Since the two copulas C^q_{min} and C^q_{max} coincide on $[q,1]^2$, it is straightforward to check that C_α belongs to $\mathscr{C}(q)$. By Assumption 3.2.1, δ is such that the function $\alpha \mapsto \delta(C_\alpha, F_1, F_2)$ is a continuous function of α. Thus, we have $\delta(C_0, F_1, F_2) = \delta_{min}$

and $\delta(C_1, F_1, F_2) = \delta_{max}$. Thanks to the continuity of $\delta(C_\alpha, F_1, F_2)$ w.r.t. α we can use the intermediate value theorem and conclude that for all $\overline{\delta} \in [\delta_{min}, \delta_{max}]$ there exists $\overline{\alpha} \in [0, 1]$ such that and $\delta(C_{\overline{\alpha}}, F_1, F_2) = \overline{\delta}$.

Using the results we just proved, it is clear that for each $d \in [\delta_{min}, \delta_{max}]$ there exists a copula C in $\mathscr{C}(q)$ that satisfies the constraint $\delta(C, F_1, F_2) = d$. Thanks to Proposition 3.2.1, we know that all copulas in $\mathscr{C}(q)$ attain the unconstrained VaR$^+$ upper bound, and thus $\overline{\text{VaR}}_q^d = \overline{\text{VaR}}_q$. □

3.5.3 Proof of Proposition 3.2.4

Proof. Assume the hypotheses in Proposition 3.2.4 are satisfied. Fix $q \in (0, 1)$ and $d \in (\delta_{min}(q), \delta_{max}(q))$. First, let us denote with $\widetilde{\text{VaR}}_q^d$ the solution to the problem we are studying, namely

$$\begin{aligned} \widetilde{\text{VaR}}_q^d := \sup \quad & \text{VaR}_q(X_1 + X_2) \\ \text{subject to} \quad & X_j \sim F_j, \ j = 1, 2 \\ & \delta(X_1, X_2) = d \end{aligned} \tag{3.5.4}$$

Observe that for given $q \in (0, 1)$ and $q^* \in (0, q)$ it holds that

$$\overline{\text{VaR}}_{q^*}^d \leqslant \widetilde{\text{VaR}}_q^d \leqslant \overline{\text{VaR}}_q^d. \tag{3.5.5}$$

where

$$\begin{aligned} \overline{\text{VaR}}_{q^*}^d := \sup \quad & \text{VaR}_{q^*}^+(X_1 + X_2) \\ \text{subject to} \quad & X_j \sim F_j, \ j = 1, 2 \\ & \delta(X_1, X_2) = d \end{aligned} \tag{3.5.6}$$

Since $d \in (\delta_{min}(q), \delta_{max}(q))$, we can use Theorem 3.2.3 to conclude $\overline{\text{VaR}}_q^d = \overline{\text{VaR}}_q$. Furthermore, since $d \in (\delta_{min}(q), \delta_{max}(q))$ and since the mappings $\delta_{min} : q \mapsto \delta(C^q_{min}, F_1, F_2)$ and $\delta_{max} : q \mapsto \delta(C^q_{max}, F_1, F_2)$ are continuous, it is clear that there exists $q^d \in (0, q)$ such that $d \in (\delta_{min}(q^*), \delta_{max}(q^*))$, for all $q^* \in (q^d, q)$. Using again Theorem 3.2.3, for all $q^* \in (q^d, q)$ we obtain $\overline{\text{VaR}}_{q^*}^d = \overline{\text{VaR}}_{q^*}$.

These last observations together with the inequalities in (3.5.5) imply that there exists $q^d \in (0, q)$ such that for all $q^* \in (q^d, q)$ the following inequalities hold:

$$\overline{\text{VaR}}_{q^*} \leqslant \widetilde{\text{VaR}}_q^d \leqslant \overline{\text{VaR}}_q. \tag{3.5.7}$$

Thus, it is clear that

$$\lim_{q^* \nearrow q} \overline{\text{VaR}}_{q^*} \leqslant \widetilde{\text{VaR}}_q^d \leqslant \overline{\text{VaR}}_q. \tag{3.5.8}$$

where $\lim_{q^* \nearrow q} \overline{\text{VaR}}_{q^*}$ denotes the right limit of $\overline{\text{VaR}}_{q^*}$ as q^* approached q from below. From Lemma 4.4 in Bernard et al. (2013), the mapping $q \mapsto \overline{\text{VaR}}_q$ is continuous on $(0, 1)$ if $F_1^{-1}(q)$ and $F_2^{-1}(q)$ are continuous for $q \in (0, 1)$. Hence, we get $\lim_{q^* \nearrow q} \overline{\text{VaR}}_{q^*} = \overline{\text{VaR}}_q$, which completes the proof.

□

3.5.4 Proof of Lemma 3.2.5

Proof. To prove (3.2.19) and (3.2.20) it is useful to split the hypercube $[0,1]^2$ into four subsets, namely

$$D_1 = [q,1]^2,\ D_2 = [0,q]^2,\ D_3 = [0,q] \times [q,1] \text{ and } D_4 = [1,q] \times [0,q].$$

Using these subsets, for any function f integrable on $[0,1]^2$, we can write

$$\iint_{[0,1]^2} f(u,v)dudv = \sum_{i=1}^{4} \iint_{Di} f(u,v)dudv$$

The technique adopted here is to integrate separately on each D_i, and then to sum up the results. For example, to prove ρ_{max} in(3.2.19), for any fixed $q \in [0,1]$, we want to express ρ_{max} as a function of q. To do so, we calculate the following integral

$$\rho_{max} = 12 \iint_{[0,1]^2} C^q_{max}(u,v)dudv - 3. \tag{3.5.9}$$

The details are available upon request and we only report the values for the four integrals over D_1, D_2, D_3 and D_4. We find that

$$\iint_{D_1} C^q_{max}(u,v)dudv = \iint_{[q,1]^2} q + \max(u+v-1-q,0)dudv = q(1-q)^2 - \frac{1}{6}(q-1)^3$$

$$\iint_{D_2} C^q_{max}(u,v)dudv = \int_0^q \int_0^q \min(u,v)dudv = \int_0^q qu - \frac{u^2}{2}du = \frac{q^3}{3}.$$

$$\iint_{D_4} C^q_{max}(u,v)dudv = \iint_{D_3} C^q_{max}(u,v)dudv = \int_0^q \int_q^1 \min(u,v)dudv = \frac{q^2}{2}(1-q)$$

After summing up the results of integrals on D_1, D_2, D_3 and D_4 and simplifying we obtain the result in (3.2.19). The calculations to obtain ρ_{min}, τ_{max} and τ_{min} are very similar, hence we do not report them here.

□

3.5.5 Proof of Proposition 3.2.11

Proof. The fact $C \in \mathscr{C}^L(q) \implies \mathrm{VaR}_q(X_1 + X_2) = \underline{\mathrm{VaR}}_q$ follows from Lemma 3.2.8. We shall prove now the opposite implication. Since $X_i \sim \mathcal{U}(0,1)$ for $i = 1,2$, we have $\underline{\mathrm{VaR}}_q = q$. It is well known that given the marginals distributions the LTVaR is minimal for comonotonic risks. Thus, with uniform marginals we have has the following lower bound,

$$\mathrm{LTVaR}_q(S) = \frac{1}{q}\int_0^q \mathrm{VaR}_\gamma(S)d\gamma \geqslant \mathrm{LTVaR}_q(S^c) = q. \tag{3.5.10}$$

Assume now that $\text{VaR}_q(S) = q$. Since the VaR in non-decreasing w.r.t to the probability level, we have $\text{VaR}_\gamma(S) \leqslant q$, for all $\gamma \in (0, q)$. Note that if we assume that there exists $\gamma^* \in (0, q)$ such that $\text{VaR}_{\gamma^*}(S) < q$ the corresponding LTVaR would be strictly lower than q and this would lead to a contradiction of (3.5.10). Thus, it is clear that if a quantile function of $S = X_1 + X_2$ satisfies $\text{VaR}_q(S) = q$, then it must satisfy $\text{VaR}_\gamma(S) = q$, for all $\gamma \in (0, q)$. Observe now that if $\text{VaR}_q(S) = q$, then $\text{LTVaR}_q(S) = q$ and thus $\text{TVaR}_q(S) = 1 + q = \text{TVaR}_q(X_1) + \text{TVaR}_q(X_2)$, which is the maximum TVaR attainable with these marginals. From Lemma 3.2.13, the maximum $\text{TVaR}_q(S)$ is achieved if and only if the copula C is q-concentrated. Finally, if C is q-concentrated and makes the quantile function of $S = X_1 + X_2$ constant and equal to q on $(0, q)$, then C must belong to $\mathscr{C}^L(q)$. □

3.5.6 Proof of Lemma 3.2.15

Proof. We recall that C^q_{min} and C^q_{max}, defined in (3.2.11) and (3.2.12), both belong to $\mathscr{C}(q)$ and from Lemma 3.2.2 these copulas are the pointwise lowest and highest copulas in $\mathscr{C}(q)$, respectively. Since $C^q_{min}(q, q) = q$ and $C^q_{max}(q, q) = q$, we conclude that $\mathscr{C}(q) \subseteq \mathscr{C}_{TVaR}(q)$. Observe, for example, that for each $q \in (0, 1)$ the comonotonic copula is in $\mathscr{C}_{TVaR}(q)$ but not in $\mathscr{C}(q)$, and therefore $\mathscr{C}(q) \subset \mathscr{C}_{TVaR}(q)$.

$\mathscr{C}_{TVaR}(q)$ is obviously convex: let C_1 and C_2 be two copulas belonging to $\mathscr{C}_{TVaR}(q)$. Then, for each $\alpha \in [0, 1]$, the copula $C_\alpha = \alpha C_1 + (1-\alpha) C_2$ satisfies $C_\alpha(q, q) = \alpha C_1(q, q) + (1-\alpha) C_2(q, q) = q$ and therefore belongs to $\mathscr{C}_{TVaR}(q)$. □

3.5.7 Proof of Theorem 3.2.16

The following lemma will be useful in the proof of Theorem 3.2.16.

Lemma 3.5.1. *Let* $q \in (0, 1)$, *and* C *be a copula in set* $\mathscr{C}_{TVaR}(q)$. *Then,*

$$C(u, v) \geqslant C^q_{min}(u, v),\ \forall (u, v) \in [0, 1]^2. \tag{3.5.11}$$

Proof. To prove that C^q_{min} is the pointwise lower bound copula in $\mathscr{C}_{TVaR}(q)$ it is convenient to split the square $[0, 1]^2$ into the following subsets:

- For $(u, v) \in [0, q[^2$, the proof is the same as the beginning of the proof of Lemma 3.2.2, where we proved that C^q_{min} is the pointwise lower bound copula in $\mathscr{C}(q)$.

- For $(u, v) \in [0, q[\times]q, 1]$ and $(u, v) \in]q, 1]\times]0, q]$ all the copulas in $\mathscr{C}_{TVaR}(q)$ assume the same values, see Remark 3.2.2.

- For $(u, v) \in [q, 1]^2$ we need to further split this region as follows.

  – If $(u, v) \in [q, 1]^2$ and $u + v \leqslant 1 + q$, then by equation (3.2.46), $C \in \mathscr{C}_{TVaR}(q) \iff C(q, q) = q$, and for the points we are considering it must hold that $C(u, v) \geqslant C(q, q) = q = C^q_{min}(u, v)$.

– If $(u,v) \in [q,1]^2$ and $u+v > 1+q$, then $C^q_{min}(u,v) = u+v-1 = C^a(u,v)$, and C^a (the anti-monotonic copula) is the pointwise lowest copula in general, which ends the proof.

□

The proof of Theorem 3.2.16 now simply follows from Lemma 3.2.14, Lemma 3.2.15, Lemma 3.5.1 and an argument similar to the proof of Theorem 3.2.3.

3.5.8 Proof of Proposition 3.3.4

Proof. The proof is organized as follows. First, we are going to derive RVaR bounds for given marginals in the case when a constraint on the variance sum is assumed. Second, we use the one-to-one relationship between the variance sum and the average correlation to express the bounds in terms of average correlation. For n given marginal distributions $F_1, F_2, ..., F_n$ and $s^2 \geqslant 0$, denote by $\mathcal{F}_{s^2}$ the set of possible distributions for the sum S that satisfy $var(S) \leqslant s^2$, that is

$$\mathcal{F}_{s^2} = \mathcal{F}_{s^2}(F_1, ..., F_n) = \left\{ \text{cdf of } S = X_1 + ... + X_n : \ X_i \sim F_i \text{ for } i = 1, .., n, \ var(S) \leqslant s^2 \right\}$$

Let now $\mathcal{S}_{s^2}$ be the set of random variables having a distribution in $\mathcal{F}_{s^2}$,

$$\mathcal{S}_{s^2} = \mathcal{S}_{s^2}(F_1, ..., F_n) = \left\{ X : F_X \in \mathcal{F}_{s^2} \right\}.$$

Let us denote with $V_{\leqslant}(\mu, s^2)$ the set of random variables having a mean equal to μ and a variance lower or equal than s^2:

$$V_{\leqslant}(\mu, s^2) = \left\{ X : \mathbb{E}(X) = \mu, \ var(X) \leqslant s^2, \right\}.$$

The inequality $V(\mu, s^2) \subset V_{\leqslant}(\mu, s^2)$ is trivial.

Let $\mu = \sum_{i=1}^n \mathbb{E}(X_i)$, it is clear that $\mathcal{S}_{s^2} \subseteq V_{\leqslant}(\mu, s^2)$. Thanks to Proposition 3.3.3, we deduce that for any $0 < q < q' < 1$, the maximum and minimum RVaRs attainable in $V_{\leqslant}(\mu, s^2)$ are achieved by two distributions that belong also to $V(\mu, s^2)$. Consequently,

$$\sup_{X \in \mathcal{S}_{s^2}} \text{RVaR}_{q,q'}(X) \leqslant \max_{X \in V_{\leqslant}(\mu, s^2)} \text{RVaR}_{q,q'}(X) = \max_{X \in V(\mu, s^2)} \text{RVaR}_{q,q'}(X) = \mu + s\sqrt{\frac{q}{1-q}}, \quad (3.5.12)$$

$$\inf_{X \in \mathcal{S}_{s^2}} \text{RVaR}_{q,q'}(X) \geqslant \min_{X \in V_{\leqslant}(\mu, s^2)} \text{RVaR}_{q,q'}(X) = \min_{X \in V(\mu, s^2)} \text{RVaR}_{q,q'}(X) = \mu - s\sqrt{\frac{1-q'}{q'}}. \quad (3.5.13)$$

Combining the inequalities in (3.5.12) and (3.5.13) with Proposition 3.3.2, we find that

$$\sup_{X \in \mathcal{S}_{s^2}} \text{RVaR}_{q,q'}(X) \leqslant \min \left(\mu + s\sqrt{\frac{q}{1-q}}, B(q) \right) \quad (3.5.14)$$

$$\inf_{X \in \mathcal{S}_{s^2}} \text{RVaR}_{q,q'}(X) \geqslant \max\left(\mu - s\sqrt{\frac{1-q'}{q'}}, A(q')\right). \tag{3.5.15}$$

We will show now that if $s^2 \geqslant q'\left(A(q') - \mu\right)^2 + (1-q')\left(B(q') - \mu\right)^2$, then $l(q') = A(q')$ (the proof of $s^2 \geqslant q\left(A(q) - \mu\right)^2 + (1-q)\left(B(q) - \mu\right)^2 \implies b(q) = B(q)$ is very similar).

Let $X^* \sim D_{q'}(A(q'), B(q'))$, then

$$\begin{aligned}
&\mathbb{E}(X^*) = q'A(q') + (1-q')B(q') = \mu,\\
&var(X^*) = q'\left(A(q') - \mu\right)^2 + (1-q')\left(B(q') - \mu\right)^2,\\
&\text{RVaR}_{q,q'}(X^*) = A(q') = \sum_{i=1}^{n} \text{LTVaR}_{q'}(X_i).
\end{aligned}$$

Therefore, if the variance constraint s^2 is higher than the threshold $q'\left(A(q') - \mu\right)^2 + (1-q')\left(B(q') - \mu\right)^2$, one has $X^* \in V_{\leqslant}(\mu, s^2)$ and

$$A(q') = \text{RVaR}_{q,q'}(X^*) \geqslant \min_{X \in V_{\leqslant}(\mu, s^2)} \text{RVaR}_{q,q'}(X) = \mu - s\sqrt{\frac{1-q'}{q'}}.$$

To complete this part of the proof, one only needs to use the one-to-one relationship between the variance of the sum and the average correlation expressed in equation (3.3.2). The bounds in (3.5.14) and (3.5.15) and the condition $s^2 \geqslant q'\left(A(q') - \mu\right)^2 + (1-q')\left(B(q') - \mu\right)^2$ are expressed in terms of the standard deviation of the sum s. By rewriting the bounds in terms of the average correlation constraint d, we obtain the bounds $l(q')$ and $u(q)$. Using the same approach, one finds the values $c(q)$ and $c(q')$.

We shall now prove that the equality $\overline{\text{RVaR}}^{d}_{q,q'} = u(q)$ holds if there exists a copula such that acorr$(\mathbf{X}) \leqslant d$ and S is mixing in the upper q-part of its distribution with mixing constant $u(q)$. If the hypothesis is satisfied, the mixability of S in its q-upper part with mixing constant $u(q)$ implies that

$$\text{VaR}_p(S) = u(q), \forall p \in (q, 1).$$

Therefore,

$$\text{RVaR}_{q,q'}(S) = \frac{1}{q'-q}\int_q^{q'} \text{VaR}_p(S)dp = \frac{1}{q'-q}\int_q^{q'} u(q)dp = u(q).$$

The proof for the lower bound is very similar and we omit it. □

Chapter 4

Robust assessment of life insurance products

4.1 Introduction

The actuarial literature related to model risk assessment has experienced rapid growth in recent years. The idea behind this field of research is that any actuarial evaluation is prone to error in that the underlying loss distribution is typically only partially known. As illustrated in Chapter 3, there is extensive literature on finding bounds on a risk measure. Clearly, a main quantity of actuarial interest that can be affected by model misspecification is the premium to be paid for an insurance contract. The net premium of a life insurance contract depends essentially on two ingredients: the residual lifetime distribution function (actuarial risk) and the discount curve (financial risk). The goal of the present Chapter is to develop a framework that can help the insurer to deal with model risk that arises from a misspecified residual lifetime distribution function. Recent studies have shown that in a low-interest rate environment longevity risk becomes the major risk-driver of the life insurance business; see Haberman et al. (2011), Antolin (2007), Rabitti and Borgonovo (2020), and some of the references therein.

Several competing longevity models (or mortality forecast models) have been proposed in the literature. For an overview, see for instance Pitacco et al. (2009). Nonetheless, as is the case for any statistical procedure's output, a projected life table can never be completely trusted. In the life insurance literature, the study of the effect of model misspecification on the price of a contract is usually conducted using a parametric approach: it is assumed that the residual lifetime distribution belongs to one or more given families of probability distributions with uncertain parameters. Olivieri (2001), Olivieri and Pitacco (2011), and Cairns (2000) offer discussions of this approach and provide some case studies. We propose to tackle this problem using a non-parametric approach. The estimated probability distribution that arises from the life-table will be considered as a reference distribution function. One can wonder about the range of the price of an insurance contract when we consider all the distribution functions that are somehow close to the reference distribution function, but not necessarily coming from the same parametric family. In order to formalize the notion of closeness between probability distributions, we will make use of the L^2

metric between two distribution functions (in a similar spirit as in Pichler (2014)).

There is growing interest in the impact of model risk on insurance pricing. Escobar and Pflug (2020) study the sensitivity and continuity of the distortion premium with respect to the Wasserstein distance. In the context of reinsurance pricing of large catastrophic events, Dietz and Walker (2019) and Dietz and Niehörster (2021) offer analyses that link theory on model risk developed in actuarial science with recent results on decision making under ambiguity derived in the economic literature. In particular, these authors develop a theory that formally motivates premium loadings due to ambiguity. Finally, in the context of life insurance pricing, Pichler (2014) derives upper bounds on a coherent risk measure of a life insurance contract when the lifetime distribution function is known to sit in a Wasserstein ball built around a reference distribution function. Since this latter paper studies a problem closely related to the one we consider here, in Section 4.6 we compare our results to those obtained in this paper.

Section 4.2 is devoted to the mathematical formulation of the problem at hand and to the modelling of ambiguity using the L^2 metric. We start with a review of some basic life insurance concepts, such as the equivalence principle, and then move to the definition of the premium bounds, whose computation and properties are the main goal of the present analysis.

Section 4.3 contains most of our results regarding the properties and computation of the premium bounds. We show that best- and worst-case values for the premium can be explicitly computed using a convex Quadratically Constrained Linear Program (QCLP). By studying the properties of this linear program we prove that premium bounds enjoy desirable properties, such as continuity with respect to the L^2 distance constraint. In some cases, we are able to derive the analytic expressions of the feasible probability distributions that maximize and minimize the premium.

Section 4.4 provides several numerical examples that illustrate how our results can be used to obtain a robust assessment of an annuity contract. Specifically, the examples we provide show how our results can be useful in studying the net premium bounds, the probability distributions attaining the bounds, the relationship between model risk and interest rates, and the robust expected discounted utility maximization.

In Section 4.5, we show that the framework we developed is flexible enough to handle additional constraints other than the L^2 distance constraint. Specifically, we show how the framework makes it possible to deal with the important case in which the feasible distribution functions are assumed to be unimodal.

Section 4.6 offers a detailed comparison of our results with those obtained in Pichler (2014). In particular, we highlight the main advantages of describing uncertainty via the L^2 metric as compared to the Wasserstein distance used in Pichler (2014).

4.2 Problem formulation

Let T_x be the residual lifetime of an individual at age x and let be K_x be the *curtate remaining lifetime*, defined as the integer part of T_x, $K_x = \lfloor T_x \rfloor$.

It is common to assume that the individual residual lifetime cannot exceed a certain value ω. Hence, $\mathcal{H} = \{0, 1, 2, ..., \omega - x\}$ is the set containing all possible outcomes of the random variable

K_x. We denote by $\boldsymbol{q} = \left({}_{0|1}q_x, {}_{1|1}q_x, \ldots, {}_{\omega-x|1}q_x\right)^t$ the column vector describing the probability distribution of K_x, namely ${}_{h|1}q_x = P(K_x = h)$, for $h \in \mathcal{H}$. The quantities ${}_{h|1}q_x$ are usually estimated and reported in a life table. Finally, let $g : \mathcal{H} \to \mathbb{R}$ be the *payoff function* of an insurance contract. To be more specific, the function g associates to each $h \in \mathcal{H}$ the present value of all future payments that the insurer must pay to the beneficiaries of the contract if $K_x = h$.

4.2.1 Equivalence principle in life insurance

Several pricing principles have been proposed in the actuarial literature. Notable examples are the mean-variance pricing principle, the distortion premium principle, and the utility equivalence principle. The interested reader can find an overview in Young (2014). The present analysis focuses mainly on the equivalence principle. Since we will consider contracts having fixed benefits, i.e., contracts in which the amount of benefits is stated at policy issue, the equivalence principle can be defined as follows.

Definition 4.2.1 (Equivalence principle with fixed benefits)**.** *Given the distribution of K_x and a contract payoff function $g(\cdot)$, the equivalence principle sets the net premium π of the insurance contract equal to*

$$\pi = \mathbb{E}(g(K_x)).$$

For further details on the equivalence principle and its application in insurance pricing, we refer to Chapter 4 of Olivieri and Pitacco (2011) and Chapter 6 of Dickson et al. (2009). In this framework, the net premium of an insurance contract is set equal to the expected present value of the benefits, also called actuarial value in the life insurance literature. The equivalence pricing principle is conceptually simple and well established in the industry practice. Note that π depends solely on the payoff function and the probability distribution of K_x, that is π is law-invariant. We report here some concrete examples of payoff functions $g(\cdot)$ [1].

- Pure endowment: the amount S will be paid to the beneficiaries after m years if the insured is alive at that time. Since ${}_m p_x = P(K_x \geqslant m) = \sum_{h=m}^{\omega-x} {}_{h|1}q_x$, the net premium

$$\pi = S(1+r)^{-m}\ {}_m p_x = S(1+r)^{-m} \sum_{h=m}^{\omega-x} {}_{h|1}q_x$$

  can be written as $\pi = \mathbb{E}\left(g\left(K_x\right)\right)$ with

$$g(h) = \begin{cases} 0 & \text{, if } h < m \\ S(1+r)^{-m} & \text{, if } h \geqslant m. \end{cases}$$

[1] For notational convenience, the examples of payoff functions reported in the sequel are written considering the case of a constant annual discounting interest rate r, corresponding to a flat yield curve. This leads to a discount function in the form $v(h) = (1+r)^{-h}$, for all $h \in \mathcal{H}$. Nonetheless, all our results remain valid for any possible shape of the yield curve adopted to compute the present value of future cash flows.

- Endowment insurance: the amount C_h will be paid at time h to the beneficiaries if the insured dies between time $h-1$ and h, for $h \leqslant m$, where m denotes the policy term. The amount S will be paid after m years to the beneficiaries if the insured is alive at that time. The net premium of this contract

$$\pi = \sum_{h=1}^{m} C_h (1+r)^{-h} {}_{h-1|1}q_x + S(1+r)^{-m} \sum_{h=m}^{\omega-x} {}_{h|1}q_x,$$

can be written as $\pi = \mathbb{E}\left(g\left(K_x\right)\right)$ with

$$g(h) = \begin{cases} C_{h+1}(1+r)^{-(h+1)} & \text{, if } h < m \\ S(1+r)^{-m} & \text{, if } h \geqslant m. \end{cases}$$

- Life annuities: the amount b_h is paid to the beneficiaries at each time $h = 1, 2, ..., \omega - x$ as long as the insured is alive. Note that there exist many payment structures for life annuities, such as constant ($b_h = b$ for $h = 1, 2, ..., \omega - x$), arithmetically increasing ($b_h = b_1(1 + (h-1)\alpha)$), or geometrically increasing ($b_h = b_1(1+\alpha)^{h-1}$). In any case, we have

$$\pi = \sum_{h=1}^{\omega-x} \sum_{j=1}^{h} b_j (1+r)^{-j} {}_{h|1}q_x.$$

This can be written as $\pi = \mathbb{E}\left(g\left(K_x\right)\right)$ with

$$g(h) = \begin{cases} 0 & \text{, if } h = 0 \\ \sum_{j=1}^{h} b_j (1+r)^{-j} & \text{, if } h = 1, 2, ..., \omega - x. \end{cases}$$

Observe that, in the setting we consider, the net premium of a life insurance contract depends solely on the distribution of K_x. However, longevity trends have proven to be quite unpredictable, i.e., K_x is subject to distributional uncertainty, and several competing methodologies have been proposed to estimate its probability distribution.

4.2.2 Premium bounds

Any law invariant pricing principle is affected by possible errors made in the evaluation of the probability distribution of interest. Following Escobar and Pflug (2020), we denote with $\mathfrak{F}$ a general ambiguity set, i.e., a collection of probability distributions that are compatible with the available information.

Given a payoff function g and an ambiguity set $\mathfrak{F}$, we will study the upper- and lower-bound for the net premium computed according to the equivalence principle. These bounds will be denoted as $\overline{\pi}_g^{\mathfrak{F}}$ and $\underline{\pi}_g^{\mathfrak{F}}$ and are defined as

$$\overline{\pi}_g^{\mathfrak{F}} = \sup\{\mathbb{E}(g(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}. \tag{4.2.1}$$

$$\underline{\pi}_g^{\mathfrak{F}} = \inf\{\mathbb{E}(g(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}. \tag{4.2.2}$$

The interval $[\underline{\pi}_g^{\mathfrak{F}}, \overline{\pi}_g^{\mathfrak{F}}]$ will be used to measure the impact of model misspecification (ambiguity) on the price of the insurance contract identified by the payoff function g. As insurance companies tend to be conservative in their evaluations, it seems natural that in the presence of the distributional uncertainty described by $\mathfrak{F}$, the upper-bound $\overline{\pi}_g^{\mathfrak{F}}$ could be instrumental in setting the commercial premium. In this regard, $\overline{\pi}_g^{\mathfrak{F}}$ satisfies the following basic properties that are considered desirable for a premium principle in the life insurance context, and this holds true regardless of the specific form of the ambiguity set $\mathfrak{F}$.

Proposition 4.2.2. *For any ambiguity set $\mathfrak{F}$, given two payoff functions g_1 and g_2, the following properties hold:*

1. *Monotonicity: if $g_1(\cdot) \leqslant g_2(\cdot)$, then $\overline{\pi}_{g_1}^{\mathfrak{F}} \leqslant \overline{\pi}_{g_2}^{\mathfrak{F}}$.*

2. *Translational invariance: if $g_1(\cdot) = c + g_2(\cdot)$ with $c \in \mathbb{R}$, then $\overline{\pi}_{g_1}^{\mathfrak{F}} = c + \overline{\pi}_{g_2}^{\mathfrak{F}}$.*

3. *Positive homogeneity: if $g_1(\cdot) = \lambda g_2(\cdot)$ with $\lambda \geqslant 0$, then $\overline{\pi}_{g_1}^{\mathfrak{F}} = \lambda \overline{\pi}_{g_2}^{\mathfrak{F}}$.*

4. *Convexity: if $g(\cdot) = \alpha g_1(\cdot) + (1-\alpha) g_2(\cdot)$ with $\alpha \in [0,1]$, then $\overline{\pi}_g^{\mathfrak{F}} \leqslant \alpha \overline{\pi}_{g_1}^{\mathfrak{F}} + (1-\alpha)\overline{\pi}_{g_2}^{\mathfrak{F}}$.*

The proof is straightforward and is thus relegated to Appendix 4.8.1. Proposition 4.2.2 partially extends the results of Theorem 16 in Pichler (2014). On the one hand, Proposition 4.2.2 is more general than Theorem 16 in Pichler (2014) in that our result is valid for any arbitrary ambiguity set, whereas Pichler (2014) considers a more specific case of ambiguity sets that are Wasserstein balls. On the other hand, we consider only the expectation of $g(K_x)$, whereas Theorem 16 of Pichler (2014) considers a general concave distortion risk measure and thus is indeed more general from this point of view.

Hereafter, given $\tilde{F} \in \mathfrak{F}$ and a payoff function $g(\cdot)$, we denote with $\pi_{\tilde{F}}$ the net premium computed with the distribution function $\tilde{F}$, i.e., $\pi_{\tilde{F}} = \mathbb{E}(g(K_x))$ with $K_x \sim \tilde{F}$.

4.2.3 Modelling distributional uncertainty

In order to compute the premium bounds (4.2.1) and (4.2.2), we need to provide a precise definition of the ambiguity set $\mathfrak{F}$. In this section, we propose to describe ambiguity (uncertainty) using the L^2 distance between distribution functions. For a general introduction to the L^p metric and its financial applications see Rachev et al. (2008). For applications in actuarial science, see for instance López-Díaz et al. (2012) and Yang et al. (2014). In particular, we assume that the distribution of K_x belongs to a subset of an L^2-ball built around a target distribution denoted with F. Interpretatively, F represents the current best estimate for the df of K_x, although we agree it is subject to model misspecification. We denote as $\boldsymbol{f} = (f_0, f_1, ..., f_{\omega-x})^t$ its reference probability distribution, i.e. $f_h = P(K_x = h)$ for $h = 0, 1, 2, ..., \omega - x$ under the model $K_x \sim F$. The L^2-ball of radius $\sqrt{\varepsilon}$ and center F is defined as

$$\mathcal{M}_\varepsilon(F) = \left\{ \tilde{F} \;\middle|\; d(\tilde{F}, F) \leqslant \sqrt{\varepsilon} \right\}, \tag{4.2.3}$$

in which $d(\tilde{F}, F)$ is the L^2 distance between two distribution functions $\tilde{F}$ and F, i.e.,

$$d(\tilde{F}, F) = \sqrt{\int_{-\infty}^{+\infty} \left(\tilde{F}(t) - F(t)\right)^2 dt}. \tag{4.2.4}$$

Sometimes we will refer to the L^2-ball as a L^2 ambiguity set. Since K_x is a discrete random variable taking values $\{0, 1, 2, ..., \omega - x\}$, any candidate df of K_x is a step function that can have jumps only at the values $\{0, 1, 2, ..., \omega - x\}$, satisfies $\tilde{F}(t) = 0$ for $t < 0$ and $\tilde{F}(t) = 1$, for $t \geqslant \omega - x$. Thus, the distributional uncertainty regarding K_x can be described by considering all distributions $\tilde{F}$ belonging to $\mathcal{M}_\varepsilon(F)$ that are piecewise constant and with jumps at the points $\{0, 1, 2, ..., \omega - x\}$. Although this set is a subset of $\mathcal{M}_\varepsilon(F)$, in what follows we use the same notation for it.

If $\tilde{F}$ and F are both piecewise constant, then the function $t \longrightarrow \left(\tilde{F}(t) - F(t)\right)^2$ is also a step function with jumps at the points $\{0, 1, 2, ..., \omega - x\}$. Therefore, when we restrict ourself to these distributions, the L^2 distance expression $d(\tilde{F}, F)$ can be expressed as

$$d(\tilde{F}, F) = \sqrt{\sum_{h=0}^{\omega - x} (\tilde{F}_h - F_h)^2}, \tag{4.2.5}$$

where $\tilde{F}_h$ and F_h are the constant values taken by $\tilde{F}$ and F on the interval $[h, h+1)$, for $h = 0, 1, ..., \omega - x$. A distribution function $\tilde{F} \in \mathcal{M}_\varepsilon(F)$ is uniquely determined by its probability distribution $\boldsymbol{q} = \left({}_{0|1}q_x, {}_{1|1}q_x, ..., {}_{\omega - x|1}q_x\right)^t$. Therefore, we sometimes write $d(\boldsymbol{q}, F)$ to denote the L^2 distance between F and the distribution function $\tilde{F}$ such that $\tilde{F}(h) = \sum_{j=0}^{h} {}_{j|1}q_x$. Going forward, we focus on the case in which $\mathfrak{F} = \mathcal{M}_\varepsilon(F)$, as described above. In order to clarify that we are dealing with this ambiguity set, which depends on the parameter ε, we denote the worst- and best-case prices as $\overline{\pi_\varepsilon}$ and $\underline{\pi_\varepsilon}$, respectively. Thus, the problems we aim to study in the remainder of this analysis can be formulated as follows:

$$\begin{aligned} \overline{\pi_\varepsilon} = \max \quad & \mathbb{E}\left(g\left(K_x\right)\right) \\ \text{subject to} \quad & K_x \sim \tilde{F}, \\ & \tilde{F} \in \mathcal{M}_\varepsilon(F). \end{aligned} \tag{4.2.6}$$

$$\begin{aligned} \underline{\pi_\varepsilon} = \min \quad & \mathbb{E}\left(g\left(K_x\right)\right) \\ \text{subject to} \quad & K_x \sim \tilde{F}, \\ & \tilde{F} \in \mathcal{M}_\varepsilon(F). \end{aligned} \tag{4.2.7}$$

and we also study the distributions attaining the bounds in (4.2.6) and (4.2.7).

4.3 Computing premium bounds

In this section we focus on the computation and properties of $\overline{\pi}_\varepsilon$ and $\underline{\pi}_\varepsilon$ in (4.2.6) and (4.2.7). Moreover, we show how to compute the distributions that achieve the bounds. As a by-product of the analysis conducted here, we highlight the convenience of using the L^2 distance to describe distributional uncertainty in this context. As for notation, in what follows, given a vector $\boldsymbol{a} = (a_1, a_2, ..., a_n) \in \mathbb{R}^n$ with $a_j > 0$ for $j = 1, 2, ..., n$, we write $\boldsymbol{a} > 0$.

4.3.1 Problem reformulation

Observe that the L^2 ambiguity set can be equivalently expressed in terms of vectors $({}_{0|1}q_x, ..., {}_{\omega-x|1}q_x)$, i.e., in terms of probability distributions such that $\sum_{j=0}^{h} {}_{j|1}q_x = \tilde{F}_h$. Thus, Problems (4.2.6) and (4.2.7) can be rewritten as

$$\begin{aligned}
\min_{\boldsymbol{q}} \quad & \langle \boldsymbol{y}, \boldsymbol{q} \rangle = \sum_{h=0}^{\omega-x} y_h \, {}_{h|1}q_x \\
\text{subject to} \quad & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 \leqslant \varepsilon, \\
& \sum_{h=0}^{\omega-x} {}_{h|1}q_x = 1, \\
& {}_{h|1}q_x \geqslant 0, h = 0, ..., \omega - x.
\end{aligned} \tag{4.3.1}$$

where the vector $\boldsymbol{y} = (y_0, y_1, ..., y_{\omega-x})$ can be identified according to the payoff function g considered. Specifically, Problem (4.2.6) corresponds to the case in which $y_h = -g(h)$, and Problem (4.2.7) corresponds to the case $y_h = g(h)$. In the language of operational research, Problem (4.3.1) can be classified as a Quadratically Constrained Linear Program, a class of problem for which a general closed form solution is not presently available. The following results state the well-posedness and the convexity of our problems.

Proposition 4.3.1. *Given $\varepsilon > 0$, Problem* (4.3.1) *is well-posed and its feasible region is a non-empty, compact, and convex set.*

The proof is given in Appendix 4.8.2.

Proposition 4.3.2. *Problems* (4.2.6) *and* (4.2.7) *are well-posed. Moreover, for any $c \in [\, \underline{\pi}_\varepsilon, \, \overline{\pi}_\varepsilon \,]$ there exists $\tilde{F} \in \mathcal{M}_\varepsilon(F)$ such that $\pi_{\tilde{F}} = c$.*

The proof follows immediately from Proposition 4.3.1. Since in Problem (4.3.1) we are looking for the minimum of a convex function over a convex set, Problem (4.3.1) can be classified as a convex problem.

Remark 4.3.1. *Numerical solutions to Problem* (4.3.1) *(and therefore to Problems* (4.2.6) *and* (4.2.7)*) and its optimizing distribution can be easily obtained. We remind the reader that for a convex problem, any local minimum is a global minimum and there exist many efficient algorithms to obtain numerical solutions for this class of optimization problem. See Boyd and Vandenberghe (2004).*

As for any procedure based on an optimization problem, it is important that small perturbations of one constraint (in our case, for example, the radius for the L^2-ball) do not significantly affect the solutions. See for example Escobar and Pflug (2020) for a detailed study of continuity of distortion risk measures with respect to the Wasserstein distance. In the following proposition, we ensure the continuity of premium bounds with respect to the L^2 distance constraint.

Proposition 4.3.3 (Continuity of premium bounds)**.** *Given* $\varepsilon > 0$, *let* $\overline{\pi_\varepsilon}$ *and* $\underline{\pi_\varepsilon}$ *be defined as in* (4.2.6) *and* (4.2.7)*. Then, the mappings*

$$\varepsilon \mapsto \overline{\pi_\varepsilon} \quad \textit{and} \quad \varepsilon \mapsto \underline{\pi_\varepsilon} \tag{4.3.2}$$

are concave and convex, respectively, and thus continuous.

The proof of Proposition 4.3.3 is given in Appendix 4.8.3. Let us make the following two observations on the solutions to Problem (4.3.1). First, when ε is "too big", the L^2 distance constraint in Problem (4.3.1) may be redundant and in this case the optimizing distribution function is a degenerate distribution, i.e., a distribution concentrating all probability mass in one point, which is clearly not acceptable for applications in the life insurance context. Hence, in Proposition 4.3.4 we derive a sufficient condition on ε that prevents this situation. Second, we focus on certain cases in which it is possible to derive an explicit analytic expression for the optimizing distribution of Problem (4.3.1), and this is done in Theorem 4.3.6.

Proposition 4.3.4. *Let* $\boldsymbol{f} > 0$*. Then, an optimizing distribution of Problem* (4.3.1) *is a degenerate probability distribution if and only if*

$$\varepsilon \geqslant \min_{h \in \mathcal{H}^{\boldsymbol{y}}_{min}} \sum_{j=0}^{h-1} F_j^2 + \sum_{j=h}^{\omega-x} (F_j - 1)^2, \tag{4.3.3}$$

with $\mathcal{H}^{\boldsymbol{y}}_{min} = \{h \in \mathcal{H} \mid y_h = y_{min}\}$*, where* $y_{min} = \min\{y_0, y_1, ..., y_{\omega-x}\}$.

The proof is given in Appendix 4.8.4. Proposition 4.3.4 provides a sufficient and necessary condition on ε such that the probability distributions attaining the bounds in (4.2.6) and (4.2.7) are not degenerate. We underline that asking that no degenerate distribution belongs to the feasible set $\mathcal{M}_\varepsilon(F)$ is a strictly stronger condition than the one in the statement of Proposition 4.3.4. In what follows, we present a case in which the solution is not degenerate and we can derive an explicit formula. This can be useful for example to check that the numerical solutions to Problem (4.3.1) are accurate in these cases. In order to derive this analytical expression, we need some conditions regarding the reference distribution F (or, equivalently, on the reference probability distribution $\boldsymbol{f}$) and the feasible distributions functions of Problem (4.3.1). However, we do not need any assumption on $\boldsymbol{y}$ or, equivalently, on the payoff function g. First, we need a lemma.

Lemma 4.3.5. *Let* $\boldsymbol{f} > 0$*. Then, there exists* ε *such that any feasible probability distribution of Problem* (4.3.1) *satisfies* $\boldsymbol{q} > 0$, *i.e.,* ${}_{h|1}q_x > 0$ *for all* $h \in \mathcal{H}$*. In particular,* $\boldsymbol{q} > 0$ *holds for any* $\varepsilon \in \left(0,\ \min\left\{ \frac{f_h^2}{2} \,\middle|\, h = 0, 1, ..., \omega - x \right\}\right)$.

The proof is given in Appendix 4.8.5. Lemma 4.3.5 gives us a sufficient condition on ε such that the assumptions in the next theorem are satisfied. In particular, if $\boldsymbol{f} > 0$ (i.e. $f_h > 0$ for $h = 0, 1, ..., \omega - x$), one can simply choose $\varepsilon < \min\left\{ \frac{f_h^2}{2} \,\middle|\, h = 0, 1, ..., \omega - x \right\}$ and the assumptions of Theorem 4.3.6 are satisfied. This theorem implies that if ε is small enough, the distribution attaining the bounds in Problems (4.2.6) and (4.2.7) are unique and admit a closed-form representation.

Theorem 4.3.6. *Let $\boldsymbol{f} > 0$ and $\varepsilon > 0$ such that any feasible probability distribution satisfies $\boldsymbol{q} > 0$. Then, the optimizing distribution of Problem* (4.3.1) *is unique and given by the vector $\boldsymbol{q}^*$ obtained as*

$$\begin{aligned} {}_{0|1}q_x^* &= f_0 + \frac{y_1 - y_0}{2\lambda^*}, \\ {}_{h|1}q_x^* &= f_h + \frac{y_{h-1} - 2y_h + y_{h+1}}{2\lambda^*} \quad \text{for } h = 1, 2, ..., \omega - x - 1, \\ {}_{\omega-x|1}q_x^* &= f_{\omega-x} + \frac{y_{\omega-x-1} - y_{\omega-x}}{2\lambda^*}. \end{aligned} \tag{4.3.4}$$

where $\lambda^* = \sqrt{\frac{\sum_{h=0}^{\omega-x-1}\left(y_{h+1}-y_h\right)^2}{4\varepsilon}}$. Furthermore, $\boldsymbol{q}^*$ satisfies $d(\boldsymbol{q}^*, F) = \sqrt{\varepsilon}$.

The proof is given in Appendix 4.8.6. Theorem 4.3.6 provides an explicit formula for the minimizing distribution of Problem (4.3.1). Furthermore, under the assumptions of Theorem 4.3.6, the distributions attaining the bounds in Problems (4.2.6) and (4.2.7) are unique and have the maximal L^2 distance admissible from the reference probability distribution.

To give an example, consider an endowment insurance in which the amount C_h will be paid at time h to the beneficiaries if the insured dies between time $h-1$ and h, where $h \leqslant m$ and m denotes the policy term. The amount S will be paid at time m to the beneficiaries if the insured is alive at that time. The payoff function of this contract writes as

$$g(h) = \begin{cases} C_{h+1}(1+r)^{-(h+1)} & , \text{ if } h < m \\ S(1+r)^{-m} & , \text{ if } h \geqslant m. \end{cases}$$

Under the assumption of Theorem 4.3.6, we can find the probability distribution solving Problem (4.2.6) by using formula (4.3.4) with $y_h = -g(h)$, for $h \in \mathcal{H}$. The following two remarks complete this section.

Remark 4.3.2. *(How to choose ε?)*

The value of ε reflects the level of ambiguity regarding the curtate residual lifetime distribution. One practical method that can be used to choose a specific value of ε is the following. Consider the case in which the reference distribution F is estimated but there is also a finite set of alternative candidate distributions that are deemed reasonable. In this situation, one can compute the L^2 distance between the reference distribution and each other candidate and set $\sqrt{\varepsilon}$ equal to maximal L^2 distance observed. This procedure ensures that all other distributions deemed reasonable belong to the ambiguity set $\mathcal{M}_\varepsilon(F)$.

Remark 4.3.3. *(Further fields of application.)*

The aim of this remark is to underline that the results obtained so far to compute the bounds of the net premium can also be used to compute the bounds of other functionals of the residual lifetime K_x.

- *Stop-loss premium bounds. So far, we have considered premiums computed using the equivalence principle (Definition 4.2.1). Another well-established premium principle in the actuarial literature is the stop-loss premium. For an introduction and properties of the stop-loss premium and the related stop-loss order, see for example Kaas (1993). If $d \in \mathbb{R}$ and $g(K_x)$ is the random payoff of the contract, then we denote with π_d the stop-loss premium with threshold d, i.e.,*

$$\pi_d = \mathbb{E}\left((g(K_x) - d)_+\right), \tag{4.3.5}$$

  where $(g(K_x) - d)_+ = \max(g(K_x) - d, 0)$. Clearly, π_d can be seen as the expected value of $t(K_x)$ where $t(\cdot) = \max(g(\cdot) - d, 0)$ is a deterministic function of K_x. Thus, the bounds of π_d can be computed and studied using the results obtained in Section 4.3.

- *Robust expected discounted utility. Consider an individual who aims to buy an annuity but has to decide which payment structure is the best. Let $\boldsymbol{b} = (b_0, b_1, ..., b_{\omega-x})$ be the vector that identifies the payment structure of an annuity, i.e., for $h = 1, 2, ..., \omega - x$, b_h is the amount that is paid at policy anniversary h, while b_0 can be set equal to $-\pi$ where π is the premium that the insurance company charges for this annuity. The standard tool adopted in the economic theory of inter-temporal choices to compare consumption plans (such as annuities) is the expected discounted utility. See for example the seminal paper of Yaari (1965) and the stream of literature that followed. Since here we are comparing insurance contracts that offer payments only at policy anniversary h for $h \in \{0, 1, 2, ..., \omega - x\}$, the value of the discounted utility at time h for an annuity with payment structure $\boldsymbol{b} = (b_0, b_1, ..., b_{\omega-x})$ can be written as*

$$D_{\boldsymbol{b}}(h) = \sum_{j=0}^{h} t(j)u(b_j), \tag{4.3.6}$$

  where $u(\cdot)$ is the utility function describing the individual preferences and $t(\cdot)$ is the subjective discount function. Thus, the expected discounted utility of an annuity with payment structure $\boldsymbol{b}$ becomes

$$\mathbb{E}(D_{\boldsymbol{b}}(K_x)) = \mathbb{E}\left(\sum_{j=0}^{K_x} t(j)u(b_j)\right) = \sum_{h=0}^{\omega-x} {}_{h|1}q_x \sum_{j=0}^{h} t(j)u(b_j). \tag{4.3.7}$$

  Observe that the function $D_{\boldsymbol{b}}(\cdot)$ is merely a deterministic function of K_x. Hence, if the ambiguity regarding the distribution of K_x is described using L^2-balls, all the results obtained in this section can be used to derive the bounds of the expected discounted utility in (4.3.7)*, regardless of the characteristics of the subjective discount function $t(\cdot)$ and of the utility function $u(\cdot)$. A well-established result in decision theory under ambiguity is that an ambiguity adverse decision maker should choose the option that maximizes her expected utility* *under the worst-case scenario. This was proven in a decision theoretic setting in the seminal paper of Gilboa and Schmeidler (1989). See for example d'Albis and Thibault (2012) for a discussion of this approach in the context of the optimal annuitization problem. Being able to compute upper and lower bounds for the expected discounted utility can help an ambiguity averse decision maker to choose which insurance contract is optimal for her.*

Specifically, given an ambiguity set $\mathfrak{F}$ and a set $\mathcal{S}$ of possible payment structures, an ambiguity adverse decision maker with a utility function $u(\cdot)$ and a subjective discount function $t(\cdot)$ faces the following problem:

$$\max_{\boldsymbol{b}\in\mathcal{S}} \min_{K_x\sim\tilde{F},\, \tilde{F}\in\mathfrak{F}} \mathbb{E}(D_{\boldsymbol{b}}(K_x)). \tag{4.3.8}$$

If the ambiguity set $\mathfrak{F}$ is described using an L^2-ball built around a reference distribution and if the set $\mathcal{S}$ of considered payment structures is finite, then our results can be used to compute $\min_{K_x\sim\tilde{F},\tilde{F}\in\mathcal{M}} \mathbb{E}(D_{\boldsymbol{b}}(K_x))$ for all $\boldsymbol{b}\in\mathcal{S}$ and the solution of (4.3.8) *can then be found.*

4.4 Examples of robust assessment for annuities

In this section we propose several numerical examples that illustrate how the results obtained in Section 4.3 can be used to obtain a robust assessment of annuities.

In the following examples, the reference probability distribution F of K_x is $Binomial(n,p)$, with $n=\omega-x+1$. Unless otherwise specified, the parameters adopted in the numerical examples are taken from Table 4.1.

Table 4.1: Parameters used in the numerical examples.

r	x	ω	p
0.025	65	120	0.35

4.4.1 Standard annuity

The first case considered is a standard (paid in advance) whole life annuity, paying 1 euro per year as long as the insured is alive. Assuming a constant interest rate r, the corresponding payoff function $g(h)$ is given as

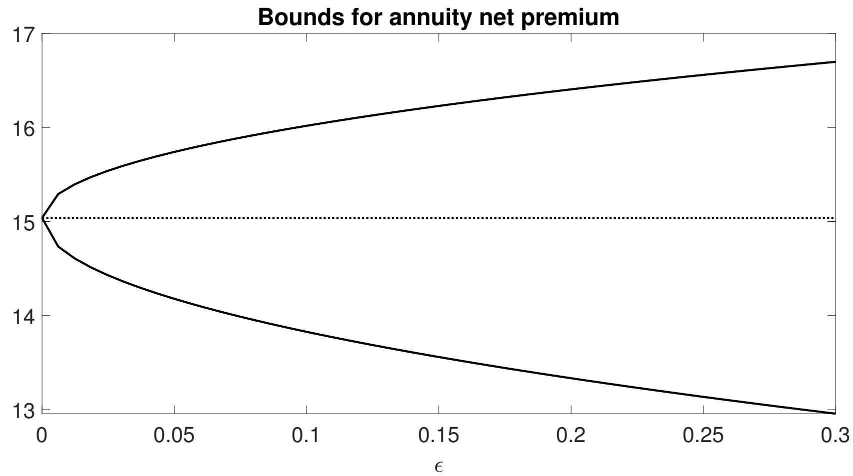
$$g(h) = \begin{cases} 0 & \text{, if } h = 0 \\ a_{\overline{h}|} & \text{, if } h = 1, 2, ..., \omega - x, \end{cases} \tag{4.4.1}$$

where $a_{\overline{h}|} = \sum_{j=1}^{h}(1+r)^{-j}$, and the net premium writes as $\pi = \mathbb{E}\left(g(K_x)\right) = \sum_{h=1}^{\omega-x} a_{\overline{h}|}\ {}_{h|1}q_x$.

The premium upper and lower bounds are displayed in Figure 4.1. Here we find a numerical confirmation of what was anticipated in Proposition 4.3.1, which states that the premium upper bound is convex with respect to ε while the lower bound is concave.

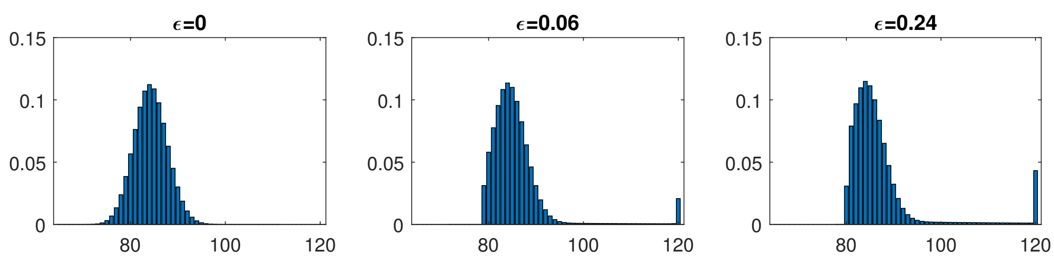
Figure 4.2 shows the evolution of the probability distribution for K_x attaining the premium upper bound corresponding to various levels of ε. Clearly, the worst-case distribution is obtained by

Figure 4.1: Premium bounds with respect to ε, for ε varying between $\varepsilon = 0$ and $\varepsilon = 0.3$. The dashed line describes the premium level computed using the reference distribution. The black lines represents the upper and lower bounds corresponding to each value of ε. The annuity payoff function is given in (4.4.1). The other parameters are taken from Table 4.1.



moving as much probability mass as possible from early to late ages, in particular to the maximal attainable age. The reason behind this shape is that we are maximizing the expected value of an increasing payoff function taking its highest value in $\omega - x$ (corresponding to the case in which the insured reaches the maximal attainable age ω). Note that the distributions reported in Figure 4.2 appear to be somewhat odd. On the one hand, this is due to the fact that we are adopting a

Figure 4.2: Worst-case probability distribution for K_x. The annuity payoff function is given in (4.4.1). The parameters of the reference probability distribution and the interest rate are reported in Table 4.1.

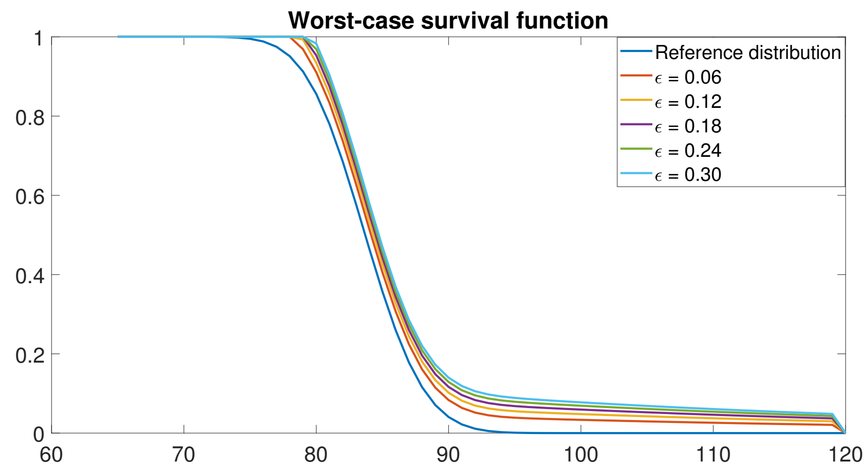


non-parametric approach: we describe uncertainty using a probability metric and hence we do not force the solution to belong to a certain parametric model. The use of a non-parametric approach is justified when the actuary is willing to consider the possibility that future mortality could exhibit some features that standard parametric models are not able to take into account. The distributions displayed in Figure 4.2 correspond to the worst-case scenario in this setup. On the other hand, the graphs in Figure 4.2 can have the following interpretation. By fixing ω, the actuary is essentially fixing the maximal age she believes an insured can reach. Figure 4.2 is then suggesting that this biological bound will not be overcome, but, as worst case scenario, this maximal age could be reached by a higher share of the portfolio population, for example thanks to medical and quality of life improvements. This observation is a direct consequence of our assumption to set to ω the maximal attainable age.

Nonetheless, in Section 4.5 we illustrate how an additional unimodality constraint on the set of feasible distributions can be useful in those situations in which the actuary does not deem realistic to have a high probability mass concentrated at the maximal attainable age.

Figure 4.3 looks at the worst-case probability distribution of K_x from another point of view. Interestingly, as ε increases, the survival function moves toward a rectangular shape. The survival function rectangularization is a well-know phenomenon in actuarial science, illustrated for example in Pitacco et al. (2009).

Figure 4.3: Worst-case survival function of K_x. The annuity payoff function is given in (4.4.1). The parameters of the reference probability distribution and the interest rate are reported in Table 4.1.

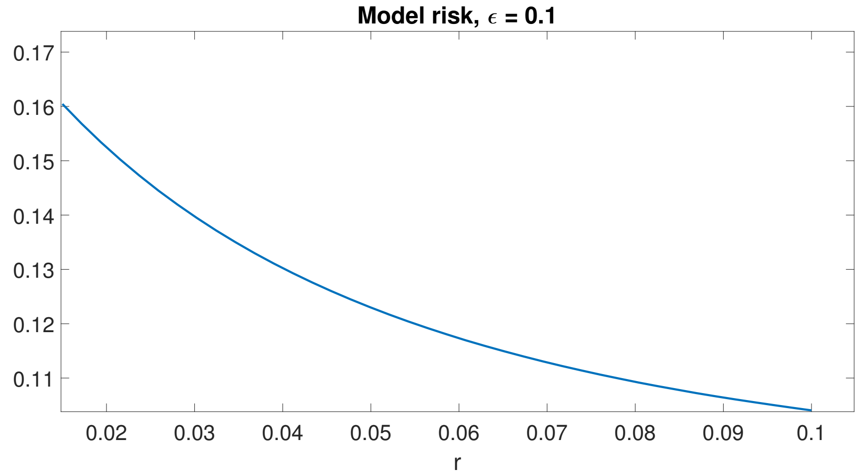


We now fix the ambiguity set (reference distribution and radius of L^2-ball are given) and we study the model risk as a function of the interest rate r. The measure of model risk we adopt here is defined as

$$\mathrm{MR} = \frac{\overline{\pi}_\varepsilon - \underline{\pi}_\varepsilon}{\pi_F}, \tag{4.4.2}$$

where π_F is the net premium computed using the reference distribution F. The quantity MR in (4.4.2) is sometimes called the normalised length of the bounds and is a standard tool in the model risk literature to assess the robustness of a risk measure under distributional uncertainty. See for example Barrieu and Scandolo (2015) and Bernard et al. (2020b) for an introduction to this concept and practical examples. Figure 4.4 displays the MR for several values of the interest rate r. Note that a change in the interest rate r implies a change for the payoff function g in (4.4.1) and thus leads to different values of $\overline{\pi}_\varepsilon$, $\underline{\pi}_\varepsilon$, and π_F. Clearly, this graph displays a decreasing relationship between model risk (MR) and the interest rate r. In the case considered, a higher interest rate implies an increasing robustness for the annuity premium with respect to changes in the underlying residual lifetime distribution. This observation is in line, for example, with the conclusions of Rabitti and Borgonovo (2020), where the authors investigate the relative importance of mortality and financial components on annuity premiums using a sensitivity analysis approach.

Figure 4.4: Model risk as defined in (4.4.2) **with respect to** r**, for** r **varying between** 0.015 **and** 0.1**.** Given $\varepsilon = 0.1$, for each value of the interest rate r, we compute the length of the normalized bounds as defined in (4.4.2). The annuity payoff function is given in (4.4.1). The other parameters are taken from Table 4.1.



4.4.2 Comparison of annuities with different payment structure

To conclude this numerical section, we provide an example of how our results can be used to compare the robustness of an annuity net premium under different payment structures. In particular, we compare the annuity with constant payments in (4.4.1) with annuities having arithmetically increasing payments that, under the reference probability distribution F, have the same net premium. This comparison is conducted on two levels. First, in Figure 4.5 we compare the robustness of the net premium of these contracts with respect to ambiguity. Second, in Figure 4.6 we compare these payment structures using the robust expected discounted utility approach described in Remark 4.3.3. Assuming a constant discounting interest rate r, the payoff function of an annuity with arithmetically increasing payments with initial payment b and rate of increase α is given as

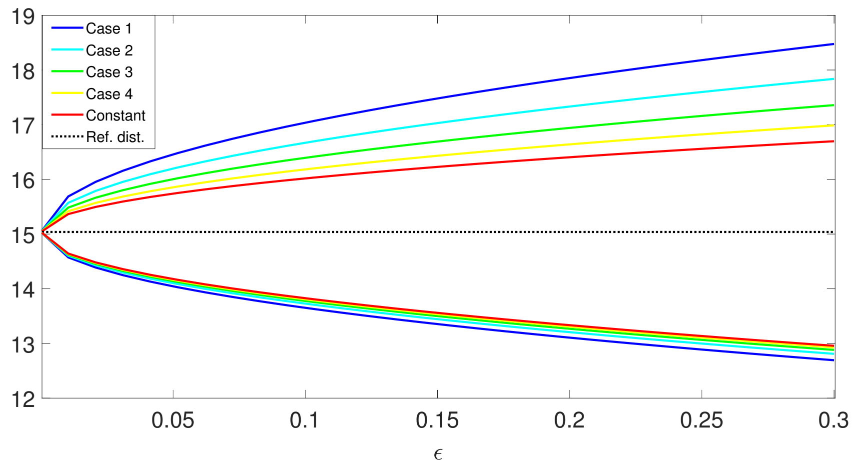
$$g_{b,\alpha}(h) = \begin{cases} 0 & \text{, if } h = 0 \\ \sum_{j=1}^{h} b(1 + (j-1))\alpha \, (1+r)^{-j} & \text{, if } h = 1, 2, ..., \omega - x. \end{cases} \tag{4.4.3}$$

Table 4.2: Parameters of arithmetically increasing payments. The parameters b and α are fixed in a way such that the net premium of the annuities with a payoff function as in (4.4.3) in Cases 1-4 coincides with the net premium of the annuity with constant payments (4.4.1), using the reference distribution from Table 4.1.

	Case 1	Case 2	Case 3	Case 4
b	0.5389	0.7003	0.8238	0.9212
α	0.100	0.050	0.025	0.010

Figure 4.5 shows that, in the case considered, the net premium of the constant annuity is more robust with respect to model risk. The net premium computed with the reference distribution will

Figure 4.5: Upper and lower bounds for the net premium of annuities with arithmetically increasing payments as a function of ε. The payoff function for this class of annuities is given in (4.4.3). The parameters that identify the payment structure of Cases 1-4 are reported in Table 4.2.



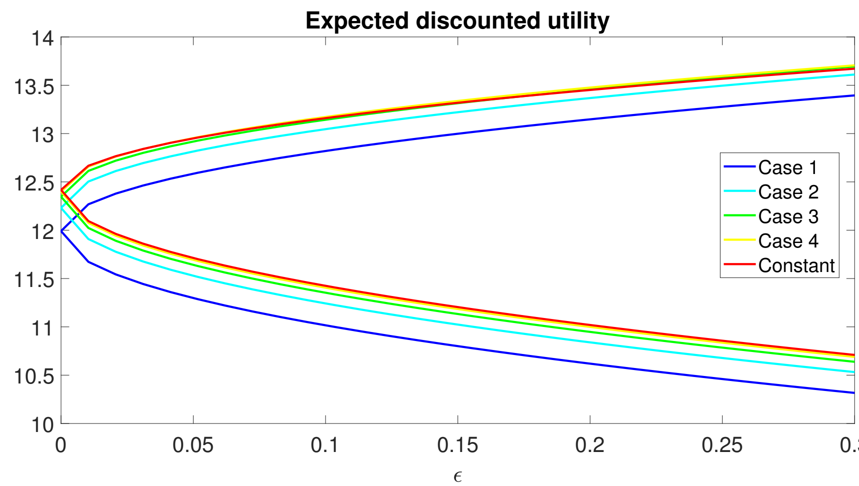
be the same for all these contracts, but the annuity with constant payments exhibits an upper bound for the expected losses that is always lower than in the case of arithmetically increasing payments.

From the point of view of the insured, a comparison of these different payment structures can be obtained using the robust expected discounted utility method described in Remark 4.3.3. In particular, we assume that the preferences of the policyholder can be described by an exponential utility function $u(x) = 1 - e^{-xa}$ in which $a > 0$ is the risk aversion coefficient. Moreover, we consider the case of a constant subjective discount factor $0 < v < 1$, which allows us to write the subjective discount function in the form $t(h) = v^h$ for $h = 0, 1, 2..., \omega - x$. Figure 4.6 displays the bounds for the expected discounted utility given in (4.3.7) as a function of the ambiguity level ε. Observe that for all levels of ε considered, Figure 4.6 shows that the annuity with constant payments is the one leading to the highest lower bound for the insured's expected discounted utility. Figures 4.5 and 4.6 suggest that, at least for the specific cases under consideration, the annuity with constant payments is preferred to the annuities with arithmetically increasing payments if both insurer and insured are ambiguity adverse.

4.5 Unimodality constraint

The goal of this section is to show that the framework we developed is flexible enough to accommodate further constraints on the structure of the probability distribution of K_x, i.e., on $P(K_x = h) = {}_{h|1}q_x$ for $h = 0, 1, ..., \omega - x$, without impairing the bounds' numerical tractability. This is done by considering an additional mode preserving constraint on the set of feasible distributions. In the life insurance context, the mode of the residual life time distribution is sometimes referred as the Lexis point (Pitacco et al. (2009)). An unimodality constraint is not new in the literature related to risk bounds. Li et al. (2018) and Bernard et al. (2020a) derive moment bounds on VaR under a unimodality constraint. However, to the best our knowledge, our study is the first to consider an

Figure 4.6: Bounds for expected discounted utility of annuities given in (4.3.7) with respect to ε. The risk aversion parameter of exponential utility is $a = 2$ and constant subjective discount factor is $v = 0.97$. The payoff function for this class of annuities is given in (4.4.3). The parameters that identify the payment structure of Cases 1-4 are reported in Table 4.2.



additional unimodality constraint when ambiguity is described using a probability metric. First, let us introduce the definition of unimodality that we consider in this analysis.

Definition 4.5.1. *(Keilson and Gerber (1971)) Given a discrete random variable X taking value on $\mathcal{H}$, we say that its distribution is unimodal if there exists at least one $h_m \in \mathcal{H}$ (called the mode) such that $P(X = h)$ is non-decreasing for $h = 0, 1, 2..., h_m$ and $P(X = h)$ is non-increasing for $h = h_m, h_m + 1, ..., \omega - x$.*

For example, a random variable with a discrete uniform distribution on $\mathcal{H}$ is unimodal. Consider now the case in which one is interested in finding the bounds for the net premium of an insurance contract having a payoff function g, considering all distributions of K_x that satisfy an L^2 distance constraint from a reference distribution F and that additionally are unimodal with mode equal to h_m. The probability distributions attaining these bounds can be found by solving the following problem:

$$
\begin{aligned}
\min_{\boldsymbol{q}} \quad & \langle \boldsymbol{y}, \boldsymbol{q} \rangle = \sum_{h=0}^{\omega-x} y_h \; {}_{h|1}q_x \\
\text{subject to} \quad & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 \leqslant \varepsilon, \\
& \sum_{h=0}^{\omega-x} {}_{h|1}q_x = 1, \\
& {}_{h|1}q_x \geqslant 0, \; h = 0, ..., \omega - x, \\
& {}_{h|1}q_x \geqslant {}_{h+1|1}q_x, \; h = h_m, ..., \omega - x - 1, \\
& {}_{h|1}q_x \geqslant {}_{h-1|1}q_x, \; h = 2, ..., h_m,
\end{aligned}
\tag{4.5.1}
$$

where the vector $\boldsymbol{y} = (y_0, y_1, ..., y_{\omega-x})$ can be defined as $y_h = -g(h)$ and $y_h = g(h)$ for $h \in \mathcal{H}$ in order to derive the distributions that attain the upper and lower bound for $\mathbb{E}(g(K_x))$, respectively. The following proposition ensures that if the reference distribution is unimodal, an additional mode preserving constraint does not affect the numerical tractability of the problem at hand.

Proposition 4.5.2. *Let the reference distribution be unimodal with modal value h_m and $\varepsilon > 0$. Then, Problem* (4.5.1) *is well-posed and its feasible region is a non-empty, compact, and convex set.*

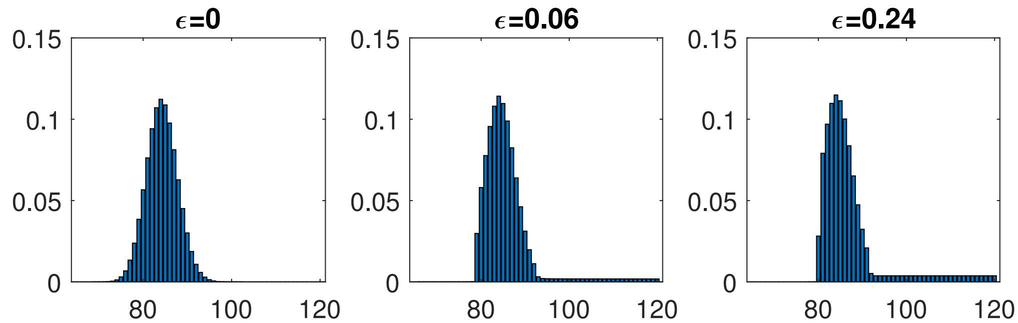
The proof of Proposition 4.5.2 follows from an argument that is similar to the proof of Proposition 4.3.1 and thus is omitted. Figure 4.7 points out an interesting aspect regarding the solutions to problems (4.3.1) and (4.5.1). If one compares Figure 4.7 to Figure 4.2, it is clear that the unimodality constraint can have an impact on the shape of the distribution attaining the upper bound. In particular, this constraint on the modal value prevents the worst-case probability distribution from having a high probability mass concentrated at maximal attainable age, a feature that we observed in Figure 4.2. Thus, a unimodality constraint can be useful in those situations in which probability distributions in the form represented in Figure 4.2 are not considered acceptable as worst-case scenarios. It would be interesting to study whether it is possible to obtain solutions for Problem (4.5.1) when the reference distribution is not unimodal or when one imposes a modal value that does not coincide with that of the reference distribution, but we leave this issue to future research.

4.6 Comparison with Pichler (2014)

Given a coherent risk measure ϱ, Pichler (2014) studies the upper bound of the values that $\varrho(g(K_x))$[2] can take when there is ambiguity regarding the distribution of K_x. In the present analysis we focus on the upper and lower bounds on the expectation of $g(K_x)$, meaning that we consider the simple

[2] Note Pichler (2014) denotes the payoff function using the symbol L, while we denote the payoff function using g.

Figure 4.7: Worst-case probability distribution for K_x under L^2 distance and unimodality constraints with modal value at age 85. The annuity payoff function is given in (4.4.1). The reference probability distribution parameters and the interest rate are reported in Table 4.1.



case in which $\varrho(\cdot) = \mathbb{E}(\cdot)$, and thus there is apparently an overlap between our analysis and that of Pichler (2014). However, our approach is quite different from a methodological point of view. The aim of this section is to point out some methodological differences and to list the advantages of our set-up.

First, in the present analysis we assume that the residual lifetime K_x is a bounded random variable; that is, there exists ω such that $P(K_x \leqslant \omega - x) = 1$, while Pichler (2014) implicitly assumes that K_x is unbounded from the right. We believe that our set-up is consistent with usual practice, since the life-tables used by insurance companies typically consider only a finite number of possible ages that an individual can reach.

Second, Pichler (2014) describes the ambiguity around a reference distribution using the Wasserstein distance, whereas we consider the L^2 distance defined in (4.2.4). For any given $p \geqslant 0$, the Wasserstein distance of order p between two distributions on the real line writes as

$$d_{W_p}(\widetilde{F}, F) = \left(\int_0^1 \left| \widetilde{F}^{-1}(\alpha) - F^{-1}(\alpha) \right|^p d\alpha \right)^{\frac{1}{p}}. \tag{4.6.1}$$

For $p = 1$, $d_{W_1}(\widetilde{F}, F)$ coincides with the L^1 distance between distribution functions, but correspondence between Wasserstein distances of order p and L^p distances it is not true in general for $p \neq 1$. Thus, even if one fixes the same reference distribution F and the same radius, the premium bounds obtained using the Wasserstein distance or using the L^2 distance do not coincide in general.

A priori, it is not clear which metric one should use to describe distributional uncertainty in the context of life insurance pricing. See for example Chapter 3 in Rachev et al. (2008) for a discussion of the characteristics of various probability metrics with financial applications. A main advantage of our approach is that it makes the problem much more tractable from the numerical and mathematical points of view.

From a mathematical point of view, Pichler (2014) obtains most of his results under the assumption of Hölder continuity of the payoff function (see Definition 19 in Pichler (2014)), and the bounds obtained in Theorems 20 and 22 require the computation of the Hölder constant. Observe that the assumption of Hölder continuity needs to be verified case by case, and this can be difficult if one has a non-constant term structure or a contract with non-constant payments. We do not make any assumption on the payoff function $g(\cdot)$, but we are still able to derive premium bounds and analytic expressions of their attaining distributions (Theorem 4.3.6), and to prove that the premium bounds are continuous with respect to the parameter ε (Proposition 4.3.3). This feature allows us, for example, to extend the applicability of the results obtained in Section 4.3 to other contexts, as in Remark 4.3.3.

From a numerical point of view, we show that with the L^2 metric the bounds on the price can be reformulated as an easy-to-solve convex problem, while Pichler (2014) states that computing the bounds numerically using the Wasserstein distance requires solving an involved bilinear optimization. Finally, in Section 4.5 we show that our approach is flexible enough to handle additional constraints such as a constraint on the modal value, and this was not considered in Pichler (2014).

4.7 Final remarks

In this Chapter we develop new tools that insurance companies can use to assess the impact of model risk on the net premium of standard life insurance products. Our analysis focuses on those situations in which uncertainty regarding future mortality trends is present and the net premium is determined using the equivalence principle. We study the case in which the distributional uncertainty regarding the underlying residual lifetime distribution is described via L^2 distance. In particular, the use of this metric makes it possible to reformulate the premium bounds' problem as a Quadratically Constrained Linear Program, which is numerically tractable (convex) and easy-to-implement. We further study the properties of this linear program and show that in some cases explicit formulas for the premium bounds and their attaining distribution functions can be obtained. A numerical section illustrates how the results we obtain can be useful in assessing the robustness of a life insurance contract from various points of view. Finally, we show that additional constraints, such as the important case of an unimodality restriction, can be easily incorporated into our framework.

4.8 Appendix

4.8.1 Proof of Proposition 4.2.2

Proof. Given an ambiguity set $\mathfrak{F}$,

1. Let $\tilde{F} \in \mathfrak{F}$, $g_1(\cdot) \leqslant g_2(\cdot)$ and $K_x \sim \tilde{F}$. Then, $g_2(K_x)$ is

   first order stochastically larger than $g_1(K_x)$, see Definition 2.2.1 in Chapter **??**. This implies $\mathbb{E}(g_1(K_x)) \leqslant \mathbb{E}(g_2(K_x))$ for all $\tilde{F} \in \mathfrak{F}$, which brings us to $\pi_{g_1}^{\mathfrak{F}} \leqslant \pi_{g_2}^{\mathfrak{F}}$.

2. If $g_1(\cdot) = c + g_2(\cdot)$ with $c \in \mathbb{R}$, then

$$\begin{aligned}\pi_{g_1}^{\mathfrak{F}} &= \sup\{\mathbb{E}(g_1(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} = \sup\{c + \mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} \\ &= c + \sup\{\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} = c + \pi_{g_2}^{\mathfrak{F}}.\end{aligned}$$

3. if $g_1(\cdot) = \lambda g_2(\cdot)$ with $\lambda \geqslant 0$, then

$$\begin{aligned}\pi_{g_1}^{\mathfrak{F}} &= \sup\{\mathbb{E}(g_1(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} = \sup\{\lambda\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} \\ &= \lambda\pi_{g_2}^{\mathfrak{F}}.\end{aligned}$$

4. if $g(\cdot) = \alpha g_1(\cdot) + (1-\alpha)g_2(\cdot)$ with $\alpha \in [0,1]$ then,

$$\begin{aligned}\pi_{g}^{\mathfrak{F}} &= \sup\{\alpha\mathbb{E}(g_1(K_x)) + (1-\alpha)\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} \\ &\leqslant \alpha\sup\{\mathbb{E}(g_1(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} + (1-\alpha)\sup\{\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} \\ &= \alpha\pi_{g_1}^{\mathfrak{F}} + (1-\alpha)\pi_{g_2}^{\mathfrak{F}}.\end{aligned}$$

□

4.8.2 Proof of Proposition 4.3.1

Proof. Let us denote with $\mathcal{D}_\varepsilon$ the feasible region of Problem (4.3.1), i.e.

$$\mathcal{D}_\varepsilon = \mathcal{D}_1 \cap \mathcal{D}_2 \cap \mathcal{D}_3, \tag{4.8.1}$$

where

$$\mathcal{D}_1 = \left\{\boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \middle| \sum_{h=0}^{\omega-x}\left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h\right)^2 \leqslant \varepsilon\right\}, \tag{4.8.2}$$

$$\mathcal{D}_2 = \left\{\boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \middle| \boldsymbol{e}^t\boldsymbol{q} = 1\right\} \text{ with } \boldsymbol{e} = (1,1,...,1), \tag{4.8.3}$$

$$\mathcal{D}_3 = \left\{\boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \middle| \boldsymbol{q} \geqslant 0\right\}. \tag{4.8.4}$$

Observe that $\mathcal{D}_2 \cap \mathcal{D}_3$ is clearly a convex and closed set. The set $\mathcal{D}_1$ is defined as the ε-sublevel set of the function $d^2 : \boldsymbol{q} \to \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2$, which is the sum of $\omega - x + 1$ functions defined by

$$d_h : \mathbb{R}^{h+1} \to \mathbb{R}, \text{ with } d_h({}_{0|1}q_x, ..., {}_{h|1}q_x) = \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 .$$

For any given h, the Hessian matrix of d_h is a $h+1$ by $h+1$ matrix whose elements are all equal to 2. It is thus a positive semidefinite matrix. Hence, all functions d_h are convex, which means that d^2 is also a convex function of $\boldsymbol{q}$. The continuity and convexity of d^2 imply that its ε-sublevel set $\mathcal{D}_1$ is closed and convex. To conclude, the feasible set $\mathcal{D}_\varepsilon$ is obtained as the intersection of two closed and convex sets and hence it is closed and convex. Since $\mathcal{D}_\varepsilon$ is a subset of $[0, 1]^{\omega-x+1}$, $\mathcal{D}_\varepsilon$ is also bounded and therefore compact. Finally, $\mathcal{D}_\varepsilon$ cannot be empty since the vector describing the probability distribution induced by the reference df F always belongs to $\mathcal{D}_\varepsilon$.

Second, we prove that Problem (4.3.1) admits at least one solution. Note that in Problem (4.3.1) we are looking for the minimum of the function $\boldsymbol{q} \mapsto \langle \boldsymbol{y}, \boldsymbol{q} \rangle$ that is linear with respect to $\boldsymbol{q}$. Therefore, the image of $\mathcal{D}_\varepsilon$ through this function is a closed interval and hence there exists at least one element in $\mathcal{D}_\varepsilon$ that minimizes $\langle \boldsymbol{y}, \boldsymbol{q} \rangle$. □

4.8.3 Proof of Proposition 4.3.3

Proof. To prove the statement, it is sufficient to show that the mapping $\varepsilon \mapsto \langle \boldsymbol{y}, \boldsymbol{q}_\varepsilon^* \rangle$ is convex, where $\boldsymbol{q}_\varepsilon^*$ is an optimizing distribution for Problem (4.3.1). Fix a reference distribution F, and let D_ε be defined as in (4.8.1). From Proposition 4.3.1, D_ε is convex and compact for any $\varepsilon > 0$. Observe that

$$\langle \boldsymbol{y}, \boldsymbol{q}_\varepsilon^* \rangle = \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q} \rangle \;\middle|\; \boldsymbol{q} \in \mathcal{D}_\varepsilon \right\}.$$

Fix now $\lambda \in (0, 1)$, $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$. Let us define the following set,

$$\mathcal{D}_{\varepsilon_1,\varepsilon_2}^\lambda = \left\{ \boldsymbol{q}_\lambda \;\middle|\; \boldsymbol{q}_\lambda = \lambda \boldsymbol{q}_1 + (1-\lambda)\boldsymbol{q}_2,\ \boldsymbol{q}_1 \in \mathcal{D}_{\varepsilon_1},\ \boldsymbol{q}_2 \in \mathcal{D}_{\varepsilon_2} \right\}.$$

In Appendix 4.8.2 we showed that the function $d^2 : \boldsymbol{q} \to \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2$ is convex, and thus we have

$$d^2(\boldsymbol{q}_\lambda, F) \leqslant \lambda d^2(\boldsymbol{q}_1, F) + (1-\lambda) d^2(\boldsymbol{q}_2, F) = \lambda \varepsilon_1 + (1-\lambda)\varepsilon_2.$$

It is then clear that $\mathcal{D}_{\varepsilon_1,\varepsilon_2}^\lambda \subseteq \mathcal{D}_{\lambda\varepsilon_1+(1-\lambda)\varepsilon_2}$. Finally, we have

$$\begin{aligned}
\langle \boldsymbol{y}, \boldsymbol{q}_{\lambda\varepsilon_1+(1-\lambda)\varepsilon_2}^* \rangle &= \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q} \rangle \;\middle|\; \boldsymbol{q} \in \mathcal{D}_{\lambda\varepsilon_1+(1-\lambda)\varepsilon_2} \right\} \leqslant \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q}_\lambda \rangle \;\middle|\; \boldsymbol{q}_\lambda \in \mathcal{D}_{\varepsilon_1,\varepsilon_2}^\lambda \right\} \\
&= \lambda \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q}_1 \rangle \;\middle|\; \boldsymbol{q}_1 \in \mathcal{D}_{\varepsilon_1} \right\} + (1-\lambda) \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q}_2 \rangle \;\middle|\; \boldsymbol{q}_2 \in \mathcal{D}_{\varepsilon_2} \right\} \\
&= \lambda \langle \boldsymbol{y}, \boldsymbol{q}_{\varepsilon_1}^* \rangle + (1-\lambda) \langle \boldsymbol{y}, \boldsymbol{q}_{\varepsilon_2}^* \rangle.
\end{aligned}$$

□

4.8.4 Proof of Proposition 4.3.4

Proof. Consider the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{q}} \quad & \sum_{h=0}^{\omega-x} y_h \ {}_{h|1}q_x \\ \text{subject to} \quad & \sum_{j=0}^{\omega-x} {}_{j|1}q_x = 1, \\ & {}_{h|1}q_x \geqslant 0, \forall h = 0, ..., \omega - x. \end{aligned} \tag{4.8.5}$$

Note that the solution of Problem (4.8.5) coincides with y_{min}, trivially. Furthermore, a degenerate probability distribution attains the equality $y_{min} = \sum_{h=0}^{\omega-x} y_h \ {}_{h|1}q_x$ if and only if ${}_{h^*|1}q_x = 1$ for one $h^* \in \mathcal{H}^{\boldsymbol{y}}_{min}$ and ${}_{h^*|1}q_x = 0$ for $h \neq h^*$. Thus, the number of degenerate probability distributions solving Problem (4.8.5) coincides with the number of elements in $\mathcal{H}^{\boldsymbol{y}}_{min}$. Let us now denote with $\boldsymbol{q}_{h^*}$ a degenerated probability distribution such that ${}_{h^*|1}q_x = 1$ and ${}_{h|1}q_x = 0$ for $h \neq h^*$. The square of the L^2 distance between the df of $\boldsymbol{q}_{h^*}$ and the reference distribution F is given by

$$d(\boldsymbol{q}_{h^*}, F)^2 = \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\omega-x} (1 - F_h)^2.$$

It is then clear that if $\varepsilon \geqslant \min_{h \in \mathcal{H}^{\boldsymbol{y}}_{min}} \sum_{j=0}^{h-1} F_j^2 + \sum_{j=h}^{\omega-x} (F_j - 1)^2$, then there exists $h^* \in \mathcal{H}^{\boldsymbol{y}}_{min}$ such that $\boldsymbol{q}_{h^*}$ belongs to the feasible set of Problem (4.3.1), and thus Problem (4.3.1) admits degenerate solutions.

We shall now prove that if $\varepsilon < \min_{h \in \mathcal{H}^{\boldsymbol{y}}_{min}} \sum_{j=0}^{h-1} F_j^2 + \sum_{j=h}^{\omega-x} (F_j - 1)^2$, a solution of Problem (4.3.1) cannot have all probability mass concentrated in one point. Consider h^* in $\mathcal{H} \setminus \mathcal{H}^{\boldsymbol{y}}_{min}$, and consider the distribution $\boldsymbol{q}_{h^*}$ such that ${}_{h^*|1}q_x = 1$ and ${}_{h|1}q_x = 0$ for $h \neq h^*$. Let $y^* = \langle \boldsymbol{y}, \boldsymbol{q}_{h^*} \rangle$. Assume that $\boldsymbol{q}_{h^*}$ is a distribution attaining the minimum of Problem (4.3.1), and thus $d(\boldsymbol{q}_{h^*}, F) \leqslant \varepsilon$. Fix now $\tilde{h} \in \mathcal{H}^{\boldsymbol{y}}_{min}$.

First, we focus on the case where $\tilde{h} > h^*$. Consider the distribution $\tilde{\boldsymbol{q}}$ such that

$$\begin{aligned} {}_{h|1}\tilde{q}_x &= 1 - \delta, \text{ for } h = h^*, \\ {}_{h|1}\tilde{q}_x &= \delta, \text{ for } h = \tilde{h}, \\ {}_{h|1}\tilde{q}_x &= 0, \text{ elsewhere}, \end{aligned}$$

with $0 < \delta < 1$. Since $\tilde{h} \in \mathcal{H}^{\boldsymbol{y}}_{min}$, it is clear that $\langle \boldsymbol{y}, \tilde{\boldsymbol{q}} \rangle < \langle \boldsymbol{y}, \boldsymbol{q}_{h^*} \rangle$. Fix $\delta = \frac{1 - F_{\omega-x-1}}{2}$, so that $0 < 1 - \delta - F_h < 1 - F_h$ holds for any $h \leqslant \omega - x - 1$. Since $\boldsymbol{f} > 0$ and δ is strictly positive, we have

$$
\begin{aligned}
d(\tilde{\boldsymbol{q}}, F)^2 &= \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\tilde{h}-1} (1-\delta-F_h)^2 + \sum_{h=\tilde{h}}^{\omega-x} (1-F_h)^2 \\
&< \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\omega-x} (1-F_h)^2 = d(\boldsymbol{q}_{h^*}, F)^2 \leqslant \varepsilon.
\end{aligned}
$$

Thus, $\tilde{\boldsymbol{q}}$ belongs to the feasible set of Problem (4.3.1) and leads to a strictly lower value of the objective function. This shows that the degenerate distribution $\boldsymbol{q}_{h^*}$ cannot be a solution of Problem (4.3.1).

Let us now consider the case $\tilde{h} < h^*$. Again, we define the distribution $\tilde{\boldsymbol{q}}$ such that

$$
\begin{aligned}
{}_{h|1}\tilde{q}_x &= 1-\delta, \text{ for } h = h^*, \\
{}_{h|1}\tilde{q}_x &= \delta, \text{ for } h = \tilde{h}, \\
{}_{h|1}\tilde{q}_x &= 0, \text{ elsewhere.}
\end{aligned}
$$

with $0 < \delta < 1$. Since $\tilde{h} \in \mathcal{H}^{\boldsymbol{y}}_{min}$, it is clear that $\langle \boldsymbol{y}, \tilde{\boldsymbol{q}} \rangle < \langle \boldsymbol{y}, \boldsymbol{q}_{h^*} \rangle$. Fix $\delta = \frac{F_{\tilde{h}}}{2}$. Since $\boldsymbol{f} > 0$ and δ is strictly positive, we have

$$
0 < F_h - \delta < F_h \implies \left(\sum_{j=0}^{h} {}_{h|1}\tilde{q}_x - F_h \right)^2 = (F_h - \delta)^2 < F_h^2, \text{ for } \tilde{h} \leqslant h < h^* - 1.
$$

Thus,

$$
\begin{aligned}
d(\tilde{\boldsymbol{q}}, F)^2 &= \sum_{h=0}^{\tilde{h}-1} F_h^2 + \sum_{h=\tilde{h}}^{h^*-1} (F_h-\delta)^2 + \sum_{h=\tilde{h}}^{\omega-x} (1-F_h)^2 \\
&< \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\omega-x} (1-F_h)^2 = d(\boldsymbol{q}_{h^*}, F)^2 \leqslant \varepsilon.
\end{aligned}
$$

Thus, $\tilde{\boldsymbol{q}}$ belongs to the feasible set of Problem (4.3.1) and leads to a strictly lower value of the objective function. This shows that a degenerate distribution $\boldsymbol{q}_{h^*}$ cannot be a solution of Problem (4.3.1). □

4.8.5 Proof of Lemma 4.3.5

Proof. Let $\boldsymbol{f} > 0$ and $\boldsymbol{q}$ be a probability distribution such that there exists $k \in \mathcal{H}$ for which ${}_{h|1}q_x = 0$. Then, $d(\boldsymbol{q}, F)^2 \geqslant \varepsilon^*$ with

$$\begin{aligned} \varepsilon^* = \min_{\boldsymbol{z}} \quad & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} z_j - F_h \right)^2 \\ \text{subject to} \quad & \boldsymbol{z} \in \mathbb{R}^{\omega-x+1}, \\ & z_k = 0. \end{aligned} \tag{4.8.6}$$

Assume $0 < k < \omega - x$. In order to solve (4.8.6), we adopt the standard Lagrangian multiplier method with a Lagrangian function defined as

$$\mathcal{L}(\boldsymbol{z}, \lambda) = \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} z_j - F_h \right)^2 - \lambda \left(z_k - 0 \right). \tag{4.8.7}$$

After differentiating with respect to z_h, we get the following system of equations:

$$\begin{cases} \sum_{h=i}^{\omega-x} \sum_{j=0}^{h} z_j = \sum_{h=i}^{\omega-x} F_h, \text{ for } i \neq k, \\ \sum_{h=k}^{\omega-x} \sum_{j=0}^{h} z_j = \sum_{h=k}^{\omega-x} F_h - \frac{\lambda}{2}. \end{cases} \tag{4.8.8}$$

The system in (4.8.8) is a system of linear equations. It can thus be solved using a standard substitution method, which leads to the following solution:

$$\begin{cases} z_i = f_i, \text{ for } i \neq k-1, k, k+1, \\ z_i = f_i - \frac{\lambda}{2}, \text{ for } i = k-1, k+1, \\ z_k = f_k + \lambda. \end{cases} \tag{4.8.9}$$

To satisfy the constraint in (4.8.6), we get $\lambda = -f_k$. At this point we can easily compute the solution of (4.8.6), which is $\varepsilon^* = \frac{f_k^2}{2}$. Thus, if the value of ε in Problem (4.3.1) is lower than $\frac{f_k^2}{2}$, no feasible distribution can satisfy ${}_{k|1}q_x = 0$. Using a similar argument, one can show that if $k = 0$ or $k = \omega - x$, the solutions of (4.8.6) are $\varepsilon^* = \frac{f_{\omega-x}^2}{2}$ and $\varepsilon^* = f_0^2$, respectively. □

4.8.6 Proof of Theorem 4.3.6

Proof. Recall that from Proposition 4.3.1 we know that Problem (4.3.1) admits at least one optimizing distribution. First, we show that if $\boldsymbol{f} > 0$ and ε is such that any feasible probability distribution satisfies $\boldsymbol{q} > 0$, then an optimizing distribution $\boldsymbol{q}^*$ of Problem (4.3.1) must satisfy $d(\boldsymbol{q}^*, F)^2 = \varepsilon$. Let $\boldsymbol{q}$ be a feasible distribution such that $d(\boldsymbol{q}, F)^2 < \varepsilon$. Fix h_{min} in $\mathcal{H}_{min}^{\boldsymbol{y}}$, $\tilde{h}$ in

$\mathcal{H} \setminus \mathcal{H}^{\boldsymbol{y}}_{min}$, and let us define $\boldsymbol{q}^*$ as follows:

$$\begin{aligned}&{}_{h_{min}|1}q^*_x = {}_{h_{min}|1}q_x + \delta,\\&{}_{\tilde{h}|1}q^*_x = {}_{\tilde{h}|1}q_x - \delta,\\&{}_{h|1}q^*_x = {}_{h|1}q_x,\ \text{for } h \neq h_{min}, \tilde{h}.\end{aligned}$$

Observe that for any $\delta > 0$, we have $\langle \boldsymbol{y}, \boldsymbol{q}^* \rangle < \langle \boldsymbol{y}, \boldsymbol{q} \rangle$. Clearly, $d(\boldsymbol{q}^*, F)^2$ is a continuous function of δ and for $\delta = 0$ we have $d(\boldsymbol{q}^*, F)^2 = d(\boldsymbol{q}, F)^2 < \varepsilon$. Thanks to the continuity of $d(\boldsymbol{q}^*, F)^2$ w.r.t. δ, we know that there exists $\delta > 0$ such that $d(\boldsymbol{q}^*, F)^2 \leqslant \varepsilon$. This implies that $\boldsymbol{q}$ can not be an optimizing distribution for Problem (4.3.1). Thus, any optimizing distribution must have the maximal L^2 distance from the reference distribution.

Second, we look for the optimizing distributions of Problem (4.3.1) by means of KKT conditions. This is justified by the fact that Problem (4.3.1) has differentiable objective and constraint functions. Slater's conditions are satisfied since the reference probability distribution $\boldsymbol{f}$ is strictly feasible in that $\boldsymbol{f}$ has strictly positive probabilities by assumption and $d\left(\boldsymbol{f}, F\right)^2 = 0 < \varepsilon$. Thus, strong duality holds. As a consequence, any optimizing distribution of Problem (4.3.1) must satisfy the KKT conditions. Moreover, from Proposition 4.3.1 we know that Problem (4.3.1) is convex and therefore we conclude that a feasible point is an optimizing distribution of Problem (4.3.1) if and only if it satisfies the KKT conditions. For a detailed explanation of Slaters's condition, strong duality, and KKT conditions we refer to Boyd and Vandenberghe (2004), sections 5.2.3 and 5.5.3, respectively. The KKT conditions of Problem (4.3.1) write as follows:

1. $d(\boldsymbol{q}^*, F)^2 - \varepsilon \leqslant 0,$
2. $\sum_{h=0}^{\omega-x} {}_{h|1}q^*_x = 1,$
3. ${}_{h|1}q^*_x \geqslant 0,\ \text{for } h = 0, 1, ..., \omega - x,$
4. $\lambda^* \geqslant 0,$
5. $\nu^*_h \geqslant 0,\ \text{for } h = 0, 1, ..., \omega - x,$
6. $\lambda^* \left(d(\boldsymbol{q}^*, F)^2 - \varepsilon \right) = 0,$
7. ${}_{h|1}q^*_x\, \nu^*_h = 0,\ \text{for } h = 0, 1, ..., \omega - x,$
8. $\nabla_{\boldsymbol{q}^*} \mathcal{L}\left(\boldsymbol{q}^*, \lambda^*, \mu^*, \boldsymbol{\nu}^*\right) = \boldsymbol{0},$

where $\mathcal{L}\left(\boldsymbol{q}, \lambda, \mu, \boldsymbol{\nu}\right)$ is the Lagrangian function corresponding to Problem (4.3.1),

$$\mathcal{L}\left(\boldsymbol{q}, \lambda, \mu, \boldsymbol{\nu}\right) = \sum_{h=0}^{\omega-x} y_h\ {}_{h|1}q_x + \lambda \left(\sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 - \varepsilon \right) + \mu \left(\sum_{h=0}^{\omega-x} {}_{h|1}q^*_x - 1 \right) + \sum_{h=0}^{\omega-x} {}_{h|1}q_x\ \nu_h. \tag{4.8.10}$$

and the KKT condition 8 is obtained by equating to 0 the following partial derivatives,

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{q},\lambda,\mu,\boldsymbol{\nu})}{\partial_{i|1}q_x} &= y_i + 2\lambda\left(\sum_{h=i}^{\omega-x}\left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h\right)\right) + \mu + \nu_i \\ &= y_i + 2\lambda\left(\sum_{h=i}^{\omega-x}\sum_{j=0}^{h} {}_{j|1}q_x\right) + 2\lambda\sum_{h=i}^{\omega-x} F_h + \mu + \nu_i.\end{aligned}$$

Hence, for any given λ, μ and $\boldsymbol{\nu}$ we can write the system of equations $\nabla_{\boldsymbol{q}}\mathcal{L}(\boldsymbol{q},\lambda,\mu,\boldsymbol{\nu}) = \mathbf{0}$ in the following compact form

$$\boldsymbol{y} + 2\lambda\left(M\boldsymbol{q} - \boldsymbol{F}_{cum}\right) + \mu + \boldsymbol{\nu} = \mathbf{0} \tag{4.8.11}$$

where $\boldsymbol{F}_{cum}$ is the column vector such that $\boldsymbol{F}_{cum}(k) = \sum_{h=k}^{\omega-x} F_h$, for $k = 0, 1, ..., \omega - x$ and M is a $\omega - x + 1 \times \omega - x + 1$ symmetric matrix whose components are $M(k,j) = \omega - x + 2 - \max(k,j)$, for $k, j = 1, 2, ..., \omega - x + 1$. Using the standard Gauss Algorithm, one can easily check that M has full rank, and hence it is an invertible matrix and we deduce that the system of equations in (4.8.11) has a unique solution in the form

$$\boldsymbol{q}^* = M^{-1}\left(\boldsymbol{F}_{cum} - \frac{\boldsymbol{y} + \mu + \boldsymbol{\nu}}{2\lambda}\right). \tag{4.8.12}$$

After some calculations, we find

$$\begin{aligned}{}_{0|1}q_x^* &= f_0 + \frac{y_1 - y_0 + \nu_1 - \nu_0}{2\lambda^*}, \\ {}_{h|1}q_x^* &= f_h + \frac{y_{h-1} - 2y_h + y_{h+1} + \nu_{h-1} - 2\nu_h + \nu_{h+1}}{2\lambda^*} \quad \text{for } h = 1, 2, ..., \omega - x - 1, \\ {}_{\omega-x|1}q_x^* &= f_{\omega-x} + \frac{y_{\omega-x-1} - 2y_{\omega-x} + \nu_{\omega-x-1} - 2\nu_{\omega-x} - \mu}{2\lambda^*}.\end{aligned}$$

At this stage, we just need to find the values of λ, μ, and $\boldsymbol{\nu}$ that satisfy KKT conditions 1-7.

First, $\boldsymbol{\nu}$: since we are considering a case in which any feasible point must satisfy $\boldsymbol{q} > 0$, the constraints in the form ${}_{h|1}q_x \geqslant 0$ cannot be active at the optimal point, and therefore in our problem KKT conditions 3, 5, and 7 are satisfied if and only if $\boldsymbol{\nu} = \mathbf{0}$.

Second, μ: observe that $\sum_{h=0}^{\omega-x} {}_{h|1}q_x^* = \sum_{h=0}^{\omega-x} f_h + \frac{-y_n-\mu}{2\lambda}$. Hence, to satisfy the KKT condition 2 we need

$$\sum_{h=0}^{\omega-x} {}_{h|1}q_x^* = 1 \iff \mu^* = -y_n.$$

Third, λ: we have already shown that any optimizing distribution of Problem (4.3.1) must satisfy $d(\boldsymbol{q}^*, F)^2 - \varepsilon = 0$. This constraint is therefore active at the solution, and we just need to find the corresponding $\lambda > 0$. Using the previous results on $\boldsymbol{\nu}^*$ and μ^* we have

$$\begin{aligned} d(\boldsymbol{q}^*, F)^2 = \varepsilon \iff & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x^* - F_h \right)^2 = \varepsilon \\ \iff & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} \left(f_j + \frac{c_j}{2\lambda} \right) - F_h \right)^2 = \varepsilon \\ \iff & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} \frac{c_j}{2\lambda} \right)^2 = \varepsilon \end{aligned}$$

where $\sum_{j=0}^{h} c_j = y_{h+1} - y_h$ for $h = 0, 1, ..., \omega - x - 1$ and $\sum_{j=0}^{\omega-x} c_j = 0$. We conclude that the only λ that satisfies both condition 1 and condition 4 is $\lambda^* = \sqrt{\frac{\sum_{h=0}^{\omega-x-1} \left(y_{h+1} - y_h \right)^2}{4\varepsilon}}$, which completes the proof. □

Chapter 5

Discussion and outlook

The results obtained in this thesis can be extended in several directions. In Chapter 2, we derived quite general sufficient conditions on the joint distribution of the asset returns leading to a two- and three-fund separation theorem, under the sole assumption that the investor's preferences are consistent with the first-order stochastic dominance rule. Moreover, we discuss the implications of these results for portfolio optimization problems. Note that when the asset returns distribution is determined by two parameters (location-scale) we obtained a two-fund theorem, and when it is a function of three parameters (location-scale mixture) a three-fund separation holds. Therefore, it would be interesting to study a similar problem under more general distributional assumptions, and verify for example if a distributional assumption involving four parameters leads to a four-fund theorem. The general idea would be to establish if there exists a link between the number of parameters used to identify the asset returns distribution and the number of market funds in which the investor should allocate her initial wealth. Regarding our distributional assumptions, an additional level of generalization could be reached by studying the case in which the joint distribution of the asset returns is not completely specified. The stream of literature related to this class of problems goes by the name of robust portfolio selection; see Kim et al. (2014) and Xidonas et al. (2020) for a review of this research line. A first result linking the fund separation theorems with robust portfolio selection was obtained in Baviera and Bianchi (2021). These authors studied the portfolio optimization problem of a mean-variance investor assuming that the joints distribution of the asset returns belongs to an ambiguity set obtained by considering all joint distributions that satisfy a Kullback-Leibler divergence constraint from a joint Gaussian distribution, and obtained a closed-form formula for the worst-case optimal portfolio that exhibits a three-fund separation. In future investigations, it might be possible to extend this result under more general conditions on the investor's objective function and on the benchmark joint distribution.

Another aspect that could be further investigated is the following. It is well-known that when the joint distribution of the asset returns is Gaussian, for each mean-variance efficient portfolio there exists an expected utility investor for which this portfolio is optimal. See Sargent (1987) pages 154-155 for a proof. It would be interesting to see if this result still holds for the "mean-skewness-variance" efficient frontier we derived in Chapter 2, under the location-scale mixture assumption for the returns joint distribution.

In Chapter 3, we studied how the knowledge of a dependence measure can affect the (Range)

Value-at-Risk bounds, with respect to the case in which no dependence information is given, under the assumption that the risks' marginal distributions are known. In particular, we showed that a constraint on a dependence measure such as Pearson correlation or Spearman's rho may not be effective when it comes to reducing the worst-case value of a tail-risk measure for the sum of two risks. A similar conclusion holds for the sum of three or more risks when the dependence information is summarized using the average correlation. In the case of the sum of two risks, a natural progression of this work is to investigate if the best-possible VaR bounds given $d \notin [\delta_{min}, \delta_{max}]$ can still coincide with the bounds obtained in absence of any dependence information. Further work needs to be done to establish how to compute best-possible (R)VaR bounds for an arbitrary number of risks, assuming that the entire correlation matrix and the marginal distributions are given. This problem seems mathematically very challenging, but the output of this research line could give an important contribution to the debate about use of correlation matrices to set capital requirements for banks and insurance companies. In wider terms, the identification of the sources of partial dependence information that can be easily inferred from the available data but that are also able to reduce risks bounds requires further work. The results obtained in the present thesis suggest that summarising all dependence information in one number, e.g., the Spearman's rho for $n = 2$ or the average correlation for $n \geqslant 3$, may not be sufficient to have an improvement in term of risk bounds. As mentioned before, having a full correlation matrix could give better results, but it is possible to imagine other sources of dependence constraints. For instance, a positive tail dependence constraint, for example upper corner comotonicity, seems to be a promising starting point to obtain a reduction of the worst-case VaR.

A third possible extension of the results presented in Chapter 3 is to consider a risk aggregation problem in which the marginal distributions are not completely given. Note that the assumption of given marginal distributions is essentially omnipresent in the literature related to risk aggregation under dependence uncertainty. Nonetheless, practical observations suggests that this hypothesis can be challenged since also the estimation of the marginal distributions comes with a certain level of uncertainty. A promising approach that could be adopted in order to solve these above mentioned problems is to develop machine learning based numerical techniques. These numerical techniques have proved to be a valid choice when it comes to tackle complex risk aggregation problems for which a theoretical solution seems out of reach, as illustrated for example in Eckstein et al. (2020).

Chapter 4 studies the bounds on the net premium of life insurance contracts, when the estimated residual lifetime distribution is not fully trusted. Specifically, we provide a methodology for deriving the upper and lower bound on the net premium of life insurance contracts given that the residual lifetime distribution function lies in a certain neighbourhood (measured with the L^2 distance) of a benchmark distribution function. The results obtained in Chapter 4 highlight the convenience of using the L^2 distance to describe distributional uncertainty in this context, in that it allows to reformulate the premium bounds' problem as numerically tractable linear program. A numerical analysis illustrate how our results can be used also to solve robust expected utility maximization problems and to measure the impact of distributional uncertainty on the net premium under various financial scenarios. In Chapter 4 we consider life insurance contracts having fixed benefits, i.e., contracts in which the amount of benefits is stated at policy issue. Future research should therefore try to extend the results we obtained to the case of life insurance contracts with varying benefits. This extension could include, for example, those contracts having benefits that are adjusted to inflation or to the insurer investments performance, such as variable annuities. This would require to consider the ambiguity regarding the joint distribution of the financial and mortality components that determine the contract's net premium.

Chapter 4 focuses on the net premium bounds of a life insurance contract under the assumption that the net premium is computed according to the equivalence principle. This justifies the study of the worst- and best-case scenarios for $\mathbb{E}(g(K_x))$, assuming that the distribution of K_x is only partially specified. Nonetheless, it would be of theoretical and practical interest to verify if it is possible to obtain numerically tractable bounds for $\varrho(g(K_x))$, where ϱ is a given risk measure, such as VaR or TVaR. This could extend the applicability of the results obtained in Chapter 4 to the computation of the reserves that an insurer needs to allocate for an existing contract.

Finally, our analysis is focused on life insurance contracts, but it would be interesting to investigate if a similar approach can be adopted to study premium bounds for health insurance contracts. For instance, one can wonder if the use of the L^2 metric still allows to obtain numerically tractable bounds for the net premium of a disability insurance cover when the estimated transition probabilities among the active, disability and death states are not fully trusted.

Bibliography

American Academy of Actuaries, 2019. Model risk management.

Antolin, P., 2007. Longevity risk and private pensions. OECD Working Papers on Insurance and Private Pensions.

Barndorff-Nielsen, O., 1978. Hyperbolic distributions and distributions on hyperbolae. Scandinavian Journal of Statistics, 151–157.

Barndorff-Nielsen, O. E., 1997. Normal inverse Gaussian distributions and stochastic volatility modelling. Scandinavian Journal of Statistics 24 (1), 1–13.

Barrieu, P., Scandolo, G., 2015. Assessing financial model risk. European Journal of Operational Research 242 (2), 546–556.

Baucells, M., Heukamp, F., 2006. Stochastic dominance and cumulative prospect theory. Management Science 52 (9), 1409–1423.

Baviera, R., Bianchi, G., 2021. Model risk in mean-variance portfolio selection: an analytic solution to the worst-case approach. Journal of Global Optimization 81, 469–491.

Bellman, R. E., 2015. Adaptive control processes: a guided tour. Vol. 2045. Princeton university press.

Bernard, C., Bondarenko, O., Vanduffel, S., 2018a. Rearrangement algorithm and maximum entropy. Annals of Operations Research 261 (1), 107–134.

Bernard, C., Chen, J. S., Vanduffel, S., 2015. Rationalizing investors' choices. Journal of Mathematical Economics 59, 10–23.

Bernard, C., De Vecchi, C., Vanduffel, S., 2021. When do two- or three-fund separation theorems hold? Quantitative Finance 21 (11).

Bernard, C., De Vecchi, C., Vanduffel, S., 2022a. The impact of correlation on (range) value-at-risk. Working paper.

Bernard, C., De Vecchi, C., Vanduffel, S., 2022b. Robust assessment of life insurance products. Working paper.

Bernard, C., Denuit, M., Vanduffel, S., 2018b. Measuring portfolio risk under partial dependence information. Journal of Risk and Insurance 85 (3), 843–863.

Bernard, C., Jiang, X., Wang, R., 2013. Risk aggregation with dependence uncertainty. Insurance: Mathematics and Economics 54, 93–108.

Bernard, C., Kazzi, R., Vanduffel, S., 2020a. Range value-at-risk bounds for unimodal distributions under partial information. Insurance: Mathematics and Economics 94, 9–24.

Bernard, C., McLeish, D., 2016. Algorithms for finding copulas minimizing convex functions of sums. Asia-Pacific Journal of Operational Research 33 (5).

Bernard, C., Pesenti, S., Vanduffel, S., 2020b. Robust distortion risk measures. Working Paper.

Bernard, C., Rüschendorf, L., Vanduffel, S., 2017. Value-at-Risk bounds with variance constraints. Journal of Risk and Insurance 84 (3), 923–959.

Bernard, C., Vanduffel, S., 2015. A new approach to assessing model risk in high dimensions. Journal of Banking & Finance 58, 166 – 178.

Bertsimas, D., Lauprete, G. J., Samarov, A., 2004. Shortfall as a risk measure: properties, optimization and applications. Journal of Economic Dynamics and control 28 (7), 1353–1381.

Bignozzi, V., Puccetti, G., Rüschendorf, L., 2015. Reducing model risk via positive and negative dependence assumptions. Insurance: Mathematics and Economics 61, 17 – 26.

Birge, J. R., Chavez-Bedoya, L., 2016. Portfolio optimization under a generalized hyperbolic skewed t distribution and exponential utility. Quantitative Finance 16 (7), 1019–1036.

Birge, J. R., Chavez-Bedoya, L., 2020. Portfolio optimization under the generalized hyperbolic distribution: optimal allocation, performance and tail behavior. Quantitative Finance 0 (0), 1–21.

Birnbaum, M., 1997. Violations of monotonicity in judgment and decision making. Choice, decision, and measurement, 73–100.

Birnbaum, M., Navarrette, J., 1998. Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. Journal of Risk and Uncertainty 17, 49–78.

Black, R., Tsanakas, A., Smith, A. D., Beck, M. B., Maclugash, I. D., Grewal, J., Witts, L., Morjaria, N., Green, R. J., Lim, Z., et al., 2018. Model risk: illuminating the black box. British Actuarial Journal 23.

Blaesild, P., Jensen, J. L., 1981. Multivariate distributions of hyperbolic type. In: Statistical Distributions in Scientific Work. Springer, pp. 45–66.

Board of Governors of the Federal Reserve System, 2011. Supervisory guidance on model risk management.

Boudt, K., Jakobsons, E., Vanduffel, S., 2018. Block rearranging elements within matrix columns to minimize the variability of the row sums. 4OR 16 (1), 31–50.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.

Cairns, A. J., 2000. A discussion of parameter and model uncertainty in insurance. Insurance: Mathematics and Economics 27 (3), 313–330.

Cantelli, F. P., 1910. Intorno a un teorema fondamentale della teoria del rischio. Bollettino Attuari Italiano 24, 1–23.

Carr, P., Geman, H., Madan, D. B., Yor, M., 2002. The fine structure of asset returns: an empirical investigation. Journal of Business 75 (2), 305–332.

Cesari, R., Cremonini, D., 2003. Benchmarking, portfolio insurance and technical analysis: a monte carlo comparison of dynamic strategies of asset allocation. Journal of Economic Dynamics and Control 27 (6), 987–1011.

Chamberlain, G., 1983. A characterization of the distributions that imply mean-variance utility functions. Journal of Economic Theory 29 (1), 185–201.

Chernih, A., Henrard, L., Vanduffel, S., 2010. Reconciling credit correlations. Journal of Risk Model Validation 4 (2).

Christiansen, M. C., Denuit, M. M., Lazar, D., 2012. The Solvency II square-root formula for systematic biometric risk. Insurance: Mathematics and Economics 50 (2), 257–265.

Cont, R., Deguest, R., Scandolo, G., 2010. Robustness and sensitivity analysis of risk measurement procedures. Quantitative finance 10 (6), 593–606.

Cornilly, D., Puccetti, G., Rüschendorf, L., Vanduffel, S., 2022. Fair allocation of indivisible goods with minimum inequality or minimum envy. European Journal of Operational Research 297 (2), 741–752.

Cornilly, D., Rüschendorf, L., Vanduffel, S., 2018. Bounds for strictly concave distortion risk measures on moment spaces. Insurance: Mathematics and Economics 82 (2), 141–151.

d'Albis, H., Thibault, E., 2012. Optimal annuitization, uncertain survival probabilities, and maxmin preferences. Economics Letters 115 (2), 296–299.

De Giorgi, E., Hens, T., Mayer, J., 2011. A note on reward-risk portfolio selection and two-fund separation. Finance Research Letters 8 (2), 52–58.

De Schepper, A., Heijnen, B., 2010. How to estimate the value at risk under incomplete information. Journal of Computational and Applied Mathematics 233 (9), 2213–2226.

Dhaene, J., Vanduffel, S., Goovaerts, M., Kaas, R., Tang, Q., Vyncke, D., 2006. Risk measures and comonotonicity: A review. Stochastic Models 22, 573–606.

Dickson, D. C., Hardy, M. R., Waters, H. R., 2009. Actuarial mathematics for life contingent risks. Cambridge University Press.

Dietz, S., Niehörster, F., 2021. Pricing ambiguity in catastrophe risk insurance. The Geneva Risk and Insurance Review 46, 112 – 132.

Dietz, S., Walker, O., 2019. Ambiguity and insurance: capital requirements and premiums. Journal of Risk and Insurance 86 (1), 213 – 235.

Eberlein, E., Keller, U., 1995. Hyperbolic distributions in finance. Bernoulli 1 (3), 281–299.

Eckstein, S., Kupper, M., Pohl, M., 2020. Robust risk aggregation with neural networks. Mathematical Finance 30 (4), 1229–1272.

EIOPA, 2016. Discussion paper on the review of specific items in the Solvency II delegated regulation.

Embrechts, P., Höing, A., Juri, A., 2003. Using copulae to bound the Value-at-Risk for functions of dependent risks. Finance and Stochastics 7 (2), 145–167.

Embrechts, P., Höing, A., Puccetti, G., 2005. Worst var scenarios. Insurance: Mathematics and Economics 37 (1), 115–134, papers presented at the DeMoSTAFI Conference, Québec, 20-22 May 2004.

Embrechts, P., Liu, H., Wang, R., 07 2018. Quantile-based risk sharing. Operations Research 66.

Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. Risk management: value at risk and beyond 1, 176–223.

Embrechts, P., Puccetti, G., 10 2006. Bounds for Functions of Dependent Risks. Finance and Stochastics 10 (3), 341–352.

Embrechts, P., Puccetti, G., Rüschendorf, L., 2013. Model uncertainty and VaR aggregation. Journal of Banking & Finance 37 (8), 2750 – 2764.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., Beleraj, A., 2014. An academic response to basel 3.5. Risks 2 (1), 25–48.

Escobar, D., Pflug, G. C., 2020. The distortion principle for insurance pricing: properties, identification and robustness. Annals of Operations Research 292 (3), 771 – 794.

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. Journal of Mathematical Economics 18 (2), 141–153.

Haberman, S., Khalaf-Allah, M., Verrall, R., 2011. Entropy, longevity and the cost of annuities. Insurance: Mathematics and Economics 48 (2), 197–204.

Hofert, M., Memartoluie, A., Saunders, D., Wirjanto, T., 2017. Improved algorithms for computing worst value-at-risk. Statistics & Risk Modeling 34 (1-2), 13–31.

Huang, C.-f., Litzenberger, R. H., 1988. Foundations for Financial Economics. North-Holland.

Hürliman, W., 1998. Extremal moment methods and stochastic orders. Application in Actuarial Science. Monograph manuscript (available from the author).

Hürlimann, W., 2002. Analytical bounds for two value-at-risk functionals. ASTIN Bulletin 32 (2), 235–265.

Jansen, K., Haezendonck, J., Goovaerts, M. J., 1986. Upper bounds on stop-loss premiums in case of known moments up to the fourth order. Insurance: Mathematics and Economics 5 (4), 315–334.

Kaas, R., 1993. How to (and how not to) compute stop-loss premiums in practice. Insurance: Mathematics and Economics 13 (3), 241–254.

Kaas, R., Goovaerts, M. J., 1986. Extremal values of stop-loss premiums under moment constraints. Insurance: Mathematics and Economics 5 (4), 279–283.

Kaas, R., Laeven, R. J., Nelsen, R. B., 2009. Worst VaR scenarios with given marginals and measures of association. Insurance: Mathematics and Economics 44 (2), 146 – 158.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–291.

Keilson, J., Gerber, H., 1971. Some results for discrete unimodality. Journal of the American Statistical Association 66 (334), 386–389.

Kim, J. H., Kim, W. C., Fabozzi, F. J., 2014. Recent developments in robust portfolios with a worst-case approach. Journal of Optimization Theory and Applications 161, 103–121.

Küchler, U., Neumann, K., Sørensen, M., Streller, A., 1999. Stock returns and hyperbolic distributions. Mathematical and Computer Modelling 29 (10-12), 1–15.

Kwak, M., Pirvu, T. A., 2018. Cumulative prospect theory with generalized hyperbolic skewed t distribution. SIAM Journal on Financial Mathematics 9 (1), 54–89.

Levy, H., Levy, M., 2004. Prospect theory and mean-variance analysis. Review of Financial Studies 17 (4), 1015–1041.

Levy, H., Markowitz, H., 1979. Approximating expected utility by a function of mean and variance. American Economic Review 69, 308–317.

Li, L., Shao, H., Wang, R., Yang, J., 1 2018. Worst-case Range Value-at-Risk with partial information. SIAM Journal on Financial Mathematics 09, 146 – 158.

López-Díaz, M., Sordo, M. A., Suárez-Llorens, A., 2012. On the lp-metric between a probability distribution and its distortion. Insurance: Mathematics and Economics 51 (2), 257–264.

Makarov, G. D., 1981. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. Theory of Probability & Its Applications 26, 803–806.

Mao, T., Wang, B., Wang, R., 2019. Sums of standard uniform random variables. Journal of Applied Probability 56 (3), 918–936.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

Markowitz, H., 2014. Mean–variance approximations to expected utility. European Journal of Operational Research 234 (2), 346–355.

May, J.-F., Scherer, M., 2014. Financial Engineering with Copulas Explained. Palgrave MacMillan, London.

McNeil, A. J., Frey, R., Embrechts, P., 2010. Quantitative risk management: concepts, techniques, and tools. Princeton university press.

Mencía, J., Sentana, E., 2009. Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. Journal of Econometrics 153 (3), 105–121.

Nelsen, R. B., 2010. An Introduction to Copulas. Springer Publishing Company, Incorporated.

Nelsen, R. B., Molina, J. J. Q., Lallena, J. A. R., Úbeda Flores, M., 2004. Best-possible bounds on sets of bivariate distribution functions. Journal of Multivariate Analysis 90 (2), 348 – 358.

Nelsen, R. B., Úbeda Flores, M., 2005. A comparison of bounds on sets of joint distribution functions derived from various measures of association. Communications in Statistics - Theory and Methods 33 (10), 2299–2305.

Olivieri, A., 2001. Uncertainty in mortality projections: an actuarial perspective. Insurance: Mathematics and Economics 29 (2), 231–245.

Olivieri, A., Pitacco, E., 2011. Introduction to insurance mathematics: technical and financial features of risk transfers. Springer.

Owen, J., Rabinovitch, R., 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. Journal of Finance 38 (3), 745–752.

Pichler, A., 2014. Insurance pricing under ambiguity. European Actuarial Journal 4, 335–364.

Pirvu, T. A., Schulze, K., 2012. Multi-stock portfolio optimization under prospect theory. Mathematics and Financial Economics 6 (4), 337–362.

Pitacco, E., Denuit, M., Haberman, S., Olivieri, A., 2009. Modeling Longevity Dynamics for Pensions and Annuity Business. Oxford University Press.

Puccetti, G., Rüschendorf, L., 2012. Computation of sharp bounds on the distribution of a function of dependent risks. Journal of Computational and Applied Mathematics 236 (7), 1833 – 1840.

Puccetti, G., Rüschendorf, L., 2013a. Asymptotic equivalence of conservative var-and es-based capital charges. Journal of Risk 16 (3), 3–22.

Puccetti, G., Rüschendorf, L., 2013b. Sharp bounds for sums of dependent risks. Journal of Applied Probability 50 (1), 42–53.

Puccetti, G., Wang, B., Wang, R., 2013. Complete mixability and asymptotic equivalence of worst-possible var and es estimates. Insurance: Mathematics and Economics 53 (3), 821–828.

Puccetti, G., Wang, R., 2015a. Detecting complete and joint mixability. Journal of Computational and Applied Mathematics 280, 174–187.

Puccetti, G., Wang, R., 2015b. Extremal dependence concepts. Statistical Science 30 (4), 485–517.

Quiggin, J., 1993. Generalized Expected Utility Theory - The Rank-Dependent Model. Kluwer Academic Publishers.

Rabitti, G., Borgonovo, E., 2020. Is mortality or interest rate the most important risk in annuity models? a comparison of sensitivity analysis methods. Insurance: Mathematics and Economics 95, 48–58.

Rachev, S., Stoyanov, S., Fabozzi, F., 2008. Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measu res. John Wiley & Sons, Hoboken, New Jersey.

Ryan, M. J., 2006. Risk aversion in RDEU. Journal of Mathematical Economics 42 (6), 675–697.

Rüschendorf, L., 1982. Random variables with maximum sums. Advances in Applied Probability 14 (3), 623–632.

Salmon, F., 2009. Recipe for disaster: The formula that killed wall street.

Sargent, T. J., 1987. Macroeconomic theory. Second edition ,Academic Press, Boston.

Shaked, M., Shanthikumar, J. G., 2007. Stochastic orders. Springer Science & Business Media.

The Deloitte Center for Regulatory Strategy, 2018. Model risk management, building supervisory confidence.

Tobin, J., 1958. Liquidity preference as behavior towards risk. Review of Economic Studies, 65–86.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5 (4), 297–323.

Vanduffel, S., Yao, J., 2017. A stein type lemma for the multivariate generalized hyperbolic distribution. European Journal of Operational Research 261 (2), 606–612.

Wang, B., Wang, R., 2011. The complete mixability and convex minimization problems with monotone marginal densities. Journal of Multivariate Analysis 102 (10), 1344 – 1360.

Wang, R., Peng, L., Yang, J., 2013. Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. Finance and Stochastics 17, 395–417.

Wang, R., Zikitis, R., 2020b. An axiomatic fuondation of expected shortfall. Management Science.

Xidonas, P., Steuer, R., Hassapis, C., 2020. Robust portfolio optimization: a categorized bibliographic review. Annals of Operations Research 292, 533–552.

Yaari, M., 1987. The dual theory of choice under risk. Econometrica 55 (1), 95–115.

Yaari, M. E., 1965. Uncertain lifetime, life insurance, and the theory of the consumer. Review of Economic Studies 32 (2), 137–150.

Yang, J., Zhuang, W., Hu, T., 2014. Lp-metric under the location-independent risk ordering of random variables. Insurance: Mathematics and Economics 59, 321–324.

Young, V., 2014. Premium Principles. Wiley StatsRef: Statistics Reference Online.