Bignozzi, Valeria (2012). Contributions to solvency risk measurement. (Unpublished Doctoral thesis, City University London)

CITY UNIVERSITY LONDON

City Research Online

Cass Business School
CITY UNIVERSITY LONDON

Faculty of Actuarial Science and Insurance

Contributions to solvency risk measurement

A thesis submitted for the degree of Doctor in Philosophy

September 2012

Supervisors

Dr. Andreas Tsanakas

Dr. Laura Ballotta

PhD Candidate

Valeria Bignozzi

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my greatest gratitude to the people who have helped and supported me throughout my PhD.

First and foremost I am deeply grateful to my supervisor, Dr Andreas Tsanakas, for his invaluable support. Without his constructive comments, patience and knowledge this thesis would have not been possible. I honestly could have not wished for a better supervisor.

I would also like to thank my second supervisor, Dr Laura Ballotta, for her useful comments and help. I am very grateful to Dr Enrico Biffis, for his precious advices and encouragement throughout the years. Great appreciations go to the Actuarial Profession and Chorafas Foundation for their important financial help during my last year of the PhD. I offer my regards to academic and research staff at Cass Business School, City University London. Special thanks are due to Ms. Malla Pratt and Mr. Abdul Momin for their administrative support.

I wish to thank my parents, Luisa and Stefano, and my sister, Lorenza, for all the emotional support, understanding, and caring they provided. A special thank goes to my partner and future husband, Leonardo, for being there every time I needed him. For teaching me determination (or at least for trying), for helping me get through the difficult times but especially for amusing me everyday. I would also like to thank my colleagues and friends Ioulia and Eilnaz for sharing with me all the happy, and more important, all the difficult days of this PhD. My warmest appreciation goes to my old time friends Riccardo, Raffaele and Camilla, I was so lucky to meet them.

A special thank to all my Londoner friends, Daria, Dominika, Alessandro, Sabella, Silvia, Silvano, Vanessa, Stephane, Florence, Adrien and Vincent that made these years simply amazing.

At last but not the least I want to thank my dog Big, for making me laugh everyday, for his British humor, his unconditional love and beauty. Simply The Best Friend. To him I dedicate this thesis.

Declaration

I grant powers of discretion to the university Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgment.

Abstract

The thesis focuses on risk measures used to calculate solvency capital requirements. It consists of three independent papers.

The first paper (Chapter 2) investigates time-consistency, the relation that should hold across risk measurements of the same financial position at different time points. Sufficient conditions are provided for coherent risk measures, in order to satisfy the requirements of acceptance-, rejection- and sequential consistency. It is shown that risk measures used in practice usually do not satisfy these requirements. Hence a method is provided to systematically construct sequentially consistent risk measures. It is also emphasized that current approaches to dynamic risk measurement do not consider that risk measures at different time points have different arguments. Here we briefly discuss this new setting highlighting that the notions of time consistency presented in the literature need to be reinterpreted.

The second and third papers (Chapters 3 and 4) consider respectively the risk arising from parameter and model mis-specification due to estimation from a limited amount of available data. This risk may have a substantial impact on risk measures used to quantify solvency capital requirements. We introduce a new method to quantify this impact measured as the additional capital needed to allow for randomness in the data sample used for the estimation procedure. This level of capital we call residual estimation risk.

In the second paper, for parameter uncertainty we prove the effectiveness of three approaches for reducing residual estimation risk in the case of location-scale families. These are based on (a) raising the capital requirement by adjusting the risk measure, (b) Bayesian predictive distributions under probability-matching priors and (c) residual risk estimation via parametric bootstrap. Risk measures satisfying standard properties are used, for example the popular TVaR. For more general distributions only (a) and (b) are investigated and a truncated version of TVaR is used. Numerical results obtained via Monte-Carlo simulation demonstrate that the proposed

methods perform well.

In the third paper (Chapter 4), we compare the effectiveness of four different approaches to estimate capital requirements in the presence of model uncertainty. For a given set of candidate models the model posterior weights can be obtained via a Bayesian approach. Then we consider approaches based on: (a) worst case scenario, (b) highest model posterior, (c) averaging the capital under each model according to the model posterior weights and (d) determining the predictive distribution of the financial loss and using it to calculate the capital. It is shown that all these methods are very sensitive to the set of candidate models specified. If this has been carefully selected (for instance via expert judgement) the approach based on the highest posterior performs slightly better than the others. Alternatively, if there is poor prior information on the model set the effectiveness of all these approaches decreases substantially. In particular, the worst case approach has a very low performance. It also emerges that mis-specifying the model by using distributions that are more heavy-tailed than the one generating the data, may reduce the capital and thus it is not a conservative approach.

Chapter 1

Introduction

1.1 Literature review

The correct quantification of risks faced by financial or insurance firms is a central task for both investors and regulators. *Risk* can be defined as the possibility that future events may cause adverse effects. In particular, financial risks can be modelled via random variables that assume different values accordingly to possible future events. Hence the last decades have seen an increasing interest in approaching risk assessment from a quantitative point of view.

The essential technical tools to quantify risks are *risk measures*. A risk measure is a functional that assigns to every financial position a real number. Such a number summarizes the information relative to the future monetary outcomes of the financial position and their probability. It provides insight on the level of risk of the position and suggests, according to the preferences or constraints of an investor, whether it is acceptable or not. It has been stressed by Artzner et al. (1999) that although gathering all the information of a position in a single number may be simplistic, it is consistent with the binary choice that investors and regulators have to make: accept the position or reject it.

Risk measures are very versatile and have been object of research in sev-

eral fields. Here we give a brief overview of the main areas of application and research.

Apart from the work of Markowitz (1951), where the variance of a portfolio was used as first attempt to measure its desirability, risk measures find their roots in the actuarial theory of premium calculation more than 40 years ago. In such a context, the outcome of the measurement represents the price that an insurer should charge for bearing the risk of the insured claim (or *financial loss*). Premium principles, as risk measures are called in this literature, are constructed in such a way to rank financial losses consistently with stochastic orders and can often be associated to expected utility functions. For rigorous treatments of risk measures, their properties, and stochastic orders, we refer to Bühlmann (1970), Gerber (1979), Goovaerts et al. (1984), Wang et al. (1997) and more recently to Denuit et al. (2005).

In the last two decades new applications of risk measures have emerged. During the period 1993-1996 some substantial derivative-based losses occurred (such as Orange County and Metallgesellschaft), which emphasized the need for a rigorous technique to manage and control market risk. Hence, in 1994, JP Morgan proposed the use of Value-at-Risk (VaR) to assess and interpret easily the riskiness of financial positions. Since then, the use of this risk measure in banking has become an imperative. For an in-depth treatment of VaR we refer to Jorion (1996) and Duffie and Pan (1997).

The rapidly growing financial market, together with its deregulation and globalization led also regulators to require a more detailed and systematic quantification of the risk. The Basel accord in 1988 sets the first step towards internationally coordinated regulatory capital requirements for the banks and in 1996 an amendment to Basel I allows the use of VaR-based internal models for measuring market risk. Nowadays, the projects Basel III and Solvency II systematically require financial and insurance firms to use risk measures to calculate solvency capital requirements. When used to determine capital requirement, the value of the risk measure represents the minimal amount of capital that a company is asked to hold as a buffer against unexpected

losses.

From an academic point of view, risk measures have become of primary interest in financial mathematics since the seminal paper by Artzner et al. (1999). They proposed four properties that a risk measure should satisfy, defining in this way the class of *coherent risk measures*. These requirements were then relaxed in various directions, in particular two of the axioms (sub-additivity and positive homogeneity) were substituted with a weaker one (convexity) by Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002), obtaining the class of *convex risk measures*. Detlefsen and Scandolo (2005) extended the above-mentioned theory to a dynamic context where new available information is used to update the risk assessment.

Risk measures find a natural application also in option pricing in incomplete markets. In the standard equivalent martingale measures approach, an entire interval of no arbitrage prices is available. Risk measures can be used to narrow this interval, for instance via the "good-deal" approach (Černý and Hodges, 2000) or to pick one price. For instance Föllmer and Schweizer (1991) propose to use the price corresponding to the minimal martingale measure, Frittelli (2002) to the minimal entropy martingale measure and Bellini and Frittelli (2002) to the mini-max martingale measure.

It is worth noting that several risk measures used in practice are defined as functionals on the probability distribution of the financial position to assess. Since this distribution is generally unknown and estimated from past available data, the calculation of risk measures becomes a statistical issue. Several approaches, such as Historical Simulation and Extreme Value theory have been developed to compute risk measures in such a context. For an overview of the statistical methods used and the literature we refer to McNeil et al. (2005).

It is clear then that risk measures are very versatile tools, in continuous development and object of research in several fields. Also the literature is quite wide and together with the notation used, changes according to the field of application. For instance the convention in financial mathematics

and banking is to represent with a random variable X a financial position, for instance the Profit and Loss of a company or the outcome of a portfolio of stocks. A positive value of X corresponds to a gain, while $X \leq 0$ is a loss. On the contrary in insurance the convention is to work with loss random variables, $Y = -X$, thus all the definitions and properties have to be adapted. In the following sections and Chapter 2, consistently with the literature in financial mathematics we work with financial positions. Afterwards, in Chapters 3 and 4 we will switch to loss random variables. This choice has been made in order to remain consistent with the main literature on the topic treated in each chapter. The main definitions and properties are restated in the new notation within each chapter.

1.2 Definitions and properties

1.2.1 Risk measures

In order to work with risk measures in mathematical terms, we need to introduce a sample space Ω that represents the set of all possible states of nature at a certain future date. A financial position is thus defined specifying the monetary outcome for every possible scenario, i.e. it is a real-valued function taking value on $\mathbb{R}$. We denote the set of all financial positions with $\mathcal{X}$. A risk measure is a functional $\rho : \mathcal{X} \longrightarrow \mathbb{R}$ that assigns to every financial position a real value.

Traditional examples of risk measures used in actuarial science are the expected premium principle:

$$\rho(X) := (1 + \alpha)E[-X] \quad \text{for } X \in \mathcal{X} \quad \text{and } \alpha \geq 0, \tag{1.1}$$

and the variance premium principle:

$$\rho(X) := E[-X] + \alpha Var(-X) \quad \text{for } X \in \mathcal{X} \quad \text{and } \alpha \geq 0. \tag{1.2}$$

For more examples we refer to Bühlmann (1970) and Goovaerts et al. (1984). In both cases the risk measure considers the expected loss of the position ($E[-X]$) and adds to it an extra load that acts as a buffer against unexpected losses. The variance premium principle penalizes positions with a high variance. While this seems reasonable, it does not distinguish between the signs of the variation.

A risk measure that focuses more on negative outcomes is VaR. VaR_p (level of confidence p) is the most widespread risk measure in banking and regulatory contexts. For instance, Basel II requires financial institution to measure market risk using VaR with a time horizon of 10 days and a probability level of 99%, see (Basel Committee on Banking Supervision, 2009). Solvency II, proposed as capital requirement for insurance companies, VaR at level of 99.5% with a time horizon of 1 year (European Insurance and Occupational Pensions Authority, 2009). This means that an insurance company is considered reliable if the difference between its net assets and liabilities is non-negative with a probability of at least 0.995 over a time period of 1 year. In mathematical terms, given a measurable space $(\Omega, \mathcal{F})$ and a financial position X, $\text{VaR}_p(X)$ is given by:

$$VaR_p := \inf\{m \in \mathbb{R} \ : \ \mathbb{P}(X + m \geq 0) \geq p\} \tag{1.3}$$

where $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. VaR_p thus requires companies to hold enough capital to cover their losses with probability p.

This risk measure has the advantage of being relatively easy to evaluate and to understand. This made it very popular from the practitioner point of view. However, it has some important deficiencies that should be considered. First, it does not warn about the magnitude of losses occurring in the remaining $1 - p$ probability. For instance consider two financial positions X, Y, such that $Y = \min(X, d)$ where d corresponds to the $p-$quantile of X. Such random variables are assessed in the same way by VaR although X may cause much higher losses that Y.

Another shortcoming is that VaR may penalize diversification, that is pooling together different positions does not necessarily reduce the final risk of the portfolio.

To overcome these issues Artzner et al. (1999) proposed a set of axioms that a risk measure should satisfy. The starting point is to decide what has to be considered too risky and what not. For this purpose an *acceptance set* $\mathcal{A}$ is defined as:

$$\mathcal{A} := \{X \in \mathcal{X} : X \text{ is considered an acceptable risk by the regulator or investor}\}.$$

Acceptance sets clearly depend on measurement scope and will vary with the context of application. Artzner et al. (1999) worked in a regulatory framework (this work was done at the same time of the regulatory project Basel II). Here, given an acceptance set $\mathcal{A}$, a risk measure is interpreted as the minimal amount of capital that should be safely invested and added to the financial position X in order to make it acceptable:

$$\rho(X) = \inf\{m \in \mathbb{R} \ : \ X + m \in \mathcal{A}\}. \tag{1.4}$$

Artzner et al. (1999) took into account also the return r arising from investing money in a safe instrument, in this case definition (1.4) becomes

$$\rho(X) = \inf\{m \in \mathbb{R} \ : \ X + m \cdot (1 + r) \in \mathcal{A}\}. \tag{1.5}$$

Without loss of generality, and consistently with most of the papers on this topic, we assume that our payoff corresponds to the already discounted financial position, so that (1.4) is used from now on.

1.2.2 Coherent risk measures

Artzner et al. (1999) proposed a set of axioms that risk measures should satisfy. For every $X, Y \in \mathcal{X}$

(1) *Monotonicity.* If $X \leq Y$, then $\rho(X) \geq \rho(Y)$;

(2) *Translation invariance.* If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$;

(3) *Subadditivity.* $\rho(X + Y) \leq \rho(X) + \rho(Y)$;

(4) *Positive homogeneity.* If $\lambda \geq 0$ then $\rho(\lambda X) = \lambda \rho(X)$.

A risk measure satisfying axioms (1), (2), (3) and (4) is called a *coherent risk measure.*

Monotonicity implies that if the position's payoff increases in every state of nature, then its riskiness should decrease. Translation invariance suggests that adding safe capital to a financial position, decreases the riskiness of the position by the same amount. In particular, it implies

$$\rho(X + \rho(X)) = \rho(X) - \rho(X) = 0. \tag{1.6}$$

This property again suggests the idea of risk measure as capital requirement, as $\rho(X)$ represents the amount of money that, added to the financial position X, makes it marginally acceptable.

The subadditivity axiom requires that adding two positions together should decrease the total risk. Positive homogeneity implies that the risk of a payoff increases linearly with the size of the investment. It also implies the *normalization* property, that is $\rho(0) = 0$, which is usually considered a natural condition to require. Together with the translation invariance and monotonicity requirements, normalization allows the following interpretation of the risk measure for fixed capital

$$\rho(0 + m) = -m \qquad \text{for } m \in \mathbb{R}. \tag{1.7}$$

Holding some safely invested capital m is of course not risky and m is the maximal amount of capital that can be withdrawn maintaining the position acceptable. In most of the situations normalization is required even if positive homogeneity does not hold.

Other important properties and definitions for risk measures are the following. For every X,Y in $\mathcal{X}$

(5) *Comonotone additivity.* If X and Y are *comonotone*, ie

$$(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0 \qquad \forall \omega_1, \omega_2 \in \Omega$$

then

$$\rho(X + Y) = \rho(X) + \rho(Y);$$

(6) *Continuity from above.* If $X_n \searrow X$, then $\rho(X_n) \nearrow \rho(X)$;

(7) *Continuity from below.* If $X_n \nearrow X$, then $\rho(X_n) \searrow \rho(X)$.

Comonotone additivity implies that adding comonotone random variables does not reduce the amount of capital required. Axioms (6) and (7) are technical conditions that will be used in the following sections for the representation theorems.

One of the main and most important results obtained in risk measure theory is that all the coherent risk measures admit a representation in terms of generalized scenarios. There are different versions of this theorem; we provide here the one by Delbaen (2002).

Theorem 1.2.1. *All coherent risk measures that are continuous from above, admit a representation in terms of generalized scenarios*

$$\rho(X) = \sup_{P \in \mathcal{P}} \mathbb{E}^P[-X] \tag{1.8}$$

where the set $\mathcal{P}$ is a set of probability measures on the measurable space $(\Omega, \mathcal{F})$. If the risk measure is also continuous from below, then the supremum is actually a maximum:

$$\rho(X) = \max_{P \in \mathcal{P}} \mathbb{E}^P[-X] \tag{1.9}$$

Also the contrary holds: every risk measure defined as (1.2) is coherent. This representation makes clear the idea behind coherent risk measures. Given a financial position X, its expected loss ($E^P[-X]$) is specified accordingly to different possible scenarios $P \in \mathcal{P}$. The supremum is then taken so that the capital held, $\rho(X)$, corresponds at least to the expected loss in the worst case scenario.

A representation in terms of probability measures without requiring continuity from above is possible if we work in a finite framework (i.e. Ω is a finite set) as Artzner et al. (1999) proved.

1.2.3 Distortion risk measures

Distortion risk measures represent an important class risk measures, that is obtained via *Choquet integrals*. They have been developed by Denneberg (1994b) and Wang et al. (1997), mostly within an insurance context. Here, we consider distortion risk measures that are also coherent. This class of risk measures corresponds to the *spectral risk measures*, introduced by Acerbi (2002).

Let us consider a measurable space $(\Omega, \mathcal{F})$ and a reference probability measure $\mathbb{P}$. We also assume that $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$. We start with the definition of Choquet integral for a bounded measurable function $X \in \mathcal{X}$. Given a set function or capacity $v : \mathcal{F} \longrightarrow [0, 1]$ that satisfies

(8) *Normalization.* $v(\emptyset) = 0$, $v(\Omega) = 1$;

(9) *Monotonicity.* $v(A) \leq v(B)$ for $A \subseteq B$.

the Choquet integral of $X \in \mathcal{X}$ with respect to v is defined as

$$\int X dv = \int_{-\infty}^{0} (v(X > t) - 1)dt + \int_{0}^{\infty} v(X > t)dt. \tag{1.10}$$

This integral is generally not linear, but it coincides with the Lebesgue integral when the function v is a probability measure. It satisfies the following properties

(10) *Positive homogeneity.* $\int \lambda X dv = \lambda \int X dv$ for $\lambda \geq 0$

(11) *Translation invariance* $\int (X + m)\, dv = \int X dv + m$ for $m \in \mathbb{R}$.

Now consider a *distortion function* $g : [0, 1] \to [0, 1]$ increasing and concave, with $g(0) = 0$ and $g(1) = 1$. The set function $v(A) = g(\mathbb{P}(A))$ is called *distortion* of the probability measure $\mathbb{P}$. It satisfies normalization, monotonicity and

(12) *Submodularity.* $v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$.

With these properties, for a given financial position $X \in \mathcal{X}$, the Choquet integral

$$\rho(X) = \int -X dv = -\int_{-\infty}^{0} (1 - v(-X > t))dt + \int_{0}^{\infty} v(-X > t)dt \quad (1.11)$$

defines a coherent risk measure.

It is possible to prove that a risk measure admits a representation as Choquet integral with respect to a capacity v if and only if it is a coherent risk measure satisfying comonotone additivity. For more details see Föllmer and Schied (2004) and Kusuoka (2001).

Furthermore, this risk measure admits a representation in terms of generalized scenarios

$$\rho(X) = \int (-X) dv = \sup_{Q \in \tilde{Q}} \mathbb{E}^{Q}[-X] \quad (1.12)$$

where

$$\tilde{Q} = \{Q \ll \mathbb{P} \ : \ Q(A) \leq v(\mathbb{P}(A)) \quad \forall A \in \mathcal{F}\}$$

is called *the core* of the convex distortion of $\mathbb{P}$. Here, $Q \ll \mathbb{P}$ means that Q is *absolutely continuous* with respect to $\mathbb{P}$ on the $\sigma-$ algebra $\mathcal{F}$, that is:

$$\mathbb{P}(A) = 0 \quad \Rightarrow \quad Q(A) = 0 \quad \forall A \in \mathcal{F}.$$

Alternatively, we can say that Q is absolutely continuous with respect to $\mathbb{P}$

if and only if, it exists its Radon-Nikodym derivative $\frac{dQ}{d\mathbb{P}}$, such that:

$$\int F dQ = \int F \frac{dQ}{d\mathbb{P}} d\mathbb{P} \text{ for any } \mathcal{F} - \text{measurable function } F \geq 0.$$

For details on distortion risk measures see Carlier and Dana (2003).

Example 1. A standard example of distortion risk measure is Tail Value at Risk (TVaR). It arises as a natural extension of VaR in order to avoid its deficiencies. Given a financial position X, TVaR_p is defined as:

$$\text{TVaR}_p(X) := \frac{1}{p} \int_0^p \text{VaR}_\gamma(X) d\gamma.$$

This is an average of VaR over the percentiles from 0 to p. In this way the whole tail risk is considered and included in the assessment. TVaR is a coherent risk measure, in particular it satisfies the axiom of subadditivity that VaR fails and also satisfies comonotone additivity. Its representation in terms of generalized scenario is obtained as

$$TVaR_p(X) := \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q[-X] \tag{1.13}$$

where

$$\mathcal{Q} := \{Q \ll \mathbb{P} \ : \ \frac{dQ}{d\mathbb{P}} \leq p^{-1}\}$$

The set $\mathcal{Q}$ has a specific meaning: given a reference probability measure $\mathbb{P}$, we aim at distorting it in order to emphasize negative events and have a more conservative assessment of the risk. At the same time we want to avoid probability measures that differ to much from the physical one, thus presenting a completely unrealistic scenario. To solve this issue, a bound on the Radon-Nikodym derivative is imposed so that the probability measure $\mathcal{Q}$ is not too "distant" from $\mathbb{P}$. TVaR_p arises as a distortion risk measure, considering the distortion function $g(s) := \min(p^{-1}s, 1)$. It has been proposed by several authors independently and can be found in the literature under different names such as Expected Shortfall (ES), Tail Conditional Expectation

(TCE), Conditional Value at Risk (CVaR). For an overview of the different definitions and how they are related we refer to Acerbi and Tasche (2002).

1.2.4 Convex risk measures

We now discuss one of the main extensions made to coherent risk measure theory. Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002) in independent papers proposed to drop the positive homogeneity and subadditivity axioms and introduced the following. For any X,Y in $\mathcal{X}$ and $0 \leq \delta \geq 1$

(13) *Convexity.* $\rho(\delta X + (1-\delta)Y) \leq \delta\rho(X) + (1-\delta)\rho(Y)$.

Risk measures satisfying axioms (1), (2) and (13) are called *convex risk measures*. Risk measures satisfying convexity and positive homogeneity requirements satisfy also subadditivity. It follows that coherent risk measures are a subclass of convex ones. Föllmer and Schied (2002) proved that also in the context of convex risk measures a representation in terms of generalized scenarios is possible. For the rest of the chapter, we will assume that a reference probability measure $\mathbb{P}$ is given and we will only consider risk measures that satisfy

$$\rho(X) = \rho(Y) \qquad \text{if } X = Y \quad \mathbb{P}-\text{a.s.} \tag{1.14}$$

meaning that financial positions that are almost surely equivalent are assessed in the same way. With assumption (1.14) the set of risks will be identified by the space $\mathcal{X} = L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$ of all bounded real valued functions. The following notation is used

$$\mathcal{M}_1 := \{Q \text{ probability measures on } (\Omega, \mathcal{F})\} \tag{1.15}$$

$$\mathcal{M}_1(\mathbb{P}) := \{Q \text{ probability measures on } (\Omega, \mathcal{F}) \ : \ Q \ll \mathbb{P}\} \tag{1.16}$$

Proposition 1.2.2. *A convex risk measure $\rho(X)$ can be represented as*

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(\mathbb{P})} (\mathbb{E}^Q[-X] - \alpha_{min}(Q)) \tag{1.17}$$

if and only if it is continuous from above, i.e.

$$\text{if} \quad X_n \searrow X \ \mathbb{P}-a.s. \quad \text{then} \quad \rho(X_n) \nearrow \rho(X).$$

Here $\alpha_{min}(Q)$ is the minimal penalty function representing ρ and corresponds to

$$\alpha_{min}(Q) := \sup_{X \in \mathcal{X}} (\mathbb{E}^Q[-X] - \rho(X)) = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}^Q[-X]$$

For the coherent case, this result leads to a representation where the supremum is actually attained.

Proposition 1.2.3. *With the assumptions of the previous theorem, a coherent risk measure admits the following representation*

$$\rho(X) = \max_{Q \in \mathcal{Q}_{max}} \mathbb{E}^Q[-X] \tag{1.18}$$

for the convex set

$$\mathcal{Q}_{max} = \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ \alpha_{min} = 0\}$$

where the minimal penalty function of a coherent risk measure can only assume values 0 or $+\infty$. The set $\mathcal{Q}_{max}$ is the largest one that permits such a representation.

Assuming that no physical probability measure is given, and denoting with $\mathcal{M}_{1,f}$ the set of all finitely additive normalized set functions on $(\Omega, \mathcal{F})$, a general result still holds

Theorem 1.2.4. *Any convex risk measure $\rho(X)$ on $\mathcal{X}$ has the form*

$$\rho(X) = \max_{Q \in \mathcal{M}_{1,f}} (\mathbb{E}^Q[-X] - \alpha_{min}(Q)) \tag{1.19}$$

where

$$\alpha_{min}(Q) := \sup_{X \in \mathcal{A}_\rho} \mathbb{E}^Q[-X] \quad \textit{for } Q \in \mathcal{M}_{1,f}.$$

Even though we will not investigate this general case, it is worth noticing that if the risk measure $\rho(X)$ is continuous from below, then the penalty function α is concentred on the set $\mathcal{M}$ of probability measures on $(\Omega, \mathcal{F})$. Therefore, not only ρ admits a representation in terms of probability measures, but it has the additional property that the supremum is actually attained. Föllmer and Schied (2002) extended further this topic in a topological setting. We will not treat this approach here, as it is far from the focus of our research.

Example 2. One of the best known examples of convex risk measure is the entropic one. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an acceptance set defined as

$$\mathcal{A} := \{X \in \mathcal{X} \ : \ \mathbb{E}[e^{-\beta X}] \leq 1\}. \tag{1.20}$$

The corresponding risk measure is

$$\rho(X) := \inf\{m \in \mathbb{R} \ : \ X+m \in \mathcal{A}\} = \inf\{m \in \mathbb{R} \ : \ \mathbb{E}[e^{-\beta(X+m)}] \leq 1\} \tag{1.21}$$

with some calculation we have

$$\rho(X) = \inf\{m \in \mathbb{R} \ : \ \mathbb{E}[e^{-\beta X}] \leq e^{\beta m}\}$$

$$= \inf\{m \in \mathbb{R} \ : \ \log(\mathbb{E}[e^{-\beta X}]) \leq \beta m\}$$

$$= \frac{1}{\beta} \cdot \log \mathbb{E}[e^{-\beta X}]$$

In this case it is possible to evaluate exactly the minimal penalty function

$\alpha_{\min}(Q)$. With few steps, we obtain

$$\begin{aligned}
\alpha_{\min}(Q) &:= \sup_{X\in\mathcal{X}}\{\mathbb{E}^Q[-X]-\rho(X)\} \\
&= \sup_{X\in\mathcal{X}}\left\{\mathbb{E}^Q[-X]-\frac{1}{\beta}\left[\ \log(\mathbb{E}[e^{-\beta(X)}])\ \right]\right\} \\
&= \frac{1}{\beta}\left[\ \sup_{X\in\mathcal{X}}\{\mathbb{E}^Q[-\beta X]-\log(\mathbb{E}[e^{-\beta X}])\}\ \right] \\
&= \frac{1}{\beta}\left[\ \sup_{Z\in\mathcal{X}}\{\mathbb{E}^Q[Z]-\log(\mathbb{E}[e^{Z}])\}\ \right] \\
&= \frac{1}{\beta}\left[\ H(Q|\mathbb{P})\ \right]
\end{aligned}$$

where we used $Z=-\beta X$ and $H(Q|\mathbb{P})=\sup_{Z\in\mathcal{X}}\{\mathbb{E}^Q[Z]-\log(\mathbb{E}[e^Z])\}$ comes from Lemma 3.29 in Föllmer and Schied (2004). The penalty function is nothing but the relative entropy of the probability measure Q with respect to the physical one $\mathbb{P}$, modified by the factor $\frac{1}{\beta}$. We recall that the relative entropy is defined as

$$\begin{cases} H(Q|\mathbb{P}) := \mathbb{E}[\frac{dQ}{d\mathbb{P}}\log(\frac{dQ}{d\mathbb{P}})] & \text{if} \quad Q\ll\mathbb{P} \\ H(Q|\mathbb{P}) := +\infty & \text{otherwise} \end{cases} \tag{1.22}$$

Again as in TVaR probability measures that are considered too far from the physical one are more penalized. In this case, the bound is not anymore on the Radon-Nikodym derivative as it was in TVaR, but it arises from the penalty function $\alpha_{\min}(Q)$ that "measures" the level of entropy (distance) from $\mathbb{P}$.

1.3 Law-invariance

The wide majority of risk measures used and known in practice are *law-invariant* or *model-dependent*. Given a probability space $(\Omega,\mathcal{F},\mathbb{P})$, a risk measure ρ is said to be *law invariant* if $\rho(X)=\rho(Y)$ whenever X and Y have the same distribution under $\mathbb{P}$. This means that the outcome of law-invariant risk measures is uniquely determined by the probability distribution (model)

of the financial position to assess. VaR and the distortion risk measures are examples of law-invariant risk measures.

For practical applications the probability distribution of a financial position is unknown (in more general terms the reference probability measure $\mathbb{P}$ is not assigned) and it is estimated from a set of available data. This set of data is of limited and often small size and it may not be sufficient to estimate properly the tail of the distribution where the extreme losses occur.

Law-invariant risk measures do not include in their assessment the risk arising from model mis-specification and estimation. Moreover their outcome depends on the estimation procedure and hence is itself subject to an error. The sensitivity of risk measures to the estimation error has been investigated, among others, by Cont et al. (2010), Gerrard and Tsanakas (2011) and Gourieroux and Liu (2006). It is discussed in more detail in Chapters 3 and 4.

Going back to the definition of coherent risk measure in (1.8) (but also convex risk measures (1.19)) we note that it requires a measurable space $(\Omega, \mathcal{F})$, without specifying a reference (or physical) probability measure $\mathbb{P}$. Such risk measures are then expressed as the worst expected loss over a set $\mathcal{P}$ of generalized scenarios. Each scenario is represented by a probability measure P on $(\Omega, \mathcal{F})$. While this approach has a specific meaning, as it recognizes that a reference probability measure is generally not given, it may be difficult to deal with and is therefore not often discussed in the literature.

1.4 Dynamic risk measures

After static risk measures were studied and understood in depth, the problem of extending the above theory to a dynamic context has naturally arisen. In many applications, for example with long-term financial positions, assessments at more than one point in time are necessary. Here, newly available information and intermediate payoffs play a central role. There is a wide literature on the topic, see Artzner et al. (2007), Riedel (2004), Roorda and

Schumacher (2007) for the coherent case, Detlefsen and Scandolo (2005), Weber (2006), Föllmer and Penner (2006), Roorda and Schumacher (2008), Cheridito et al. (2006) for more general contexts. In particular Detlefsen and Scandolo (2005) extended most of the definitions and the results known for the static case to the dynamic one, and proved similar representation theorems.

There are two main approaches available for the dynamic extension of risk measures. The first is to consider financial positions as random variables with values on Ω, the set of all possible states of nature at a future time point. New available information is described by a filtration $\mathcal{F}_{t\in[0,T]}$, where the set $[0,T]$ can be discrete, continuous, finite or infinite, depending on the context. This approach, not accommodating intermediate payoffs, is treated for instance in Detlefsen and Scandolo (2005), Roorda and Schumacher (2007), Tutsch (2008) and will be presented in detail in Chapter 2.

The second approach is to consider financial positions as cash-flows on a filtered probability space, in this way both intermediate payoffs and additional information are taken into a account. This approach is presented in Artzner et al. (2007), Riedel (2004), Weber (2006) and many others. A detailed review can also be found in Roorda and Schumacher (2007).

A major problem concerning dynamic risk measures is time consistency, that is, the relation that should hold between risk assessments of the same position at different time points. If they should be related and in what sense is still a subject of debate and captures the attention of a wide literature. The most commonly treated notion of time consistency is *dynamic consistency*, which corresponds to a generalization to risk measures of the tower law for conditional expectation. However, it has been proved that under mild conditions only a very restricted class of risk measures satisfies this requirement (Kupper and Schachermayer, 2009). Weaker notions of consistency, such as acceptance and rejection consistency, are considered by Tutsch (2008) and Weber (2006) and treated in detail by Roorda and Schumacher (2007). Chapter 2 aims at investigating in depth this topic, focusing in particular on a

weak condition of time consistency, called *sequential consistency*.

1.5 Research aims and structure

Although much progress has been made and risk measurement is now a systematic activity in every financial or insurance firm, several aspects of risk measurement still need to be analyzed and clarified.

The aim of this thesis is to develop new risk measurement approaches that address research questions concerning: time-consistency, the relation that should hold across risk measurements of the same financial position at different time points; and parameter/model risk, the risk arising from parameter/model mis-specification due to estimation from a limited amount of available data.

In Chapter 2, we investigate coherent risk measures used to calculate solvency capital for long-term portfolios. As time evolves, new information becomes available and the measurement needs to be updated consequently. Simple examples show that re-applying the same risk measure with new information could lead to over- or under-estimation of capital. In this study, we characterize and construct coherent risk measures that satisfy the notion of *sequential consistency*, introduced by Roorda and Schumacher (2007). This states that financial positions that are acceptable (resp. not acceptable) in all states of the world at some future time, should also be acceptable (resp. not acceptable) at all previous times. Satisfying this requirement is important from both regulators and financial firms point of view as it leads to an efficient management of the risk capital and can reduce the risk of insolvency. For the two most common families of risk measure updates, discussed, among others, by Tutsch (2008) and Detlefsen and Scandolo (2005), we present sufficient conditions to ensure sequential consistency. Our results show that most of the coherent risk measures used in practice (such as TVaR) satisfy only partially these conditions.

Consequently, we provide a general method of constructing sequentially

consistent risk measures, which entails modification of the risk measure used at the initial time. The technique is illustrated by building a sequentially consistent version of TVaR, of the coherent entropic risk measure and of the general class of distortion risk measures. Furthermore, we also introduce a conditional version of the coherent entropic risk measure, recently proposed by Föllmer and Knispel (2011).

The last part of the chapter investigates the time horizon of risk assessment, a topic generally not treated in the literature on risk measures. Even when exposure is to long term positions, the portfolio holder or the regulator is interested in determining the capital required at a fixed future time point. For example, the impending Solvency II framework for European insurers requires that the safely invested capital corresponds to 99.5% VaR with 1 year time horizon. If the financial position expires after the time horizon, as is typical for insurance liabilities, the risk measure is applied to the fair value of the position at the horizon, rather than to the position itself. Taking into account this consideration, we introduce a new type of coherent risk measure with a rolling time horizon. We show that in this setting, the dynamic risk measures earlier constructed are no more sequentially consistent, but the weaker requirement of *acceptance consistency* can still be preserved.

The study of dynamic risk measures and time consistency properties is very sophisticated from a mathematical and numerical point of view (see for instance Föllmer and Penner (2006), Kupper and Schachermayer (2009) and the literature therein). However, it generally relies on the assumption that the probability distribution of financial positions is well specified and known. In reality this probability is not observable and generally estimated from a limited sample of past data, creating potential for substantial parameter and model error. Parameter error arises from the deviation of estimated parameters from their true values, in the context of a correctly chosen probability distribution for the financial position. Model error arises from incorrect specification of the distribution itself. For the potential of parameter (resp. model) error occurring, we use the term parameter (resp. model) uncertainty.

Cont et al. (2010), among others, emphasize how a proper risk measurement procedure should also take into account this estimation issue.

Chapter 3 contributes to understanding the impact of parameter uncertainty on risk measures used as solvency capital requirement. We follow a parametric approach where, for a given financial position, a probability distribution is fixed while the parameters are estimated from a limited random sample. The solvency capital that a firm should hold is calculated according to a law-invariant risk measure and hence depends on the random sample used to estimate the distribution parameters. The effect of parameter uncertainty can be seen as the mismatch between the theoretical and the estimated capital. Specifically, we propose to quantify this residual estimation risk as the extra capital that needs to be added to the position in order to allow for the randomness of the estimated capital.

The residual risk in general depends on the true but unknown parameters and remains in theory unknown. However, for location-scale families, we show that this dependence can be eliminated and thus we suggest adjustments of the risk measure to compensate that residual risk. In particular, we demonstrate the effectiveness of approaches based on (a) adequate increase of capital level, (b) Bayesian predictive distributions under probability-matching priors (Severini et al., 2002) and (c) residual risk estimation via parametric bootstrap.

For transformed location-scale families, approaches (b) and a heuristic modification of (a) are investigated. It is shown that trying to quantify the residual estimation risk and to apply those methods can lead to distribution functions with infinite means, which do not allow for the evaluation of coherent risk measures such as TVaR. Consequently the effectiveness of those methods is demonstrated using of a truncated version of TVaR proposed in Cont et al. (2010).

In Chapter 4 we discuss the impact of model uncertainty on risk measures used as solvency capital requirements. We assume that the correct model for the probability distribution of a financial position is unknown and instead a

set of candidate models has been specified. In this setting the effectiveness of different approaches to estimate the capital under model uncertainty is empirically investigated. In particular we consider four methods: (a) selecting the capital according to the model that represents the worst case scenario; (b) calculating the capital according to the model with the highest posterior weight calculated according to a Bayesian perspective; (c) computing the capital under each candidate model and then averaging according to the posterior weights; (d) calculating the predictive distribution of the financial position and computing the risk measure on this distribution.

The effectiveness of these approaches is quantified via a generalization of the residual estimation risk introduced in Chapter 3. As this risk does depend on the true but unknown model, we introduce a *Test Set* of models that are used as benchmarks. By applying each capital estimation method to data generated from every model in the test set, we are able to better assess their performance. In order to compare the results, for each method we also report the average, maximum and maximum of the absolute value of the residual estimation risk across models in the test set.

We perform a Monte-Carlo simulation study, using VaR and considering two different types of candidate model sets. The first where the model set almost overlaps with the test set used. This setting aims at representing a situation where the model set has been carefully specified according to expert judgment and it is termed *Informative model set*. In the second one, the model set is rather different from the test set used and corresponds to a situation where no prior information on the candidate models is available and we call this setting *Non-informative model set*.

From our analysis it appears that all the estimation approaches considered are sensitive to the model set used. In particular, for an informative model set, approach (b) performs slightly better than the others. When no prior information is available on the model set, then it is better to use an approach that averages the capital according to different models, that is approaches (c) or (d).

Interestingly, we highlight that mis-specifying the model by choosing distributions that are more heavy-tailed than the one generating the data, may reduce the capital causing a higher residual estimation risk. Hence this is not necessarily a conservative approach.

Finally in Chapter 5 we summarize the main findings of our research and outline future possible developments.

Chapter 2

Characterization and construction of sequentially consistent risk measures

Abstract: In dynamic risk measurement the problem emerges of assessing the risk of a financial position at different times. Sufficient conditions are provided for conditional coherent risk measures, in order that the requirements of acceptance-, rejection- and sequential consistency are satisfied. It is shown that these conditions are often violated for standard methods of updating. A method is consequently proposed for constructing a sequentially consistent risk measure, which entails the modification of the set of probability measures used, to obtain the risk assessment at an initial time. Consequently, time-consistent dynamic generalizations are given for well known risk measures used in insurance, such as the TVaR and distortion / Choquet risk measures, as well as the recently introduced coherent entropic risk measure. Finally we consider the situation where the term of risk exposures is longer than the time horizon used in solvency assessment. Then, regulation such as Solvency II requires replacing the financial position itself with its fair value at the time horizon. We show that in this setting acceptance consistency can be preserved, though the same is not true about rejection consistency.

2.1 Introduction

The correct quantification of risks faced by insurance companies and other financial institutions is a central task for both investors and regulators. *Risk measures* are essential tools for quantifying financial risks. Static risk measures, where uncertainty is resolved over a single period, have been extensively studied in the insurance and finance literature. Indicatively, book-length treatments with particular focus on the relationship between risk measures, their properties and stochastic orders are Goovaerts et al. (1984) and Denuit et al. (2005), while an emphasis on axiomatic characterizations can be found in Wang et al. (1997), Artzner et al. (1999), Föllmer and Schied (2002), and Frittelli and Rosazza Gianin (2002). An important class of risk measures emerging in the literature is that of distortion / Choquet risk measures, introduced in the insurance literature by Wang (1996) satisfying the properties posed by Wang et al. (1997) as well as Artzner et al. (1999). These risk measures produce the popular TVaR measure as a special case; more technical detail on them can be found in eg Carlier and Dana (2003).

Recent years have also seen an increasing interest in a dynamic approach to risk measurement, where several time periods are considered. In a dynamic setting, several issues emerge: the impact of available information on the risk assessment, the occurrence of intermediate payoffs, the time consistency among measurements of the same position at different points in time, and the assessment time horizon. A variety of approaches have been taken in the literature. Early attempts at conditioning distortion risk measures based on the results of Denneberg (1994a) are Wang and Young (1998) and Tsanakas (2004). The link between conditioning of risk measures and BSDEs has been studied by Barrieu and El Karoui (2004), Rosazza Gianin (2006), and Stadje (2010).

The strand of literature that our work most closely relates to focuses on notions of time consistency in dynamic risk measures and corresponding characterizations. Detlefsen and Scandolo (2005) considered conditional risk

measures, where the assessment outcome depends on new information becoming available. Riedel (2004), Weber (2006), Frittelli and Scandolo (2006), Artzner et al. (2007), Cheridito et al. (2006), among others, focused on risk measurements for stochastic processes. Roorda et al. (2005), Föllmer and Penner (2006), Rosazza Gianin (2006), Tutsch (2006), Weber (2004), Roorda and Schumacher (2007) discussed different types of time consistency. In the literature on dynamic preferences, properties of time consistency were already studied by Koopmans (1960) and Epstein and Schneider (2003).

In the first part of the chapter, we discuss the time consistency between assessments of the same financial position at several times. A risk measure satisfying appropriate time consistency can lead to more efficient capital management and reduce the risk of insolvency. A key notion in this area is that of *dynamic consistency*, see for example Föllmer and Penner (2006). It states that, if two positions are assessed in the same way in every future state, then should have the same assessment at the present time as well. Roorda and Schumacher (2007) proved that this requirement is equivalent to an attractive tower law property. However, in many cases dynamic consistency leads to a risk measure that produces very high capital requirement (see Tutsch (2006), Roorda and Schumacher (2008)). Furthermore, Kupper and Schachermayer (2009) prove that, under technical conditions, law-invariance (where the risk assessment depends only on the distribution of the position) and dynamic consistency reduce the class of possible risk measures to the entropic one. In this chapter we focus on the weaker requirement of *sequential consistency* (Roorda and Schumacher, 2007), combining the ideas of *acceptance* and *rejection consistency* (Weber, 2004), (Tutsch, 2006). This states:

a) A financial position cannot be considered acceptable at an initial time if it will be unacceptable in each successor state (*acceptance consistency*).

b) If the position is rejected in any state of nature at a future time point, then it should be rejected at an earlier time as well (*rejection consistency*).

We investigate sequential consistency for two of the standard ways of updating a coherent risk measure, discussed by Detlefsen and Scandolo (2005) and Tutsch (2008). The first update is obtained assuming that the new available information reduces the set of generalized scenarios that are used to construct the corresponding static risk measure. The second type of update assumes instead that new information does not influence this set. In both cases, we present sufficient conditions to ensure sequential consistency. Our results show that standard updates of a coherent risk measure (such as TVaR) often satisfy only the conditions for either acceptance or rejection consistency.

Consequently we provide a general method of constructing sequentially consistent dynamic risk measures, which requires modification of the risk measure used at the initial time. The technique is illustrated by building a sequentially consistent version of TVaR, which essentially coincides with the one proposed by Roorda and Schumacher (2008) and then extending the method to the general class of distortion / Choquet risk measures. A consistent update is also presented for the coherent entropic risk measure recently introduced by Föllmer and Knispel (2011)

The last part of the chapter concerns the time horizon of risk assessment. Even when exposure is to long term positions, the portfolio holder or the regulator is interested in determining the capital required at a future time point δ. For example, the impending Solvency II framework for European insurers requires that the safely invested capital corresponds to 99.5% VaR with 1 year time horizon. When the financial position expires before or at the time horizon δ, all the results of the first part apply. However, for longer term exposures, typical in insurance liabilities, the risk measure is applied to the fair value of the position at time δ, rather than to the position itself. This situation is outside the usual framework in the risk measures literature, as it essentially corresponds to risk measurement with an argument that changes over time, as the fair value at δ time units after measurement changes with new information. We show that even in this setting, acceptance consistency can still be preserved, but rejection consistency will in general not hold.

The chapter is organized as follows. In Section 2.2 we review the notion of conditional coherent risk measures and discuss sequential consistency. In Section 2.3, we present technical conditions on the set of generalized scenarios to ensure acceptance and rejection consistency for two different types of update. In Section 2.4 a procedure for constructing a sequentially consistent risk measure is presented. Section 2.5 discusses the relation between sequential consistency and time horizon of risk assessment.

2.2 Conditional risk measures and time consistency

2.2.1 Conditional risk measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and define $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$ the set of all bounded financial positions. Every inequality and equality involving elements of $\mathcal{X}$ is meant as holding $\mathbb{P}$-a.s. In order to take into account the role of new information, we introduce a non-trivial σ-algebra $\mathcal{G}$, such that $\{\emptyset, \Omega\} \subset \mathcal{G} \subset \mathcal{F}$. This means that at an intermediate time point before the expiry date of the portfolio, the investor or the regulator receives additional information $\mathcal{G}$. A re-assessment of the riskiness of the position at that time becomes of interest. The outcome of the new risk measurement $\rho_\mathcal{G}$ will depend on the information contained in $\mathcal{G}$, and $\rho_\mathcal{G}(X)$ will be a $\mathcal{G}$-measurable random variable. We will often refer to the starting time of the position as time 0 and the intermediate point in time, when the information is $\mathcal{G}$ is revealed, as time 1. Let $\mathcal{X}_\mathcal{G} := \mathcal{L}^\infty_\mathcal{G}(\Omega, \mathcal{G}, \mathbb{P})$ denote the set of all bounded random variables that are $\mathcal{G}$-measurable. Detlefsen and Scandolo (2005) introduce the following definition:

Definition 1. *A map $\rho_\mathcal{G} : \mathcal{X} \longrightarrow \mathcal{X}_\mathcal{G}$ is called a conditional convex risk measure if, for every $X, Y \in \mathcal{X}$, it satisfies the following properties:*

Monotonicity: *If $X \leq Y$, then $\rho_\mathcal{G}(X) \geq \rho_\mathcal{G}(Y)$.*

Conditional cash invariance: *If* $Z \in \mathcal{X}_{\mathcal{G}}, \quad$ *then* $\rho(X+Z) = \rho(X) - Z$.

Conditional convexity: $\rho(\alpha X + (1-\alpha)Y) \leq \alpha\rho(X) + (1-\alpha)\rho(Y)$ *for* $\alpha \in \mathcal{X}_{\mathcal{G}}, \quad 0 \leq \alpha \leq 1$.

Normalization: $\rho_{\mathcal{G}}(0) = 0$.

If it also satisfies

Conditional positive homogeneity: $\rho(\beta X) = \beta\rho(X)$ *for* $\beta \in \mathcal{X}_{\mathcal{G}}$, $\beta \geq 0$,

it is called a conditional coherent risk measure.

From the above properties, we can recover the definition of static coherent and convex risk measures introduced by Artzner et al. (1999) and Föllmer and Schied (2004), by simply substituting the σ-algebra $\mathcal{G}$ with the trivial one $\{\emptyset, \Omega\}$. In this case, we simply denote the risk measure $\rho(\cdot)$.

In the next sections we will make extensive use of the following sets:

$$\mathcal{M}_1(\mathbb{P}) := \{Q \text{ is a probability measure on } (\Omega, \mathcal{F}) \ : \ Q \ll \mathbb{P}\}$$

$$\mathcal{P}_{\mathcal{G}} := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q \equiv \mathbb{P} \text{ on } \mathcal{G}\}.$$

Furthermore we recall the definition of *essential supremum* and *essential infimum* of a random variable.

Definition 2. *For a given set of random variables* Φ *on* $(\Omega, \mathcal{F}, \mathbb{P})$*, we denote the essential supremum of* Φ*, the random variable* ϕ^**, such that:*

1) $\mathbb{P}(\phi^* \geq \phi) = 1 \qquad \forall \phi \in \Phi$*; and*

2) *If* ϕ^{**} *is another random variable satisfying* 1)*, then* $\mathbb{P}(\phi^{**} \geq \phi^*) = 1$.

We write

$$\phi^* := \text{ess.sup}_{\phi \in \Phi} \phi$$

Similarly, we denote the essential infimum of Φ*, the random variable*

$$\phi_* := -\text{ess.sup}_{\phi \in \Phi}(-\phi).$$

Detlefsen and Scandolo (2005) proved that any risk measure of the form

$$\rho_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \mathcal{Q}_{\mathcal{G}}} E^Q[-X|\mathcal{G}]$$

for $\mathcal{Q}_{\mathcal{G}} \subseteq \mathcal{M}_1(\mathbb{P})$ is a conditional coherent risk measure.

A collection of conditional risk measures, with increasing level of information, is called *dynamic risk measure*. In our simple setting, the dynamic risk measure is given only by an unconditional and a conditional risk measure $(\rho, \rho_{\mathcal{G}})$. Unless otherwise specified, the conditional risk measure $\rho_{\mathcal{G}}(X)$ will be an *update* of ρ, meaning that $\rho_{\mathcal{G}} = \rho$ whenever $\mathcal{G} = \{\emptyset, \Omega\}$. There does not exist a unique update for a risk measure. For example, for a set of probability measures $\mathcal{Q} \subseteq \mathcal{M}_1(\mathbb{P})$, one can define the coherent risk measure:

$$\rho(X) = \sup_{Q \in \mathcal{Q}} E^Q[-X] \tag{2.1}$$

and the updates:

$$\hat{\rho}_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X \mid \mathcal{G}] \tag{2.2}$$

where $\hat{\mathcal{Q}}_{\mathcal{G}} \subseteq \{\mathcal{P}_{\mathcal{G}} \cap \mathcal{Q}\}$ and

$$\tilde{\rho}_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \mathcal{Q}} E^Q[-X \mid \mathcal{G}] \tag{2.3}$$

where the set $\mathcal{Q}$ remains unchanged over time. Update (2.3) is probably one of the simplest and most intuitive way of updating a risk measure, (Tutsch, 2008). Update (2.2), actually representing a class of possible updates, is more sophisticated and encompasses two key features of conditional risk measurement. First, the newly available information allows us to drop some probability measures, by the requirement $\hat{\mathcal{Q}}_{\mathcal{G}} \subseteq \mathcal{Q}$. Secondly, the property $Q \equiv \mathbb{P}$ on $\mathcal{G}$, means that at time 1 risk measurement proceeds so that the set of measures constructed is forward rather than backward looking. This way of updating a risk measure was used, among others, by Detlefsen and Scandolo

(2005). In what follows we use extensively the two updates (2.2) and (2.3).

2.2.2 Examples of conditional risk measures

For an example consider the risk measure Tail Value at Risk (TVaR), that was proposed by Artzner et al. (1999) as a way to address the shortcomings of VaR. TVaR is a coherent risk measure and admits the following representation:

$$\text{TVaR}(X) = \sup_{Q \in \mathcal{Q}} E^Q[-X] \tag{2.4}$$

where

$$\mathcal{Q} := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ \frac{dQ}{d\mathbb{P}} \leq \lambda^{-1}\}.$$

A possible update of type (2.2) for TVaR, proposed by Detlefsen and Scandolo (2005), corresponds to:

$$\widehat{\text{TVaR}}_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X \mid \mathcal{G}] \tag{2.5}$$

where

$$\hat{\mathcal{Q}}_{\mathcal{G}} := \{Q \in \mathcal{P}_{\mathcal{G}} \ : \ \frac{dQ}{d\mathbb{P}} \leq \lambda^{-1}\},$$

here $\hat{\mathcal{Q}}_{\mathcal{G}} = \{\mathcal{P}_{\mathcal{G}} \cap \mathcal{Q}\}$. Update (2.3) for TVaR corresponds to the conditional risk measure:

$$\widetilde{\text{TVaR}}_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \mathcal{Q}} E^Q[-X \mid \mathcal{G}]. \tag{2.6}$$

A conditional risk measure that is introduced in this chapter is the *conditional coherent entropic risk measure*. This arises as a natural generalization of the static coherent entropic risk measure of Föllmer and Knispel (2011). In the unconditional case, this risk measure is defined as:

$$\rho^e(X) := \sup_{\mathcal{Q}} E^Q[-X] \tag{2.7}$$

where

$$\mathcal{Q} := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ H(Q \mid \mathbb{P}) \leq c\}$$

and

$$H(Q \mid \mathbb{P}) = \begin{cases} E^Q[\ \log(\frac{dQ}{d\mathbb{P}})\ ] = E^{\mathbb{P}}[\frac{dQ}{d\mathbb{P}}\ \log(\frac{dQ}{d\mathbb{P}})\ ] & \text{if } Q \ll \mathbb{P} \\ +\infty \qquad \text{otherwise} \end{cases}$$

is the relative entropy of Q with respect to $\mathbb{P}$. A possible update, consistent with (2.2), is given by:

$$\hat{\rho}^e_{\mathcal{G}}(X) := \text{ess.sup}_{\hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X|\mathcal{G}] \tag{2.8}$$

where

$$\hat{\mathcal{Q}}_{\mathcal{G}} := \{Q \in \mathcal{P}_{\mathcal{G}} \ : \ H_{\mathcal{G}}(Q \mid \mathbb{P}) \leq c\}$$

and

$$H_{\mathcal{G}}(Q \mid \mathbb{P}) := E^Q[\log(\frac{dQ}{d\mathbb{P}}) \mid \mathcal{G}] = E^{\mathbb{P}}[\ \frac{dQ}{d\mathbb{P}}\log(\frac{dQ}{d\mathbb{P}}) \mid \mathcal{G}]$$

is the conditional relative entropy of Q with respect to $\mathbb{P}$. To verify that (2.8) is in the class of updates (2.2), we have to check if $\hat{\mathcal{Q}}_{\mathcal{G}} \subseteq \{\mathcal{P}_{\mathcal{G}} \cap \mathcal{Q}\}$. The first inclusion $\hat{\mathcal{Q}}_{\mathcal{G}} \subseteq \mathcal{P}_{\mathcal{G}}$ is given. Now consider $Q \in \hat{\mathcal{Q}}_{\mathcal{G}}$:

$$H_{\mathcal{G}}(Q \mid \mathbb{P}) \leq c \quad \Rightarrow \quad E^{\mathbb{P}}[\ H_{\mathcal{G}}(Q \mid \mathbb{P})\ ] \leq c \quad \Rightarrow$$

$$E^{\mathbb{P}}[\ E^Q[\ \log(\frac{dQ}{d\mathbb{P}})|\mathcal{G}]\ ] \leq c \quad \Rightarrow \quad E^Q[\ \log(\frac{dQ}{d\mathbb{P}})\ ] = H(Q \mid \mathbb{P}) \leq c$$

where we used $Q \equiv \mathbb{P}$ on $\mathcal{G}$.

2.2.3 Sequential consistency

It is reasonable to assume that a dynamic risk measure satisfies some notions of consistency. To illustrate this issue we present here two examples of inconsistency that are not desirable in a dynamic risk measure and that generally

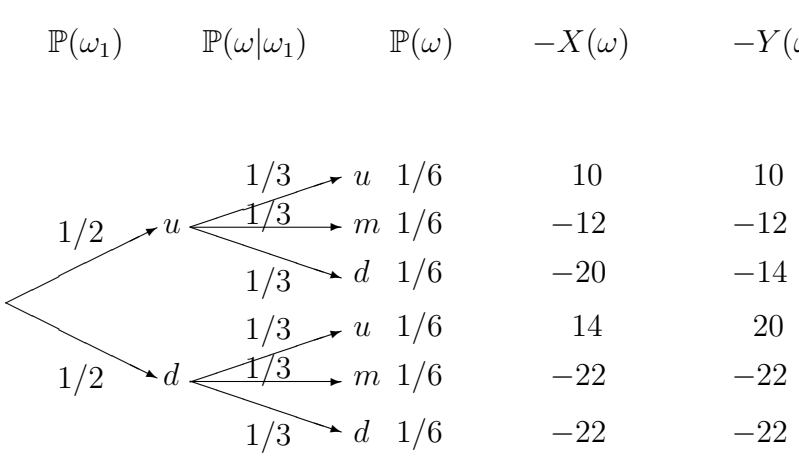

Figure 2.1: Probability distribution of $-X$ and $-Y$ under $\mathbb{P}$.

occur when we use updates (2.2) and (2.3). Example 4 was introduced by Artzner et al. (2007). By slightly modifying it, we obtained Example 3 that we use to show a different type of time inconsistency.

Example 3. Consider the dynamic risk measure $(TVaR, \widetilde{TVaR}_{\mathcal{G}})$ defined as in (2.4) and (2.6). Let $\Omega = \{uu, um, ud, du, dm, dd\}$ be the event space and $\mathbb{P}$ assign equal weight to every possible outcome as suggested by the binomial tree in figure 1. Set $\lambda = 2/3$. The set of probability measures $\mathcal{Q}$ is given by:

$$\mathcal{Q} := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q(\omega) \leq \frac{3}{2}\mathbb{P}(\omega) = \frac{1}{4} \ \ \forall \ \omega \in \Omega\}.$$

For the financial position:

$$X = [-10, 12, 20, -14, 22, 22],$$

TVaR(X) is obtained assigning the highest admissible probability (i.e. 1/4) to the worst loss, then to the second worst one and so on until the probabilities used sum up to 1. Hence:

$$\text{TVaR}(X) = (14 + 10 - 12 - 20)\frac{1}{4} = -2 \leq 0.$$

For the conditional risk measure $\widetilde{\text{TVaR}}_{\mathcal{G}}$ we seek to maximize independently the conditional expectations $E^Q[-X|\ u]$ and $E^Q[-X|\ d]$ over $Q \in \mathcal{Q}$. Note that the probability measure that maximizes the expectation in the upper

brunch (u), is generally different from the one maximizing the lower branch (d). For the upper (u) branch the probability $Q \in \mathcal{Q}$ that achieve the maximum expectation is obtained by setting $Q(uu|u) = 1$, $Q(um|u) = Q(ud|u) = 0$ and $Q(u) = \frac{1}{4}$. Then it may be assumed that $Q(du|d) = Q(dm|d) = Q(dd|d) = \frac{1}{3}$. Similarly for the lower branch (d) we can find $R \in \mathcal{Q}$ such that $\sup_{Q\in\mathcal{Q}} E^Q[-X|\ d] = E^R[-X|\ d] = 14$. Therefore we have:

$$\widetilde{\mathrm{TVaR}}_{\mathcal{G}}(X)(\omega) = \mathrm{ess.sup}_{Q\in\mathcal{Q}} E^Q[-X \mid \mathcal{G}] = \begin{cases} 10 \geq 0 & \text{if } \omega \in \{uu, um, ud\} \\ 14 \geq 0 & \text{if } \omega \in \{du, dm, dd\}. \end{cases}$$

Here, the position is acceptable at time 0 and 2 units of capital can be withdrawn from it. In contrast, at time 1, in both scenarios, the position is considered unacceptable and an amount of respectively 10 and 14 units of capital is required. This type of inconsistency, is particularly undesirable from the regulatory point of view as the risk holder may not be able to raise all the money needed (12 and 16 units in this example), leading to a possible insolvency risk. A good risk measure should detect the certainty of future capital needs, so that appropriate levels of capital can already be held at time 0.

The above example is close to the one used by Artzner et al. (2007) to illustrate a different type of inconsistency as follows:

Example 4 (Artzner et al. (2007))**.** Here we consider update (2.2) for TVaR. Using the same setting than the previous example, we have

$$\hat{\mathcal{Q}}_{\mathcal{G}} := \{Q \in \mathcal{P}_{\mathcal{G}} \ :\ Q(\omega|\omega') \leq \frac{3}{2}\mathbb{P}(\omega|\omega') = \frac{1}{2} \ \ \forall\ \omega \in \Omega \text{ and } \omega' = u, d.\}$$

For the financial position:

$$Y = [-10, 12, 14, -20, 22, 22],$$

the risk measurement at time 0 is:

$$\mathrm{TVaR}(Y) = (20 + 10 - 12 - 14)\frac{1}{4} = 1 \geq 0.$$

To calculate $\widehat{TVaR}_{\mathcal{G}}$, we need to maximize the conditional probability of adverse outcomes, under the constraint $Q(\omega|\omega') \leq \frac{1}{2}$. The probability of the upper and lower branch is already fixed to be equal to $\mathbb{P}$, so $Q(u) = Q(d) = \frac{1}{2}$ for every $Q \in \hat{\mathcal{Q}}_{\mathcal{G}}$. Then, we obtain:

$$\widehat{\mathrm{TVaR}}_{\mathcal{G}}(Y) = \begin{cases} (10-12)\frac{1}{2} = -1 \leq 0 & \text{if } \omega \in \{uu, um, ud\} \\ (20-22)\frac{1}{2} = -1 \leq 0 & \text{if } \omega \in \{du, dm, dd\}. \end{cases}$$

At time 0, the position Y is considered unacceptable and an amount of 1 unit of capital is required. At time 1, the position is considered acceptable in every state of the world and 1 unit of capital can actually be withdrawn. It means that TVaR penalizes a position that will anyway be accepted later on, requiring some capital that is not needed and that could be invested in a better way.

The same inconsistency holds for the coherent entropic risk measure $\rho^e(\cdot)$ and the update $\hat{\rho}^e_{\mathcal{G}}(\cdot)$ as it is shown in the following example.

Example 5. Assume the same setting as in Example 3 and set $c = -\ln(2/3)$. For the financial position:

$$Z = [-3, 14, 10, -9, 32, 32],$$

standard optimization techniques give $\rho^e(Z) = 0.6821 \geq 0$. The probability measure that attains the maximum in (2.7) is

$$Q = [0.2931,\ 0.0913,\ 0.1201,\ 0.4424,\ 0.0266,\ 0.0266].$$

For the conditional entropic risk measure, we have:

$$\hat{\rho}^e_{\mathcal{G}}(Z) = \begin{cases} -0.3483 \leq 0 & \text{if } \omega \in \{uu, um, ud\} \\ -0.3108 \leq 0 & \text{if } \omega \in \{du, dm, dd\}. \end{cases}$$

Again, at time 0 it is required to hold some capital, that in no-case will be asked at time 1.

To address such inconsistencies the notion of *sequential consistency* was proposed by Roorda and Schumacher (2007). It emerges as a combination of the two requirements of acceptance and rejection consistency, introduced by Weber (2004) for cash-flows and Tutsch (2006) for random variables.

Definition 3. *An unconditional and a conditional risk measure ρ and $\rho_{\mathcal{G}}$ are said to be sequentially consistent if, for every $X \in \mathcal{X}$, they satisfy:*

$$\rho_{\mathcal{G}}(X) \leq 0 \quad \Longrightarrow \quad \rho(X) \leq 0 \qquad \textit{acceptance consistency} \tag{2.9}$$

$$\rho_{\mathcal{G}}(X) \geq 0 \quad \Longrightarrow \quad \rho(X) \geq 0 \qquad \textit{rejection consistency.} \tag{2.10}$$

A risk measure satisfying (2.9) would not be subject to the inconsistencies seen in Example 4. Similarly, a risk measure satisfying (2.10) would avoid inconsistency faced in Example 3.

As stated by Roorda and Schumacher (2008) a dynamic risk measure $(\rho(X), \rho_{\mathcal{G}}(X))$ is sequentially consistent if and only if, for every $X \in \mathcal{X}$,

$$\text{ess.inf}_{\omega\in\Omega}\rho_{\mathcal{G}}(X) \;\leq\; \rho(X) \;\leq\; \text{ess.sup}_{\omega\in\Omega}\rho_{\mathcal{G}}(X). \tag{2.11}$$

This implies that the capital requirement at time 0 cannot be higher than the highest amount that could ever be asked in the future. On the other side, it cannot be smaller than the lowest amount of capital that would ever be required in the future.

2.3 Conditions for sequential consistency

2.3.1 Preliminaries

Before starting, we recall some notions that are essential for the next sections. Let $L^0(\overline{\mathbb{R}})$ be the space of extended random variables, i.e. of maps from Ω to $\overline{\mathbb{R}} := [-\infty, \infty]$.

Definition 4. *A set $\mathcal{Y} \subseteq L^0(\overline{\mathbb{R}})$ is upward directed if, for any two elements Y_1, $Y_2 \in \mathcal{Y}$, there is always a third one $Y \in \mathcal{Y}$ such that $Y \geq \max(Y_1, Y_2)$.*

For upward directed sets, the following result holds:

Lemma 2.3.1. *If $\mathcal{Y} \subseteq \mathcal{L}^0(\overline{\mathbb{R}})$ is upward directed, then*

$$E^{\mathbb{P}}[\text{ess.sup}\mathcal{Y}] = \sup_{Y \in \mathcal{Y}} E^{\mathbb{P}}[Y],$$

provided that both expectations exist.

The same holds if we replace the expectation with the expectation conditional to a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ (for a proof, see Detlefsen and Scandolo (2005)). In what follows we will extensively apply the above result to the set $\mathcal{C} := \{E^Q[-X \mid \mathcal{G}],\ Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}$. Here, for every $X \in \mathcal{X}$, each probability measure $Q \in \hat{\mathcal{Q}}_{\mathcal{G}}$ identifies a random variable $E^Q[-X \mid \mathcal{G}]$ and the essential supremum of $\mathcal{C}$ can be expressed as

$$\text{ess.sup}_{Q \in \hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X|\mathcal{G}].$$

Following Detlefsen and Scandolo (2005), consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. A regular conditional probability $Q_{\mathcal{G}}$ is defined as a map $Q_{\mathcal{G}} : \Omega \times \mathcal{F} \to [0, 1]$, that is, a version of the expected conditional value of $\mathbb{I}_A$ for any $A \in \mathcal{F}$ and a probability measure for $\omega \in \Omega$. For every $Q \in \mathcal{M}_1(\mathbb{P})$, the pasting probability $\mathbb{P}Q_{\mathcal{G}}$ is defined as

$$\mathbb{P}Q_{\mathcal{G}}(A) := E^{\mathbb{P}}[Q_{\mathcal{G}}(\cdot, A)] \qquad \forall A \in \mathcal{F},$$

where $Q_{\mathcal{G}}(\cdot, A)$ is a version of $E^Q[\mathbb{I}_A|\mathcal{G}]$.

In the case of a two-period binomial tree this concept becomes straightforward. The probability $\mathbb{P}Q_{\mathcal{G}}$ is obtained using $\mathbb{P}$ for the first period, from time 0 to time 1, and then switching to Q in the second one, from time 1 to time 2. The main property of pasting probability is:

$$E^{\mathbb{P}Q_{\mathcal{H}}}[X \mid \mathcal{G}] = E^{\mathbb{P}}[\, E^Q[X \mid \mathcal{H}] \mid \mathcal{G}] \qquad \text{for } \mathcal{H} \subseteq \mathcal{G},$$

for any σ-algebra $\mathcal{H}$ such that $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$.

2.3.2 Conditions for sequential consistency

Sequential consistency conditions for update $\hat{\rho}_{\mathcal{G}}(\cdot)$

Consider two risk measures as in (2.1) and (2.2). For the dynamic risk measure $(\rho, \hat{\rho}_{\mathcal{G}})$, the following proposition holds:

Proposition 2.3.2. *(i) If, for every $X \in \mathcal{X}$, the set $\mathcal{C} := \{E^Q[-X \mid \mathcal{G}],\ Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}$ is upward directed, then the risk measures ρ and $\hat{\rho}_{\mathcal{G}}$ are rejection consistent.*

(ii) If, for every Q in $\mathcal{Q}$, the pasting probability $\mathbb{P}Q_{\mathcal{G}}$ is in $\hat{\mathcal{Q}}_{\mathcal{G}}$, then the risk measures ρ and $\hat{\rho}_{\mathcal{G}}$ are acceptance consistent.

Hence, if (i) and (ii) hold, then ρ and $\hat{\rho}_{\mathcal{G}}$ are sequentially consistent.

Proof. (i) Let $\hat{\rho}_{\mathcal{G}}(X) \geq 0$. Since the expected value of a positive random variable is again positive, we have:

$$\hat{\rho}_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X|\mathcal{G}] \geq 0 \qquad \Longrightarrow \qquad E^{\mathbb{P}}[\, \text{ess.sup}_{Q \in \hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X|\mathcal{G}]\,] \geq 0.$$

As the set C is upward directed, Lemma 2.3.1 leads to:

$$E^{\mathbb{P}}[\, \text{ess.sup}_{Q \in \hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X \mid \mathcal{G}]\,] = \sup_{Q \in \hat{\mathcal{Q}}_{\mathcal{G}}} E^{\mathbb{P}}[\, E^Q[-X \mid \mathcal{G}]\,] \geq 0$$

but $Q \equiv \mathbb{P}$ on $\mathcal{G}$, therefore

$$E^{\mathbb{P}}[\ E^{Q}[-X|\mathcal{G}]\ ] = E^{Q}[\ E^{Q}[-X|\mathcal{G}]\ ] = E^{Q}[-X] \implies \sup_{Q\in\hat{\mathcal{Q}}_{\mathcal{G}}} E^{Q}[-X] \geq 0.$$

Since $\hat{\mathcal{Q}}_{\mathcal{G}} \subseteq \mathcal{Q}$, we obtain:

$$\rho(X) = \sup_{Q\in\mathcal{Q}} E^{Q}[-X] \geq \sup_{Q\in\hat{\mathcal{Q}}_{\mathcal{G}}} E^{Q}[-X] \geq 0$$

as desired.

(ii) If $\hat{\rho}_{\mathcal{G}}(X) \leq 0$ then

$$E^{Q}[-X \mid \mathcal{G}] \leq 0 \qquad \forall Q \in \hat{\mathcal{Q}}_{\mathcal{G}}.$$

As $\mathbb{P}Q_{\mathcal{G}} \in \hat{\mathcal{Q}}_{\mathcal{G}}$ for every $Q \in \mathcal{Q}$:

$$E^{\mathbb{P}Q_{\mathcal{G}}}[-X \mid \mathcal{G}] \leq 0 \qquad \forall Q \in \mathcal{Q}.$$

By definition of the pasting probability:

$$0 \geq E^{\mathbb{P}Q_{\mathcal{G}}}[-X \mid \mathcal{G}] = E^{\mathbb{P}}[\ E^{Q}[-X \mid \mathcal{G}] \mid \mathcal{G}] = E^{Q}[-X \mid \mathcal{G}] \qquad \forall Q \in \mathcal{Q}.$$

Hence:

$$E^{Q}[-X] = E^{Q}[\ E^{Q}[-X \mid \mathcal{G}]\ ] \leq 0 \qquad \forall Q \in \mathcal{Q}$$

and

$$\rho(X) = \sup_{Q\in\mathcal{Q}} E^{Q}[-X] \leq 0$$

as desired. □

Remark 1. (i) is a technical condition ensuring that we can exchange the essential supremum and the expectation, where (ii), instead, requires that at time 0, we only use probability measures that will be also used for the risk assessment at time 1. In this way, we avoid measuring risk using probability measures that, in any case, will not even be considered when the information

$\mathcal{G}$ is revealed.

Remark 2. Risk measures as in (2.1) and (2.2) generally fail acceptance consistency. Indeed, the proof of Prop. 2.3.2 cannot be applied in this case, because:

$$\text{ess.sup}_{Q\in\hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X|\mathcal{G}] \leq 0 \qquad \Longrightarrow \qquad \sup_{Q\in\hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X] \leq 0$$

but generally,

$$\sup_{Q\in\mathcal{Q}} E^Q[-X] \geq \sup_{Q\in\hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X] \leq 0,$$

so we cannot deduce

$$\rho(X) \leq 0.$$

Lemma 2.3.3. *$TVaR$ and $\widehat{TVaR}_{\mathcal{G}}$ are rejection consistent.*

Proof. To prove it, we need to verify that the set $\mathcal{C} := \{E^Q[-X \mid \mathcal{G}],\ Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}$ is upward directed , i.e. condition (i). Consider two probability measures Q' and Q'' in $\hat{\mathcal{Q}}_{\mathcal{G}}$ and define Q as

$$Q(B) = Q'(A \cap B) + Q''(A^c \cap B) \tag{2.12}$$

where the set $A \in \mathcal{G}$ is defined as

$$A := \{E^{Q'}[-X \mid \mathcal{G}] \geq E^{Q''}[-X \mid \mathcal{G}]\}.$$

It is not difficult to see that $Q \in \hat{\mathcal{Q}}_{\mathcal{G}}$. For every $C \in \mathcal{G}$

$$Q(C) = Q'(A \cap C) + Q''(A^c \cap C) = \mathbb{P}(A \cap C) + \mathbb{P}(A^c \cap C) = \mathbb{P}(C)$$

so $Q \equiv \mathbb{P}$ on $\mathcal{G}$. Similarly, for every $B \in \mathcal{F}$

$$Q(B) = Q'(A \cap B) + Q''(A^c \cap B) \leq \lambda^{-1}(\mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)) \leq \lambda^{-1}\mathbb{P}(B)$$

so $\frac{dQ}{d\mathbb{P}} \leq \lambda^{-1}$ and

$$E^Q[-X \mid \mathcal{G}] = \mathbb{I}_A E^{Q'}[-X \mid \mathcal{G}] + \mathbb{I}_{A^c} E^{Q''}[-X \mid \mathcal{G}] \tag{2.13}$$
$$\geq \max\{E^{Q'}[-X \mid \mathcal{G}],\ E^{Q''}[-X \mid \mathcal{G}]\}.$$

Therefore $\mathcal{C}$ is upward directed and $TVaR$ and $\widehat{TVaR}_{\mathcal{G}}$ are rejection consistent. □

TVaR and the update $\widehat{TVaR}_{\mathcal{G}}$ do not satisfy sequential consistency because, as we have already seen in Example 4, they are acceptance inconsistent. It is immediate to verify that condition (ii) of Proposition 2.3.2 is not satisfied because the probability measure $Q^* \in \mathcal{Q}$ that maximizes $\sup_{Q\in\mathcal{Q}} E^Q[-X]$ in Example 4, is such that $\mathbb{P}Q^*_{\mathcal{G}}$ does not belong to $\hat{\mathcal{Q}}_{\mathcal{G}}$.

Lemma 2.3.4. *The coherent entropic risk measures $\rho^e(\cdot)$ and $\rho^e_{\mathcal{G}}(\cdot)$ are rejection consistent.*

Proof. Again, we only need to prove that the set $\mathcal{C} := \{E^Q[-X|\mathcal{G}], Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}$ is upward directed for every $X \in \mathcal{X}$. Following the steps of Lemma 2.3.3, we define a probability measure Q as in (2.12). We already know that $Q \equiv \mathbb{P}$ on $\mathcal{P}_{\mathcal{G}}$. By definition:

$$\begin{aligned}
&H_{\mathcal{G}}(Q|\mathbb{P}) = E^Q[\log\frac{dQ}{d\mathbb{P}} \mid \mathcal{G}] \\
&= \mathbb{I}_A E^{Q'}[\log\frac{dQ'}{d\mathbb{P}}|\mathcal{G}] + \mathbb{I}_{A^c} E^{Q''}[\log\frac{dQ''}{d\mathbb{P}}|\mathcal{G}] \\
&\leq \max\{E^{Q'}[\log\frac{dQ'}{d\mathbb{P}}|\mathcal{G}], E^{Q''}[\log\frac{dQ''}{d\mathbb{P}}|\mathcal{G}]\} \leq c
\end{aligned}$$

so that $Q \in \hat{\mathcal{Q}}_{\mathcal{G}}$ and the set is upward directed. □

Also $(\rho^e(\cdot), \rho^e_{\mathcal{G}}(\cdot))$ does not satisfy sequential consistency as follows from Example 5.

Sequential consistency conditions for update $\tilde{\rho}_{\mathcal{G}}(\cdot)$

We now discuss time consistency for risk measures where the set of probability measures is not updated when new information arrives, that is, the risk measures (2.1) and (2.3).

Proposition 2.3.5. *(i) The risk measures ρ and $\tilde{\rho}$ are acceptance consistent.*

(ii) If, for every $X \in \mathcal{X}$

(a) the set $\mathcal{C} := \{E^Q[-X \mid \mathcal{G}] \ : \ Q \in \mathcal{Q}\}$ is upward directed and $\mathbb{P}Q_{\mathcal{G}} \in \mathcal{Q}$ for every Q in $\mathcal{Q}$; or

(b) the supremum in the definition of $\tilde{\rho}$ is attained, that is $\exists\ P^* \in \mathcal{Q}$:

$$E^{P^*}[-X \ |\mathcal{G}] = ess\sup_{P\in\mathcal{Q}} E^P[-X \ |\mathcal{G}] \geq E^{P'}[-X \ |\mathcal{G}], \ \forall P' \in \mathcal{Q} \tag{2.14}$$

then ρ and $\tilde{\rho}$ are rejection consistent. Hence, if (a) or (b) hold, then ρ and $\tilde{\rho}$ are sequentially consistent.

Proof. (i) is proved by Tutsch (2008). Now we show that either of the conditions (a) and (b) implies rejection consistency. For (a) the proof follows the same steps of Prop. 2.3.2(i). For (b), condition (2.14) together with

$$\tilde{\rho}_{\mathcal{G}}(X) \geq 0,$$

imply that

$$\exists P^* \in \mathcal{Q} \qquad \text{such that} \qquad E^{P^*}[-X \mid \mathcal{G}] = \tilde{\rho}_{\mathcal{G}}(X) \geq 0.$$

Then,

$$E^{P^*}[E^{P^*}[-X \mid \mathcal{G}]] = E^{P^*}[-X] \geq 0.$$

As $P^* \in \mathcal{Q}$, we have:

$$\rho(X) = \sup_{Q \in \mathcal{Q}} E^Q[-X] \geq E^{P^*}[-X] \geq 0.$$

□

Remark 3. Conditional risk measures with the additional property of being continuous from below admit a representation in terms of probability measures where the supremum is attained. Nevertheless, it is usually attained on a different set than $\mathcal{Q}$, such that condition (2.14) is not necessarily verified. For details see Bion-Nadal (2004). The situation becomes easier if we work in a setting where Ω is finite. In this case, the supremum is attained if the set $\mathcal{Q}$ is closed and convex and there exists a probability measure $P \in \mathcal{Q}$ such that

$$PQ_{\mathcal{G}} \in \mathcal{Q} \quad \text{for every} \quad Q \in \mathcal{Q}. \tag{2.15}$$

An example of such a risk measure, satisfying sequential consistency on a finite probability space, was proposed by Roorda and Schumacher (2007). Define

$$STVaR(X) = \sup_{Q \in \mathcal{Q}'} E^Q[-X]$$

where

$$\mathcal{Q}' := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ \frac{dQ}{d\mathbb{P}} \leq \lambda^{-1}, \ \frac{d\mathbb{P}Q_{\mathcal{G}}}{d\mathbb{P}} \leq \lambda^{-1}\}$$

and consider the update

$$\widetilde{STVaR}_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \mathcal{Q}'} E^Q[-X \mid \mathcal{G}].$$

When Ω is finite, the set $\mathcal{Q}'$ is a polytope and $\mathbb{P}$ satisfies condition (2.15), so, using convex analysis arguments, it is possible to show that the essential supremum is attained and the risk measure is sequentially consistent. A similar argument, where the supremum is attained, is used by Fasen and Svejda (2010) to construct a sequentially consistent version of distortion risk measures in a finite framework.

2.4 Constructing sequentially consistent risk measures

2.4.1 General construction

In the previous sections, conditions for the sequential consistency of dynamic risk measures were presented. However no risk measure considered actually satisfies these conditions on an infinite probability space. Now, drawing inspiration from Roorda and Schumacher (2007), we show that it is possible to slightly modify a dynamic risk measure in order to turn it into a sequentially consistent one. The method is applied to produce a sequentially consistent version of the coherent entropic risk measure as well as the class of coherent distortion / Choquet risk measures. A numerical example is given for TVaR.

Start again with a coherent risk measure as in (2.1). Suppose that we consider ρ suitable for our measurement purposes, but the update $\hat{\rho}_{\mathcal{G}}$ does not satisfy the conditions for sequential consistency required by Proposition 2.3.2. In order to construct a sequentially consistent risk measure, starting from the update $\hat{\rho}_{\mathcal{G}}$, we work backwards, defining a new unconditional risk measure as:

$$\hat{\rho}'(X) = \sup_{Q \in \hat{Q}'} E^Q[-X]$$

where

$$\hat{Q}' := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q \in \mathcal{Q}, \ \mathbb{P}Q_{\mathcal{G}} \in \hat{\mathcal{Q}}_{\mathcal{G}}\}.$$

Proposition 2.4.1. *If the set* $\mathcal{C} := \{E^Q[-X \ |\mathcal{G}] \ : \ Q \in \hat{\mathcal{Q}}_{\mathcal{G}} \}$ *is upward directed, then the risk measures* $\hat{\rho}'(X)$ *and* $\hat{\rho}_{\mathcal{G}}(X)$ *are sequentially consistent.*

Proof. From Prop. 2.3.2 they are sequentially consistent by construction. □

Remark 4. Notice that $\rho(X)$ and $\hat{\rho}'(X)$ admit the same update (2.2), i.e.

$$\hat{\rho}'_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \hat{\mathcal{Q}}'_{\mathcal{G}}} E^Q[-X \mid \mathcal{G}] = \hat{\rho}_{\mathcal{G}}(X)$$

where

$$\hat{\mathcal{Q}}'_{\mathcal{G}} := \{Q \in \mathcal{P}_{\mathcal{G}} \ : \ Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}.$$

Moreover, they are close in the sense that $\hat{\rho}'(X)$ requires, at time 0, all the conditions on the set of measures, required by $\rho(X)$, but in addition the conditions that will be required at time 1, when new information arrives. In other words, the condition $\mathbb{P}Q_{\mathcal{G}} \in \hat{\mathcal{Q}}_{\mathcal{G}}$ excludes, at time 0, probability measures that will not be used in the representation of the update. In this way, we avoid rejecting financial positions that would be accepted when the information in $\mathcal{G}$ is revealed.

Remark 5. Once we have constructed the new unconditional risk measure

$$\hat{\rho}'(X) = \sup_{Q \in \hat{Q}'} E^Q[-X]$$

where

$$\hat{Q}' := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q \in \mathcal{Q}, \ \mathbb{P}Q_{\mathcal{G}} \in \hat{\mathcal{Q}}_{\mathcal{G}}\},$$

we can easily see that, if the set $\mathcal{C} := \{E^Q[-X \ |\mathcal{G}] \ : \ Q \in \mathcal{Q} \ \}$ is upward directed, also the update (2.3)

$$\tilde{\rho}'_{\mathcal{G}}(X) = \text{ess.sup}_{Q \in \hat{Q}'} E^Q[-X|\mathcal{G}]$$

satisfies sequential consistency as it verifies all the conditions required in Prop. 2.3.5. We remark that in this case, the two updates are actually the same, ie:

$$\text{ess.sup}_{Q \in \hat{Q}'} E^Q[-X|\mathcal{G}] = \text{ess.sup}_{Q \in \hat{Q}'_{\mathcal{G}}} E^Q[-X|\mathcal{G}]$$

due to the structure of the probability measure sets $\hat{Q}'$ and $\hat{Q}'_{\mathcal{G}}$.

2.4.2 Examples of sequentially consistent risk measures

In this section, sequentially consistent versions of TVaR, coherent entropic, and distortion / Choquet risk measures are introduced.

We start from the static TVaR and the update $\widehat{TVaR}_{\mathcal{G}}$. From Example 4, we already know that this update fails acceptance consistency. As shown in Lemma 2.3.3, the set $\mathcal{C} := \{E^Q[-X \mid \mathcal{G}] \ : \ Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}$ is upward directed. Therefore, we can define a new unconditional risk measure, as:

$$\widehat{TVaR}'(X) = \sup_{Q \in \hat{Q}'} E^Q[-X]$$

where

$$\begin{aligned} \hat{Q}' &:= \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q \in \mathcal{Q}, \ \mathbb{P}Q_{\mathcal{G}} \in \mathcal{Q}_{\mathcal{G}}\} \\ &= \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ \frac{dQ}{d\mathbb{P}} \leq \lambda^{-1}, \frac{d\mathbb{P}Q_{\mathcal{G}}}{d\mathbb{P}} \leq \lambda^{-1}\}. \end{aligned} \tag{2.16}$$

From Prop. 2.4.1, $\widehat{TVaR}'$ and $\widehat{TVaR}_{\mathcal{G}}$ are sequentially consistent.

In the following example it is seen how the sequentially consistent version of TVaR solves the inconsistencies faced in the examples 3 and 4.

Example 6. To see this, consider again the same setting as in Example 4, where $\Omega = \{uu, um, ud, du, dd, dm\}$, $\mathbb{P}(\omega) = 1/6$ for every $\omega \in \Omega$ and $\lambda = 2/3$. The set of probability measures considered at time 0 and time 1 are respectively:

$$\hat{\mathcal{Q}}' := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q(\omega) \leq \frac{3}{2}\mathbb{P}(\omega) = \frac{1}{4},$$

$$Q(\omega|\omega') \leq \frac{3}{2}\mathbb{P}(\omega|\omega') = \frac{1}{2} \ \ \forall \ \omega \in \Omega, \ \omega' = \{u, d\}\}$$

and

$$\hat{\mathcal{Q}}_{\mathcal{G}} = \{Q \in \mathcal{P}_{\mathcal{G}} \ : \ Q(\omega|\omega') \leq \frac{3}{2}\mathbb{P}(\omega|\omega') = \frac{1}{2} \ \ \forall \ \omega \in \Omega, \ \omega' = \{u, d\}\}.$$

For the financial position:

$$Y = [-10, 12, 14, -20, 22, 22],$$

we have

$$\widehat{\mathrm{TVaR}}'(Y) = (10 - 12 + 20 - 22)\frac{1}{4} = -1 \neq TVaR(Y)$$

and

$$\widehat{\mathrm{TVaR}}_{\mathcal{G}}(Y) = \begin{cases} (10-12)\frac{1}{2} = -1 & \text{if } \omega = uu, um, ud \\ (20-22)\frac{1}{2} = -1 & \text{if } \omega = du, dm, dd \end{cases}$$

Therefore, the acceptance inconsistency has been eliminated.

The new risk measure does not present the rejection inconsistency of Example 3 either. Indeed for the random variable:

$$X = \{-10, 12, 20, -14, 22, 22\},$$

we have

$$\widehat{TVaR}'(X) = (10 - 12 + 14 - 22)\frac{1}{4} = -\frac{5}{2}$$

while

$$\widehat{\mathrm{TVaR}}_{\mathcal{G}}(X) = \begin{cases} (10-12)\frac{1}{2} = -1 & \text{if } \omega = uu, um, ud \\ (14-22)\frac{1}{2} = -4 & \text{if } \omega = du, dm, dd. \end{cases}$$

Now we show how Prop. 2.4.1 can be used to construct a sequentially consistent version of the coherent entropic risk measure. As shown in Lemma 2.3.4, the set $\mathcal{C} := \{E^Q[-X|\mathcal{G}], Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}$ is upward directed. To have sequential consistency we define a new risk measure

$$\rho'^e(X) := \sup_{\hat{Q}'} E^Q[-X] \tag{2.17}$$

where

$$\hat{Q}' := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ H(Q|\mathbb{P}) \leq c \quad \text{and} \quad H_{\mathcal{G}}(\mathbb{P}Q_{\mathcal{G}}|\mathbb{P}) \leq c\}.$$

From Prop. 2.4.1 (2.17) and (2.8) are sequentially consistent.

The same procedure can be used to obtain a sequentially consistent version of distortion risk measures. Distortion risk measures are types of Choquet integrals (Denneberg, 1994b), and can be seen as generalizations of TVaR. A similar result to what we present here was obtained by Fasen and Svejda (2010) for distortion risk measures in a finite setting. For a comprehensive discussion of distortion risk measures we refer to Carlier and Dana (2003). Here, we define a distortion risk measure as:

$$\rho^C(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q[-X],$$

where

$$\mathcal{Q} = \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q(A) \leq g(\mathbb{P}(A)) \quad \forall A \in \mathcal{F}\}$$

and $g : [0,1] \mapsto [0,1]$ is an increasing, concave function such that:

$$g(0) = 0; \qquad g(1) = 1.$$

A possible update for a distortion risk measure is the following:

$$\hat{\rho}^C_{\mathcal{G}}(X) = \sup_{\hat{\mathcal{Q}}_{\mathcal{G}}} E^Q[-X|\mathcal{G}] \tag{2.18}$$

where

$$\hat{\mathcal{Q}}_{\mathcal{G}} := \{Q \in \mathcal{P}_{\mathcal{G}} \ : \ Q_{\mathcal{G}}(\cdot, A) \leq g(\mathbb{P}_{\mathcal{G}}(\cdot, A)) \quad \forall A \in \mathcal{F}\}$$

and $Q_{\mathcal{G}}(\cdot, A)$ is a version of $E^Q[\mathbb{I}_A|\mathcal{G}]$. To see that (2.18) belongs to the class of updates (2.2), we show that $\hat{\mathcal{Q}}_{\mathcal{G}} \subseteq \mathcal{Q}$. For every $Q \in \hat{\mathcal{Q}}_{\mathcal{G}}$ and for every $A \in \mathcal{F}$:

$$Q_{\mathcal{G}}(\cdot, A) \leq g(\mathbb{P}_{\mathcal{G}}(\cdot, A)) \qquad \Rightarrow \tag{2.19}$$

$$\mathbb{P}Q_{\mathcal{G}}(\cdot, A) = E^{\mathbb{P}}[Q_{\mathcal{G}}(\cdot, A)] \leq E^{\mathbb{P}}[g(\mathbb{P}_{\mathcal{G}}(\cdot, A))] \qquad \Rightarrow \tag{2.20}$$

$$Q(A) \leq g(E^{\mathbb{P}}[\mathbb{P}_{\mathcal{G}}(\cdot, A)]) = g(\mathbb{P}(A)), \tag{2.21}$$

where, in (2.20) we used the definition of pasting probabilities and (2.21)

follows from $Q \in \mathcal{P}_{\mathcal{G}}$ and the Jensen's inequality. We have already seen from Example 3 that this kind of update generally is not acceptance consistent. Now, consider the new risk measure:

$$\hat{\rho}'^{C}(X) = \sup_{Q \in \hat{\mathcal{Q}}'} E^{Q}[-X] \tag{2.22}$$

where

$$\hat{\mathcal{Q}}' := \{Q \in \mathcal{M}_1(\mathbb{P}) \,:\, Q(A) \leq g(\mathbb{P}(A)) \text{ and } Q_{\mathcal{G}}(\cdot, A) \leq g(\mathbb{P}_{\mathcal{G}}(\cdot, A)) \quad \forall A \in \mathcal{F}\},$$

the following result holds:

Lemma 2.4.2. *The risk measures (2.22) and (2.18) are sequentially consistent.*

Proof. We need to prove that the set $\mathcal{C} := \{E^{Q}[-X|\mathcal{G}],\ Q \in \hat{\mathcal{Q}}_{\mathcal{G}}\}$ is upward directed. To see it, consider again a probability measure Q and the set A defined as in (2.12), we only need to show that $Q \in \hat{\mathcal{Q}}_{\mathcal{G}}$. If $\mathbb{P}(A) = \{0, 1\}$ the proof is immediate. Assume now that $\mathbb{P}(A) \neq \{0, 1\}$, by the definition of Q' and Q'', for every $B \in \mathcal{F}$ we have:

$$\begin{aligned} Q(B) &= \mathbb{P}(A)Q'(B|A) + \mathbb{P}(A^c)Q''(B|A^c) \\ &\leq \mathbb{P}(A)g(\mathbb{P}(B|A)) + \mathbb{P}(A^c)g(\mathbb{P}(B|A^c)) \\ &\leq g(\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)) = g(\mathbb{P}(B)), \end{aligned}$$

where we used the concavity of $g(\cdot)$. It follows from Prop. 2.4.1 that $(\hat{\rho}'^{C}, \hat{\rho}^{C}_{\mathcal{G}})$ is sequentially consistent. □

2.5 The solvency time horizon in dynamic risk measurement

2.5.1 Sequential consistency in multiple periods

Here we briefly discuss the results of Sections 2.3 and 2.4 in a multi-period setting. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}\}_{n\in[0,N]}$ a filtration with $N \in \mathbb{N}$ and $\mathcal{F} \equiv \mathcal{F}_N$. A dynamic coherent risk measure is then defined as a collection

$$(\rho_0, \rho_1, \ldots, \rho_{N-1}), \tag{2.23}$$

where:

$$\rho_n(X) := \rho_{\mathcal{F}_n}(X) = \text{ess.sup}_{Q\in\mathcal{Q}_n} E^Q[-X \mid \mathcal{F}_n] \qquad \forall n \in [0, \ldots, N-1].$$

for a certain set of measures $\mathcal{Q}_0, \ldots, \mathcal{Q}_{N-1}$.

The extension of the notion of sequential consistency to this setting is straightforward.

Definition 5. *The dynamic risk measure $(\rho_0, \rho_1, \ldots, \rho_{N-1})$ is sequentially consistent if, for every $X \in \mathcal{X}$, it satisfies:*

(i) acceptance consistency

$$\rho_n(X) \leq 0 \quad \Longrightarrow \quad \rho_{n-1}(X) \leq 0 \qquad \forall n \in [0, N-1]; \text{ and} \tag{2.24}$$

(ii) rejection consistency

$$\rho_n(X) \geq 0 \quad \Longrightarrow \quad \rho_{n-1}(X) \geq 0 \qquad \forall n \in [0, N-1] \tag{2.25}$$

Again, given a static risk measure as in (2.1) we can define the updates:

$$\hat{\rho}_n(X) = \text{ess.sup}_{Q\in\hat{\mathcal{Q}}_n} E^Q[-X|\mathcal{F}_n] \tag{2.26}$$

and

$$\tilde{\rho}_n(X) = \text{ess.sup}_{Q\in\mathcal{Q}} E^Q[-X|\mathcal{F}_n] \tag{2.27}$$

where

$$\hat{\mathcal{Q}}_n \subseteq \{Q \in \mathcal{P}_n \ \cap \ Q \in \mathcal{Q}\}$$

and

$$\mathcal{P}_n := \{Q \in \mathcal{M}_1(\mathbb{P}) \ : \ Q \equiv \mathbb{P} \ \text{ on } \ \mathcal{F}_n\}$$

The results presented in Section 2.3 still hold. Specifically we have:

Corollary 2.5.1. *(i) If, for every $X \in \mathcal{X}$, the set $\mathcal{C} := \{E^Q[-X \mid \mathcal{F}_n],\ Q \in \hat{\mathcal{Q}}_n\}$ is upward directed for every $n \in [0, \dots, N-1]$, then the risk measure $(\rho, \hat{\rho}_1, \dots, \hat{\rho}_{N-1})$ is rejection consistent.*

(ii) If, for every Q in $\mathcal{Q}$, the pasting probability $\mathbb{P}Q_{\mathcal{F}_n}$ is in $\hat{\mathcal{Q}}_n$ for every $n \in [0, \dots, N-1]$, then the risk measure $(\rho, \hat{\rho}_1, \dots, \hat{\rho}_{N-1})$ is acceptance consistent.

Hence, if (i) and (ii) hold, then $(\rho, \hat{\rho}_1, \dots, \hat{\rho}_{N-1})$ is sequentially consistent.

Corollary 2.5.2. *(i) The dynamic risk measure $(\rho, \tilde{\rho}_1, \dots, \tilde{\rho}_{N-1})$ is acceptance consistent.*

(ii) If, for every $X \in \mathcal{X}$,

(a) the set $\mathcal{C} := \{E^Q[-X \mid \mathcal{F}_n] \mid Q \in \mathcal{Q}\}$ is upward directed and $\mathbb{P}Q_n \in \mathcal{Q}$ for every Q in $\mathcal{Q}$ for every $n \in [0, \dots, N-1]$, or

(b) the supremum in (2.27) is attained, i.e. $\exists\ P^* \in \mathcal{Q} \quad s.t.:$

$$E^{P^*}[-X\,|\mathcal{F}_n] = \operatorname{ess\,sup}_{P\in\mathcal{Q}} E^P[-X\,|\mathcal{F}_n] \geq E^{P'}[-X\,|\mathcal{F}_n] \qquad \forall P' \in \mathcal{Q} \tag{2.28}$$

then $(\rho, \tilde{\rho}_1, \dots, \tilde{\rho}_{N-1})$ is rejection consistent.

Hence, if (a) or (b) hold, then $(\rho, \tilde{\rho}_1, \dots, \tilde{\rho}_{N-1})$ is sequentially consistent.

Remark 6. The procedure to build a sequentially consistent version of a dynamic coherent risk measure, is the same as the one seen in Section 2.4.

We start from the update (2.26) at time $N-1$ and we proceed backwards, adding all the conditions that we need. For a random variable X, the new risk measure will be $(\hat{\rho}'_n)_{n\in[0,N]}$, where:

$$\hat{\rho}'_n(X) = \text{ess.sup}_{Q\in\hat{\mathcal{Q}}'_n} E^Q[-X|\mathcal{F}_n]$$

and the set $\hat{\mathcal{Q}}'_n$ is defined as:

$$\hat{\mathcal{Q}}'_n := \{Q \in \mathcal{P}_n \ : \ \mathbb{P}Q_l \in \hat{\mathcal{Q}}'_l \quad \forall l \in \mathbb{N}, s.t. \quad n \leq l \leq N-1\}.$$

Note that, at the penultimate time, $\hat{\rho}'_{N-1}$ coincides with ρ_{N-1}.

2.5.2 Dynamic risk measures and solvency time horizon

Here we consider the effect of a solvency time horizon on risk measurement. Often regulatory capital requirements are specified in relation to a fixed time horizon, eg 1 year in insurance regulation such as Solvency II (or a much shorter horizon of 10 days, in banking under Basel II). When a portfolio contains long term liabilities (eg insurance contracts) that expire beyond the time horizon, the random terminal payoff has to be substituted with its (random) market consistent value at the time horizon. Valuation may be carried out either using "mark-to-market" replication arguments or, if that is not possible, using a "mark-to-model" cost of capital approach (see eg Wüthrich and Salzmann (2010)).

Here, we assume that a "mark-to-market" valuation is possible via a risk neutral measure Q^*. Hence, the position X is substituted with its price at the solvency time horizon δ. In insurance, this price is for example the price of reinsuring the position at time δ. In what follows, we will introduce a new risk measure that takes into account this aspect. Consider a random variable

$X \in \mathcal{X}$ and define the functional:

$$\rho_0^\delta(X) := \sup_{P \in \mathcal{Q}} E^P\big[-E^{Q^*}[X \mid \mathcal{F}_\delta]\ \big]$$

for a certain pricing measure Q^*. In general, for $n \in [0, N-1]$,

$$\rho_n^\delta(X) := \text{ess.sup}_{P \in \mathcal{Q}_n} E^P\big[-E^{Q^*}[X|\mathcal{F}_{n+\delta}] \mid \mathcal{F}_n\ \big].$$

For the moment, we do not specify the set $\mathcal{Q}_n$ and thus what kind of update we will be using. Note that $\rho_n^\delta(X)$ is nothing but the application of a conditional coherent risk measure

$$\rho_n(\cdot) = \text{ess.sup}_{P \in \mathcal{Q}_n} E^P[-\cdot\,|\mathcal{F}_n\ ]$$

to the conditional expectation of the position X under a certain probability measure Q^*. It is straightforward to prove that the conditional risk measure $\rho_n^\delta(X)$ is coherent.

2.5.3 Sequential consistency of $\rho^\delta(\cdot)$

We now consider whether the coherent risk measures

$$\rho_n^\delta(X) = \rho_n(E^{Q^*}[X \mid \mathcal{F}_{n+\delta}])$$

and

$$\rho_{n+1}^\delta(X) = \rho_{n+1}(E^{Q^*}[X \mid \mathcal{F}_{n+1+\delta}])$$

inherit some time consistency from ρ_n and ρ_{n+1}. For convenience, consider $\rho_0^\delta(X)$ and $\rho_1^\delta(X)$:

$$\rho_0^\delta(X) := \sup_{Q \in \mathcal{Q}} E^Q[\ E^{Q^*}[-X \mid \mathcal{F}_\delta]\ ] \tag{2.29}$$

and

$$\rho_1^\delta(X) := \text{ess.sup}_{Q \in \mathcal{Q}_1} E^Q[\ E^{Q^*}[-X \mid \mathcal{F}_{1+\delta}] \mid \mathcal{F}_1] \tag{2.30}$$

for certain sets $\mathcal{Q}$ and $\mathcal{Q}_1$.

Lemma 2.5.3. *If $\rho_0(X)$ and $\rho_1(X)$ are acceptance consistent, then so are $\rho_0^{\delta+1}(X)$ and $\rho_1^{\delta}(X)$.*

Proof. If $\rho_0(X)$ and $\rho_1(X)$ are acceptance consistent, then

$$\rho_1^{\delta}(X) = \rho_1(\, E^{Q^*}[X \mid \mathcal{F}_{1+\delta}]\,) \leq 0 \quad \Longrightarrow \quad \rho_0(\, E^{Q^*}[X \mid \mathcal{F}_{1+\delta}]\,) = \rho_0^{\delta+1}(X) \leq 0.$$

□

To establish consistency between ρ_0^{δ} and ρ_1^{δ}, we need some additional conditions on the set $\mathcal{Q}$ and the probability measure Q^*. In particular, we recall that a risk measure is *law-invariant*, if it assigns the same value to financial positions having the same distribution.

Proposition 2.5.4. *If the risk measures ρ_0 and ρ_1 are acceptance consistent, and either*

(i) The probability measure Q^* belongs to $\mathcal{Q}$ and $QQ_n^* \in \mathcal{Q}$ for every $Q \in \mathcal{Q}$ and for every $n \in [0, N-1]$; or

(ii) $Q^* \in \mathcal{M}_1(\mathbb{P})$ and the risk measure $\rho(\cdot)$ is coherent law-invariant and continuous from below,

then ρ_0^{δ} and ρ_1^{δ} are acceptance consistent.

Proof. (i) We already know that $\rho_1^{\delta}(X) \leq 0$, implies

$$\rho_0^{1+\delta}(X) = \sup_{Q \in \mathcal{Q}} E^{Q}[\, E^{Q^*}[-X \mid \mathcal{F}_{1+\delta}]\,] \leq 0$$

so

$$E^{Q}[\, E^{Q^*}[-X \mid \mathcal{F}_{1+\delta}]\,] \leq 0 \qquad \forall Q \in \mathcal{Q}$$

In particular we can choose $Q = RQ_{\delta}^* \in \mathcal{Q}$ for every $R \in \mathcal{Q}$ and obtain

$$E^{RQ_{\delta}^*}[\, E^{Q^*}[-X \mid \mathcal{F}_{1+\delta}]\,] = E^{R}[\, E^{Q^*}[-X \mid \mathcal{F}_{\delta}]\,] \leq 0 \qquad \forall R \in \mathcal{Q}$$

therefore

$$\rho_0^\delta(X) = \sup_{R\in\mathcal{Q}} E^R[\ E^{Q^*}[-X \mid \mathcal{F}_\delta]\ ] \leq 0.$$

(ii) The proof follows from Corollary 4.59 in Föllmer and Schied (2004), where is proved that ρ is monotone with respect to the second order stochastic dominance $\succeq$. From

$$E^{Q^*}[-X \mid \mathcal{F}_{1+\delta}] \succeq E^{Q^*}[-X \mid \mathcal{F}_\delta]$$

and Lemma 2.5.3, we obtain the acceptance consistency of ρ_0^δ and ρ_1^δ.

□

Therefore, acceptance consistency can still be valid when we substitute X with $E^{Q^*}[X|\mathcal{F}_\delta]$.

Remark 7. The same does not hold for rejection consistency. Even if $\rho_0(X)$ and $\rho_1(X)$ are rejection consistent, this does not imply that $\rho_0^\delta(X)$ and $\rho_1^\delta(X)$ are as well. To see it, consider

$$\rho_0^\delta(X) := \sup_{Q\in\mathcal{Q}} E^Q[\ E^{Q^*}[-X \mid \mathcal{F}_\delta]\ ] \tag{2.31}$$

and

$$\rho_1^\delta(X) := \text{ess.sup}_{Q\in\mathcal{Q}_1} E^Q[\ E^{Q^*}[-X \mid \mathcal{F}_{1+\delta}] \mid \mathcal{F}_1] \tag{2.32}$$

for certain sets $\mathcal{Q}$ and $\mathcal{Q}_1$. Again from the rejection consistency of $\rho_0(X)$ and $\rho_1(X)$, we have

$$\rho_1^\delta(X) \geq 0 \qquad \Longrightarrow \qquad \rho_0^{1+\delta(X)} \geq 0$$

but in general we do not have enough information to derive

$$\rho_0^{1+\delta}(X) \geq 0 \qquad \Longrightarrow \qquad \rho_0^\delta(X) \geq 0.$$

Then, if a position is rejected, this does not give enough information to

reject also its conditional expectation, which is generally less volatile than the position itself.

2.6 Conclusions

We contribute to the discussion of the properties of dynamic risk measures, focusing on the time consistency of conditional coherent risk measures. Technical conditions are discussed to ensure *sequential consistency* for different types of updates. These requirements are generally not satisfied by coherent risk measures, such as e.g. TVaR. Hence, it becomes sometimes necessary to modify slightly the risk measure in order to obtain consistent dynamic risk measurements. This is achieved by an adjustment to the coherent risk measure set of generalized scenarios. The procedure amounts to excluding, a priori, probability measures that will not be taken into account, in any case, when new information is available. As an example, an application of this approach to TVaR, to the coherent entropic risk measure and to the class of distortion / Choquet risk measures is presented. Finally, we discuss the role of the solvency time horizon. When the position has a long term, solvency regulation often requires that risk is measured at an earlier time horizon. In this case, the argument of the risk measure is the position's fair value at that horizon. In this changed setting, acceptance consistency can be preserved, but in general we lose rejection consistency.

Chapter 3

Quantifying and controlling residual estimation risk

Abstract: In the present contribution, we address the problem of quantifying and controlling the impact of parameter uncertainty on risk measures used to calculate solvency capital requirements. We introduce a new mechanism to quantify this impact measured as the additional capital needed to allow for randomness in the data sample. Generalizing the arguments of Gerrard and Tsanakas (2011), we show that for risk measures used in practice, parameter uncertainty originates a residual risk that can increase the probability of insolvency and the size of the shortfall. For location-scale distribution families, we prove the effectiveness of three approaches for reducing this residual risk. These are based on (a) raising the capital requirement, (b) Bayesian predictive distributions under probability-matching priors (Severini et al., 2002) and (c) residual risk estimation via parametric bootstrap. For heavy-tailed distributions it emerges that for law-invariant coherent risk measures the residual risk is hard to quantify and control. Hence, we investigate a truncated version of Tail Value at Risk (Cont et al., 2010) and investigate the effectiveness of (a) and (b) above for heavy-tailed distributions arising as increasing transformations of location-scale families. Numerical results obtained via Monte-Carlo simulations demonstrate that the proposed methods

still perform well.

3.1 Introduction

A risk measure is a map that assigns to every financial position a number summarizing the information relative to the future monetary outcomes of the financial position and their probability. Risk measures have become everyday tools in the quantification of risk and receive major interest from both practitioners and academics in finance and actuarial science. We refer, among others, to Goovaerts et al. (1984) and Wang et al. (1997) for a theoretical approach to risk measures in insurance; to Artzner et al. (1999), Delbaen (2002), Föllmer and Schied (2002), Frittelli and Rosazza Gianin (2002) and Acerbi (2002) for an in-depth analysis of coherent, spectral and convex risk measures and to McNeil et al. (2005) for an overview of the literature with focus on statistical methods.

The wide majority of risk measures used in finance and actuarial science are law-invariant (Kusuoka (2001), Acerbi (2007)), so that their outcome depends uniquely on the distribution of the financial loss itself. In practice, such distribution is unknown and estimated from a finite sample of data. The estimated risk measure is then subject to parameter/model error. There is a wide literature on parameter and model uncertainty in risk analysis, see for instance Cairns (2000) and Gibson, R. (ed.) (2000). The impact of this estimation procedure on risk measurement has been investigated by Jorion (1996) and Gerrard and Tsanakas (2011) for Value at Risk, by Dowd and Blake (2006) Gourieroux and Liu (2006), Heyde et al. (2007), Cont et al. (2010) for coherent, spectral and distortion risk measures. There is also a wide literature on quantile estimation that is related to this subject, see for example Christoffersen (1998), McNeil and Frey (2000). For statistical methods for estimating VaR and TVaR, including methods from Extreme Value Theory, see McNeil et al. (2005).

In the present contribution, we follow a parametric approach where the

loss probability distribution is known while the parameters are unknown. The randomness implicit in the parameter estimation generates an exposure to uncertainty that is not addressed by the risk measure itself. In particular, Gerrard and Tsanakas (2011) prove that parameter uncertainty increases the probability of failure when the capital is calculated according to Value-at-Risk. They demonstrate the effectiveness of two approaches based, respectively, on frequentist statistics and Bayesian prediction with probability matching priors (Severini et al., 2002). Here, we generalize their arguments to law-invariant, translation invariant and positive homogeneous risk measures and propose a frequentist approach to measuring the impact of parameter uncertainty on risk assessment. The estimated risk measure is considered as a random variable depending on a sample of available data. If there was no parameter uncertainty, the risk measure would be the minimum amount of capital required to make the position acceptable. Due to the randomness in the sample, the estimated risk measure could be higher or lower than the nominal capital. Hence we suggest measuring the impact of parameter uncertainty as the extra capital needed to account for this additional source of uncertainty, calculated under the true but unknown parameters and we call this *residual estimation risk*.

Once the residual estimation risk has been quantified, we investigate three different approaches to control it:

(a) Setting the capital requirement according to a more conservative risk measure to compensate the impact of parameter uncertainty;

(b) Calculating the capital using the Bayesian predictive distribution of the loss;

(c) Increasing the estimated risk measure, by an amount of capital corresponding to the residual estimation risk as estimated by parametric bootstrap.

For location-scale families all these methods prove to be effective and their

performance does not depend on the true but unknown parameters. In particular we show that (a) eliminates the residual estimation risk completely. The Bayesian approach (b) achieves the same result for location families. For scale families, it allows to control a ratio of the loss and estimated capital, while for location-scale families its effectiveness is demonstrated numerically. We also demonstrate that for location-scale families, every iteration of parametric bootstrap leads to a reduction of the risk and does not increase the computational expense and it works exactly for location families.

Recently there has been a debate on the difficulties in using coherent risk measures, especially for heavy-tailed distributions. Nešlehová et al. (2006) emphasize that law-invariant coherent risk measures are not defined for heavy-tailed distributions with infinite mean and hence cannot be used in the context of extreme value theory. Cont et al. (2010) prove that coherent risk measures are less robust than VaR to small changes in the data set. In Section 3.4, we join that debate and show that the impact of parameter uncertainty for heavy-tailed distributions arising as increasing transformations of location-scale families (eg Log-normal, Pareto) cannot generally be quantified nor controlled with coherent risk measures. Hence, we investigate a truncated version of TVaR_p (Cont et al., 2010), which we call here TTVaR_{p_1,p_2}. This arises from averaging quantiles starting from p_1 to $p_2 < 1$. We adapt the approaches used for location-scale families and show numerically that they still perform well.

The chapter is organized as follows. In Section 3.2 we revisit the main definitions and properties of risk measures and introduce the notion of residual estimation risk. In Section 3.3 approaches (a), (b), (c) for controlling the residual estimation risk for location-scale families are investigated and the numerical results presented. Section 3.4 discusses the impact of parameter uncertainty on heavy-tailed distributions and non location scale families. Finally, Section 3.5 contains the conclusions. Tables with the numerical results are presented at the end of the chapter.

3.2 Residual estimation risk

3.2.1 Risk measures

The financial loss of a portfolio is modelled by a random variable Y. Thus, in the event $\{Y > 0\}$ a portfolio loss occurs, while $\{Y \leq 0\}$ corresponds to a gain. Throughout the chapter we assume that the mean of Y is finite. Formally, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathcal{X} \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ represents the set of all financial losses considered. The distribution of $Y \in \mathcal{X}$ is $F(\cdot;\theta)$ where $\theta \in \Theta$ is a vector of parameters. We write $Y \sim F(\cdot;\theta)$ and assume throughout that F is continuous, invertible, with density function $f(\cdot;\theta)$.

Consistently with McNeil et al. (2005), a risk measure is a functional $\rho: \mathcal{X} \longrightarrow \mathbb{R}$ that assigns to every financial loss $Y \in \mathcal{X}$ a real number $\rho(Y)$. $\rho(Y)$ is expressed in monetary units and may represent a regulatory capital requirement, which is the interpretation we follow here. Alternatively, in the context of the actuarial literature, $\rho(Y)$ may represent the technical price of an insurance contract with liability Y (Goovaerts et al. (1984); Wang et al. (1997)). Following Artzner et al. (1999), a loss is *acceptable* if $\rho(Y) \leq 0$ and *not acceptable* if $\rho(Y) > 0$.

The risk measures considered here satisfy the following standard properties.

(1) *Translation invariance.* If $m \in \mathbb{R}$, $\rho(Y + m) = \rho(X) + m$;

(2) *Positive homogeneity.* If $\lambda \geq 0$, $\rho(\lambda Y) = \lambda \rho(Y)$;

(3) *Law-invariance.* If $Y_1 \stackrel{d}{=} Y_2$, $\rho(Y_1) = \rho(Y_2)$,

where $\stackrel{d}{=}$ denotes equality in distribution. Translation invariance reflects the requirement that adding a sure loss to the portfolio will increase its capital requirement by the same amount, while positive homogeneity implies that risk measurement is insensitive to scale. Law-invariance requires that two losses with the same distribution be subject to the same capital requirement. Because of this, a risk measure can also be evaluated as a functional of a dis-

tribution, such that for $Y \sim F(\cdot;\theta)$ we may denote $\rho(Y) \equiv \rho[F(\cdot;\theta)]$. With this notation, which will also be used in the sequel, translation invariance and positive homogeneity can be written as $\rho[F(\cdot - m;\theta)] = \rho[F(\cdot;\theta)] + m$ and $\rho[F(\frac{\cdot}{\lambda};\theta)] = \lambda\rho\big[F(\cdot;\theta)\big]$ respectively.

Three risk measures satisfying the above properties are

$$\mathrm{VaR}_p(Y) := \inf\{m \in \mathbb{R} \ : \ \mathbb{P}(Y \leq m) \geq p\}, \tag{3.1}$$

$$\mathrm{TVaR}_p(Y) := \frac{1}{1-p}\int_p^1 \mathrm{VaR}_u(Y)du, \tag{3.2}$$

$$\mathrm{TTVaR}_{p_1,p_2}(Y) = \frac{1}{p_2 - p_1}\int_{p_1}^{p_2} \mathrm{VaR}_u(Y)du. \tag{3.3}$$

The VaR_p measure, used extensively in insurance and banking regulation, is the $100p^{th}$ percentile of the loss distribution; in particular for invertible distributions such as the ones we consider here it is $\mathrm{VaR}_p[F(\cdot;\theta)] = F^{-1}(p;\theta)$. VaR_p is characterised by its insensitivity to the extreme tails of loss distributions, which is related to its violation of the *coherence* axioms of Artzner et al. (1999). TVaR_p, also termed Expected Shortfall, corrects for this defect by considering the average of all VaRs above the $100p^{th}$ percentile. However, this introduces sensitivity to extreme percentiles, which may not be reliably estimable from limited data. TTVaR_{p_1,p_2}, proposed by Cont et al. (2010), offers a compromise between those two risk measures: while it considers most of the tail, it does not reflect some very extreme losses (beyond the $100p_2^{th}$ percentile) that TVaR does consider.

3.2.2 Parameter uncertainty and residual estimation risk

For a loss $Y \sim F(\cdot;\theta)$, the value of the parameter θ is typically unknown and needs to be estimated from a sample. An i.i.d. random sample of size n from $F(\cdot;\theta)$ will be denoted by $\mathbf{X} = \{X_1, \ldots, X_n\}$; with slight abuse of notation we write $\mathbf{X} \sim F(\cdot;\theta)$ and from now on we assume that Y is also independent of $\mathbf{X}$. An estimator of θ, typically a Maximum Likelihood Estimator (MLE),

will be denoted by $\hat{\theta}$. As $\hat{\theta}$ depends on the random sample $\mathbf{X}$, it can itself be viewed as a random variable.

The capital that the holder of Y needs to hold, consistently with the risk measure ρ, will also depend on the random sample and is denoted by $\eta(\mathbf{X})$. A standard choice of capital estimator arises from applying the risk measure to the estimated loss distribution, that is, setting

$$\eta(\mathbf{X}) = \rho[F(\cdot; \hat{\theta})]. \tag{3.4}$$

We reserve the notation $\eta(\mathbf{X})$ for the capital estimator derived by MLE, but we will see in Section 3.3 that other choices of capital estimator can be appropriate. It is useful to distinguish between $\eta(\mathbf{X})$, a random variable representing the estimator of required capital, and $\eta(\mathbf{x})$, a fixed number representing the estimated required capital for a particular dataset $\mathbf{x}$ obtained as a realisation of $\mathbf{X}$.

We now introduce the key idea by which we propose to measure estimation risk. First, note that from translation invariance, we have that

$$\rho(Y - \rho(Y)) = 0, \tag{3.5}$$

such that $\rho(Y)$ is the amount of capital that needs to be subtracted from the loss Y to make it marginally acceptable. For a given estimate of the capital $\eta(\mathbf{x})$, it will generally be the case that $\eta(\mathbf{x}) \neq \rho(Y)$ and therefore $\rho(Y - \eta(\mathbf{x})) \neq 0$. Reflecting the variability in the random sample $\mathbf{X}$, we can consider $Y - \eta(\mathbf{X})$ as the random variable that represents the loss, after the (variable) capital estimator has been subtracted from it. We then define as *residual estimation risk* the quantity

$$\mathrm{RR}(\theta, \eta) = \rho(Y - \eta(\mathbf{X})). \tag{3.6}$$

Equation (3.6) is analogous (3.5), with the theoretical capital value $\rho(Y)$ substituted by the capital estimator $\eta(\mathbf{X})$. The residual estimation risk thus

reflects the amount of capital that needs to be subtracted from the loss $Y - \eta(\mathbf{X})$ in order to make it acceptable. A positive residual risk implies that the impact of parameter uncertainty is such that subtracting the capital estimator $\eta(\mathbf{X})$ from the loss Y does not produce an acceptable loss. Hence more safely invested capital needs to be held.

The residual estimation risk depends on the true but unknown parameter θ and on the capital estimator η (it also depends on the risk measure ρ, the family of distribution F, and the number of samples n, though these dependencies are suppressed in the notation). In particular, the dependence on the true parameter implies that the residual risk cannot in general be exactly calculated when the parameter θ is unknown, as is the case in any realistic application. However, it will be shown in Section 3.3 that in some cases, judicious choice of the capital estimator η can nonetheless eliminate residual estimation risk.

In the particular case where the risk measure is VaR, such that $\eta(\mathbf{X}) = \mathrm{VaR}_p[F(\cdot;\hat{\theta})]$, the following equivalence holds

$$\mathrm{VaR}_p(Y - \eta(\mathbf{X})) \geq 0 \Leftrightarrow \mathbb{P}(Y > \eta(\mathbf{X})) \geq 1 - p. \tag{3.7}$$

The right-hand-side of inequality (3.7) signifies a probability of failure (future loss exceeding the capital estimator) higher than the acceptable level $1 - p$ and was used as a measure of parameter uncertainty by Gerrard and Tsanakas (2011), whose approach the present paper thus generalises. The quantity $\mathbb{P}(Y > \eta(\mathbf{X}))$ can be interpreted as the relative frequency of exceptions when back-testing a VaR model, see for instance Chapter 13 in Christoffersen (2011). Hence our definition of residual estimation risk (3.6) can be seen as a back-testing criterion for general risk measures. Alternatively, if we interpret the randomness of the capital estimator as volatility across agents in a financial market, the failure probability can be interpreted as an expected frequency of failures (i.e. defaults) across an idealised market of agents with identical but independent exposures. Under this interpreta-

tion, (3.6) is a regulatory measure of the effectiveness across the market of the capital estimator $\eta(\mathbf{X}) = \rho[F(\cdot;\hat{\theta})]$.

We now present a simple example to illustrate the concepts above. Further examples with numerical evaluation of residual estimation risk are given in Section 3.3.5.

Example 7. To enable the derivation of simple closed form expressions, we focus on a simple normal model, where $Y, \mathbf{X} \sim \mathcal{N}(\mu, \sigma^2)$. The mean μ is unknown, but the standard deviation σ is known. Hence we write $Y \sim F(\cdot;\mu) \equiv \Phi\left(\frac{\cdot-\mu}{\sigma}\right)$, where Φ is the standard normal distribution. The standard normal density is denoted by ϕ. Also note that we can write $Y \stackrel{d}{=} \mu + \sigma Z$, where $Z \sim \mathcal{N}(0,1)$.

The location parameter μ is estimated via its MLE

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Since $\hat{\mu} \sim \mathcal{N}(\mu, \sigma^2/n)$, we can also write $\hat{\mu} \stackrel{d}{=} \mu + \frac{\sigma}{\sqrt{n}}U$, where $U \sim \mathcal{N}(0,1)$. We recall that the random variables U and Z are independent.

The risk measure $\rho \equiv \mathrm{TVaR}_p$ is used. It is well known that the TVaR for a normal variable is given by (see Example 2.18 in McNeil et al. (2005))

$$\mathrm{TVaR}_p(Y) = \mu + \sigma\frac{\phi(\Phi^{-1}(p))}{1-p}.$$

From now on we denote:

$$c(p) := \frac{\phi(\Phi^{-1}(p))}{1-p}$$

and use it throughout the chapter. The capital estimator then becomes

$$\eta(\mathbf{X}) = \mathrm{TVaR}_p[F(\cdot;\hat{\mu})] = \hat{\mu} + \sigma c(p) \stackrel{d}{=} \mu + \frac{\sigma}{\sqrt{n}}U + \sigma c(p).$$

Consequently, the residual risk can be calculated as

$$\begin{aligned}\mathrm{RR}(\mu,\eta) &= \mathrm{TVaR}_p(Y-\eta(\mathbf{X})) \\ &= \mathrm{TVaR}_p\left(\mu+\sigma Z-\mu-\frac{\sigma}{\sqrt{n}}U-\sigma c(p)\right) \\ &= \sigma\mathrm{TVaR}_p\left(Z-\frac{1}{\sqrt{n}}U\right)-\sigma c(p),\end{aligned}$$

which is independent of the unknown parameter μ. Since $Z-\frac{1}{\sqrt{n}}U\sim\mathcal{N}(0,1+1/n)$, it is

$$\begin{aligned}\mathrm{RR}(\mu,\eta) &= \sigma\sqrt{1+\frac{1}{n}}c(p)-\sigma c(p) \\ &= \sigma\left(\sqrt{1+\frac{1}{n}}-1\right)c(p).\end{aligned}$$

The formula for $\mathrm{RR}(\mu,\eta)$ shows that the residual risk is always positive, such that parameter uncertainty has always an adverse effect and necessitates holding capital in excess of what is given by the MLE-based capital estimator $\eta(\mathbf{X})$. As expected, as the sample size n increases, the residual risk decreases. Unsurprisingly, a long history of i.i.d. data tends to eliminate the residual estimation risk.

□

3.3 Controlling residual estimation risk for location-scale families

3.3.1 Location-scale distribution families

Throughout Section 3.3, we focus on distribution functions that belong to location-scale families. Such distributions, like the normal, student-t, and Laplace (double-exponential) distribution are commonly used in modelling asset returns. They admit simple parameterisations, with one parameter measuring location (e.g. mean) and another measuring scale (e.g. standard

deviation). It is noted that increasing transformations of random variables in location-scale families are useful models for insurance losses and for asset prices (rather than returns) – such distributions will be further discussed in Section 3.4.

Formally, two random variables, Y and Z belong to the same *location-scale family*, if there exist $a \in \mathbb{R}$ and $b > 0$, such that $Y \overset{d}{=} bZ + a$. Denote the parameter vector $\theta = (\mu, \sigma)$, such that any random variable in the location scale family follows $F(\cdot, (\mu, \sigma))$. We say that $Z \sim F(\cdot, (0, 1))$ has a standardised distribution and simply denote it by $F \equiv F(\cdot, (0, 1))$. Hence, we can write $Y \sim F(\cdot, (\mu, \sigma)) = F\left(\frac{\cdot - \mu}{\sigma}\right)$. For example, for the normal distribution, where μ stands for the mean and σ for the standard deviation, the standardised distribution is denoted by Φ.

The analysis of location-scale families is further aided by the simple representation of parameter estimators. If the parameter vector $\theta = (\mu, \sigma)$ is estimated via Maximum Likelihood, then standard results (see for instance Gerrard and Tsanakas (2011)) show that

$$\hat{\mu} \overset{d}{=} \mu + \sigma U, \quad \hat{\sigma} \overset{d}{=} \sigma V, \tag{3.8}$$

where U and V are random variables whose distribution depends on the sample size n, but not on θ. Hence $\hat{\mu}$ and $\hat{\sigma}$ also belong to a location-scale and scale family respectively.

Now we consider the impact of parameter uncertainty on the risk measure, via the notion of residual estimation risk. From the translation invariance, positive homogeneity, and law-invariance properties of the risk measure, it follows that for $Y \sim F(\cdot; (\mu, \sigma)),\ Z \sim F$, it is

$$\rho(Y) = \rho(\mu + \sigma Z) = \mu + \sigma\rho[F]. \tag{3.9}$$

Let the capital estimator be based on the MLE, such that $\eta(\mathbf{X}) = \rho[F(\cdot; \hat{\theta})]$,

where $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$. It follows that

$$\eta(\mathbf{X}) = \hat{\mu} + \hat{\sigma}\rho[F] \stackrel{d}{=} \mu + \sigma U + \sigma V \rho[F]. \tag{3.10}$$

Consequently, the residual estimation risk can be calculated as:

$$\begin{aligned} \mathrm{RR}(\theta, \eta) &= \rho(Y - \rho[F(\cdot, \hat{\theta})]) \\ &= \rho(\mu + \sigma Z - \mu - \sigma U - \sigma V \rho[F]) \\ &= \sigma\rho(Z - U - V\rho[F]). \end{aligned} \tag{3.11}$$

Hence, while in general the residual estimation risk remains unknown, for location-scale families it does not depend on the location parameter μ and is directly proportional to the scale one σ. In particular, the amount $\rho(Z - U - V\rho[F])$ does not depend on the unknown parameters. The simplicity of this representation enables us to provide simple approaches for bringing the residual risk close to zero, which are investigated in the remainder of Section 3.3.

3.3.2 Adjustment to the risk measure

The first approach that we investigate relates to modifying the risk measure in a way that compensates for parameter uncertainty and brings the residual estimation risk down to zero. To motivate the approach, consider another risk measure ρ_{adj}, that may be used to set capital. Under this risk measure, the capital estimator, using again MLE, will be

$$\eta_{adj}(\mathbf{X}) = \rho_{adj}[F(\cdot; \hat{\theta})] \tag{3.12}$$

Analogously with (3.11), we can write

$$\mathrm{RR}(\theta, \eta_{adj}) = \rho(Y - \rho_{adj}[F(\cdot, \hat{\theta})]) = \sigma\rho(Z - U - V\rho_{adj}[F]). \tag{3.13}$$

Noting that the quantity $\rho(Z - U - V\rho_{adj}[F])$ does not depend on the true but unknown parameters θ, it becomes apparent that we can choose the risk measure $\rho_{adj}[F])$ specifically so as to set the residual risk of (3.13) to zero. This is simple enough to do for a risk measure defined for a given security level that can be varied. For example, if $\rho = \text{TVaR}_p$, we can let $\rho_{adj} = \text{TVaR}_q$ for some $q \neq p$. The value of q (generally larger than p) is then chosen such that $\text{TVaR}_p(Z - U - V\text{TVaR}_q[F]) = 0$. The process is illustrated by the following example.

Example 8. We continue Example 7 of a normal distribution with unknown mean. Let us assume that a capital estimator

$$\eta_{adj}(\mathbf{X}) = TVaR_q[F(\cdot;\hat{\mu})] = \hat{\mu} + \sigma c(q) \overset{d}{=} \mu + \frac{\sigma}{\sqrt{n}}U + \sigma c(q)$$

is used, where q is higher than the confidence level p of the regulatory risk measure TVaR_p and again $c(q) = \frac{\phi(\Phi^{-1}(q))}{1-q}$.

The residual estimation risk is then calculated using similar arguments as in Example 7:

$$\begin{aligned}
\text{RR}(\mu, \eta_{adj}) &= \text{TVaR}_p\left(\mu + \sigma Z - \mu - \frac{\sigma}{\sqrt{n}}U - \sigma c(q)\right) \\
&= \sigma\text{TVaR}_p\left(Z - \frac{1}{\sqrt{n}}U\right) - \sigma c(q) \\
&= \sigma\sqrt{1 + \frac{1}{n}}c(p) - \sigma c(q).
\end{aligned}$$

Therefore, to achieve $\text{RR}(\mu, \eta_{adj}) = 0$, one needs to solve for q the equation

$$\sqrt{1 + \frac{1}{n}}c(p) = c(q),$$

which is easily done numerically.

The required level of q is plotted in Figure 8, against the sample size n, for $p \in \{0.95, 0.99, 0.995\}$. It can be seen that in each case $q > p$ and as the sample size increases the adjusted confidence level q decays to the nominal

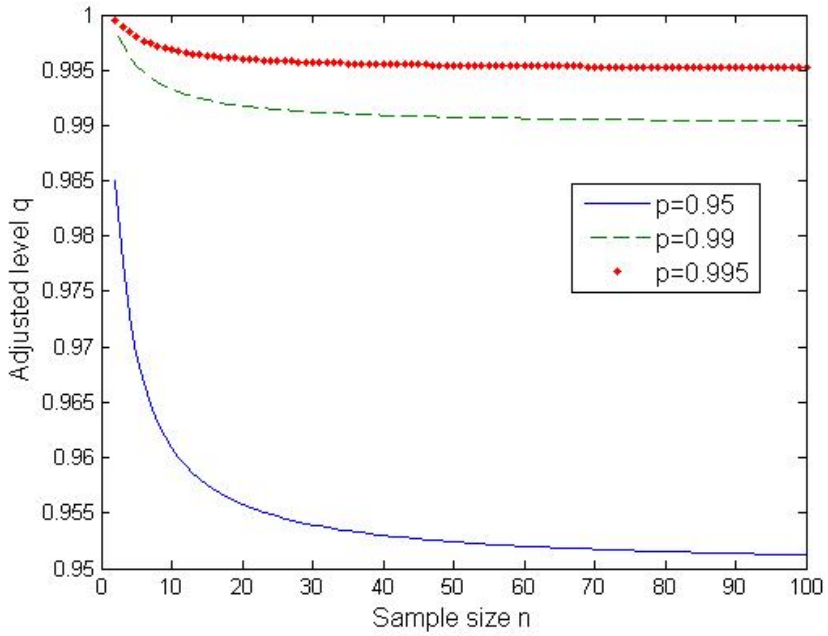

Figure 3.1: Confidence level q required to eliminate the residual estimation risk for a normal random with known scale parameter and risk measure TVaR_p.

level p. The difference $q - p$ is more pronounced for very small sample sizes, such that, if η_{adj} were adopted, portfolios with a longer history would be subject to a lower capital requirement. □

Example 8 demonstrates that, in order to eliminate residual estimation risk, the adjusted risk measure ρ_{adj} produces a more severe risk assessment than the risk measure ρ capturing regulatory preferences. However, the adjustment depends both on the family of distribution used and on the sample size. This means that the adjustment in the risk measure would be implausible in the context of a principles-based regulatory regime, as different financial firms, with different exposures and datasets, would have to calculate their capital requirement according to different risk measures imposed by the regulator. Hence, while the approach discussed in the present section is technically successful, it would be desirable to develop alternative methods where the risk measure is left unmodified and parameter uncertainty is reflected in the estimation procedure itself. The next two approaches follow

that route.

3.3.3 Capital set by a predictive distribution

The use of a Bayesian predictive distribution is a standard approach to parameter uncertainty, see Cairns (2000). Under a Bayesian approach, the parameter $\theta \in \Theta$ is considered a random variable itself with *prior* distribution $\pi(\theta)$. Once data $\mathbf{x}$ have been collected, the *posterior* of the parameter, $\pi(\theta|\mathbf{x})$, is obtained by

$$\pi(\theta|\mathbf{x}) \propto \pi(\theta) \prod_{i=1}^{n} f(x_i; \theta). \tag{3.14}$$

The *predictive* distribution of Y, given the data $\mathbf{x}$, is defined as

$$\hat{F}(\cdot|\mathbf{x}) = \int_{\theta \in \Theta} F(\cdot; \theta) \pi(\theta|\mathbf{x}) d\theta. \tag{3.15}$$

Probabilities and expectations calculated according to the predictive distribution are respectively denoted by $\hat{\mathbb{P}}(\cdot|\mathbf{x})$ and $\hat{\mathbb{E}}(\cdot|\mathbf{x})$.

We consider now whether evaluating the regulatory risk measure ρ with the predictive distribution $\hat{F}(\cdot|\mathbf{x})$ leads to an elimination of the residual risk. That is, we set

$$\eta_{bay}(\mathbf{X}) = \rho[\hat{F}(\cdot|\mathbf{X})] \tag{3.16}$$

and examine whether the residual risk $\mathrm{RR}(\theta, \eta_{bay}) = \rho(Y - \eta_{bay}(\mathbf{X}))$ is close to zero. Note the switch in (3.16) from a particular set of observations $\mathbf{x}$ to a random sample $\mathbf{X}$. This represents the fact that, while we are using a Bayesian technique, we are evaluating its performance by a frequentist criterion. In other words, in this paper we are dispensing with the deeper interpretation of Bayesian methods and consider $\hat{F}(\cdot|\mathbf{X})$ purely as an alternative estimator of the distribution $F(\cdot; \theta)$.

Using the capital estimator (3.16) is plausible, as a predictive distribution tends to be more dispersed than the MLE; as seen in (3.15) the predictive

distributions $\hat{F}(\cdot|\mathbf{x})$ emerges as the weighted average of $F(\cdot;\theta)$, over different values of θ. This increased volatility in the predictive distribution, which is more pronounced for small sample sizes n, will tend to increase the estimated capital and thus may compensate for the impact of parameter uncertainty. In fact, Gerrard and Tsanakas (2011) have shown that for distributions belonging to location-scale families (and also for distributions of random variables that are increasing transformations of random variables following location-scale distributions), it is $\mathbb{P}(Y > \mathrm{VaR}_p[\hat{F}(\cdot|\mathbf{X})]) = 1-p$, as long as the *probability matching prior* (Severini et al., 2002) $\pi(\mu,\sigma) = 1/\sigma$ is used. This implies that $\mathrm{VaR}_p(Y - \mathrm{VaR}_p[\hat{F}(\cdot|\mathbf{X})]) = 0$, such that the residual risk is completely eliminated.

For the more general class of risk measures considered here and for location-scale families, we can show that (a) when the scale parameter is known, residual risk is completely eliminated, and (b) when the location parameter is known, a quantity similar to residual risk equals zero. The proofs are given in Section 3.6.1 of the Appendix. For the case that both location and scale parameters are unknown, the simulation study of Section 3.3.5 provides some evidence that the Bayesian predictive distribution approach nearly eliminates residual estimation risk.

To clarify the above ideas, a simple example is presented below.

Example 9. Continuing from the previous examples, let Y, $\mathbf{X}$ again follow a normal distribution with unknown mean μ and standard deviation σ, such that $Y \sim F(y;\mu) = \Phi(\frac{y-\mu}{\sigma})$. Here we do not specify the risk measure used; it may be TVaR_p or indeed any other positive homogenous, translation invariant, and law-invariant risk measure ρ.

The improper prior $\pi(\mu) = 1$ is used. Using standard arguments (similar to Hogg et al. (2012) Example 11.3.1) this leads respectively to a normal posterior distribution $\pi(\mu|\mathbf{x}) \equiv \mathcal{N}(\hat{\mu}, \frac{\sigma^2}{n})$, where $\hat{\mu}$ is the sample mean of $\mathbf{X}$.

The predictive distribution can then be easily obtained as

$$\hat{F}(y|\mathbf{x}) = \int_{-\infty}^{\infty} \Phi\left(\frac{y-m}{\sigma}\right) \pi(m|\mathbf{x}) dm = \Phi\left(\frac{y-\hat{\mu}}{\sigma\sqrt{1+1/n}}\right).$$

Hence the predictive distribution is again normal $\hat{F}(\cdot|\mathbf{x}) \equiv \mathcal{N}(\hat{\mu}, \sigma^2(1+1/n))$, but it is more dispersed than the distribution estimated by MLE, which is just $F(\cdot;\hat{\mu}) \equiv \mathcal{N}(\hat{\mu}, \sigma^2)$.

A regulatory risk measure ρ is used, not necessarily TVaR. The capital estimator then becomes

$$\eta_{bay}(\mathbf{X}) = \rho[\hat{F}(\cdot|\mathbf{X})] = \hat{\mu} + \sigma\sqrt{1+\frac{1}{n}}\rho[\Phi].$$

We can now show that the residual risk when using $\eta_{bay}(\mathbf{X})$ becomes zero, which is consistent with Proposition 3.6.2 in Section 3.6.1. As before, we write $Y \overset{d}{=} \mu+\sigma Z$ and $\hat{\mu} \overset{d}{=} \mu+\frac{\sigma}{\sqrt{n}}U$, where Z, U are independent standard normal variables. We have

$$\begin{aligned}
\mathrm{RR}(\mu, \eta_{bay}) &= \rho(Y - \eta_{bay}(\mathbf{X})) \\
&= \rho\Big(\mu + \sigma Z - \mu - \frac{\sigma}{\sqrt{n}}U - \sigma\sqrt{1+\frac{1}{n}}\rho[\Phi]\Big) \\
&= \sigma\rho\Big(Z - \frac{1}{\sqrt{n}}U\Big) - \sigma\sqrt{1+\frac{1}{n}}\rho[\Phi]
\end{aligned}$$

Since $Z - \frac{1}{\sqrt{n}}U \sim \mathcal{N}(0, 1+1/n)$, it is $\rho\Big(Z - \frac{1}{\sqrt{n}}U\Big) = \sqrt{1+1/n}\rho[\Phi]$, yielding

$$\mathrm{RR}(\mu, \eta_{bay}) = \sigma\sqrt{1+\frac{1}{n}}\rho[\Phi] - \sigma\sqrt{1+\frac{1}{n}}\rho[\Phi] = 0.$$

Therefore, the residual risk is completely eliminated for any positive homogenous, translation invariant, and law-invariant risk measure ρ. □

3.3.4 Bootstrap estimation

In this section we propose to adjust the capital estimator by adding to $\eta(\mathbf{X})$, the MLE of risk measure ρ, an estimator of the residual risk under η. Hence, we are trying to correct $\eta(\mathbf{X})$, by estimating the residual risk $\mathrm{RR}(\theta, \eta)$ that it gives rise to. Insofar, the approach proposed in the current section is a form of parametric bootstrapping; for a rigorous treatment of the bootstrap see Hall (1992).

To start with, denote by

$$r_1(\theta) = \mathrm{RR}(\theta, \eta) \tag{3.17}$$

the residual estimation risk as a function of only the true parameter θ. As before, $\eta(\mathbf{X}) = \rho[F(\cdot; \hat{\theta})]$, where $\hat{\theta}$ is the MLE of θ. Then, an estimator of the residual risk itself, from the sample $\mathbf{X}$, is given by $r_1(\hat{\theta})$. Since we can interpret $r_1(\theta) = \mathrm{RR}(\theta, \eta)$ as the additional capital that needs to be subtracted from $Y - \eta(\mathbf{X})$ in order to make it acceptable, it is reasonable to propose the following *first order bootstrap* capital estimator

$$\eta_{bs1}(\mathbf{X}) = \rho[F(\cdot; \hat{\theta})] + r_1(\hat{\theta}). \tag{3.18}$$

The process can be repeated in order to refine the adjustment to the capital estimator. Let the residual risk arising from using the capital estimator η_{bs1} be

$$r_2(\theta) = \mathrm{RR}(\theta, \eta_{bs1}) \tag{3.19}$$

Consequently, we can define the *second order bootstrap* capital estimator as

$$\eta_{bs2}(\mathbf{X}) = \eta_{bs1}(\mathbf{X}) + r_2(\hat{\theta}) = \rho[F(\cdot; \hat{\theta})] + r_1(\hat{\theta}) + r_2(\hat{\theta}), \tag{3.20}$$

and the associated residual risk by

$$r_3(\theta) = \mathrm{RR}(\theta, \eta_{bs2}) \tag{3.21}$$

The process can be further repeated in order to derive bootstrap capital estimators of higher orders.

Higher order bootstrap iterations are typically computationally expensive, as they rely on nested simulations. Let $\hat{\theta}$ follow distribution $G(\cdot;\theta)$. A naive simulation-based algorithm for evaluating η_{bs2} in (3.20) would then require us to follow these steps:

1. Calculate $\hat{\theta}$ from $\mathbf{X}$ and evaluate $\rho[F(\cdot;\hat{\theta})]$.

2. i) Simulate m samples from the pair $(Y^*, \hat{\theta}^*)$ defined as $Y^*|\hat{\theta} \sim F(\cdot;\hat{\theta})$, $\hat{\theta}^*|\hat{\theta} \sim G(\cdot;\hat{\theta})$.

   ii) For each simulated pair $(Y_i^*, \hat{\theta}_i^*),\ i = 1,\ldots,m$, evaluate $s_i^* = Y_i^* - \rho[F(\cdot;\hat{\theta}_i^*)]$.

   iii) Let H^S be the empirical distribution of the sample $s_1^*,\ldots,s_m^*$. Set $r_1(\hat{\theta}) = \rho[H^S]$.

3. i) For each simulated value $\hat{\theta}_i^*$, simulate m values from the pair $(Y^{**}, \hat{\theta}^{**})$ defined as $Y^{**}|\hat{\theta}_i^* \sim F(\cdot;\hat{\theta}_i^*)$, $\hat{\theta}^{**}|\hat{\theta}_i^* \sim G(\cdot;\hat{\theta}_i^*)$.

   ii) For each simulated pair $(Y_{ik}^{**}, \hat{\theta}_{ik}^{**}),\ k = 1,\ldots,m$, evaluate $s_{ik}^{**} = Y_{ik}^{**} - \rho[F(\cdot;\hat{\theta}_{ik}^{**})]$.

   iii) Let H_i^S be the empirical distribution of the sample $s_{i1}^{**},\ldots,s_{im}^{**}$. Set $r_1(\hat{\theta}_i^*) = \rho[H_i^S]$.

   iv) Evaluate $w_i^* = s_i^* - r_1(\hat{\theta}_i^*),\ i = 1,\ldots,m$.

   v) Let H^W be the empirical distribution of the sample $w_1^*,\ldots,w_m^*$. Set $r_2(\hat{\theta}) = \rho[H^W]$.

4. Set $\eta_{bs2}(\mathbf{X}) = \rho[F(\cdot;\hat{\theta})] + r_1(\hat{\theta}) + r_2(\hat{\theta})$.

For the location-scale families studied in this section, this calculation can be substantially simplified, avoiding the need for nested simulations. Let $Y \overset{d}{=} \mu + \sigma Z \sim F(\cdot;(\mu,\sigma))$ belong to a location-scale family, with $Z \sim F$. Noting the representation (3.8) for the MLEs $(\hat{\mu},\hat{\sigma})$, we can show that

$$\eta_{bs1}(\mathbf{X}) = \hat{\mu} + \hat{\sigma}\big(\rho(Z) + \rho(Z - U - V\rho(Z))\big), \tag{3.22}$$

$$\eta_{bs2}(\mathbf{X}) = \hat{\mu}+\hat{\sigma}\Big[\rho(Z)+\rho(Z-U-V\rho(Z))+\rho\big(Z-U-V(\rho(Z)-\rho(Z-U-V\rho(Z)))\big)\Big]. \tag{3.23}$$

Formulas (3.22) and (3.23) are derived in Section 3.6.1 of the Appendix. Since the distribution of the random variables Z, U, V does not depend on the true parameters (μ, σ), formulas (3.22) and (3.23) can be evaluated from a single set of simulated values from Z, U, V.

In Section 3.3.5, it is shown via a simulation study that the residual estimation risk decreases with repeated applications of the parametric bootstrap, such that it is $\mathrm{RR}(\theta, \eta) \geq \mathrm{RR}(\theta, \eta_{bs1}) \geq \mathrm{RR}(\theta, \eta_{bs2})$. Moreover, it is proved in Section 3.6.1 that, in the particular case where the scale parameter is known, the first order bootstrap capital estimator actually reduces the residual risk to zero, that is, $\mathrm{RR}(\theta, \eta_{bs1}) = 0$. This result is demonstrated via the simple normal example that follows.

Example 10. We continue from previous examples of a normal distribution with an unknown mean and a risk measure ρ. Similarly to Example 7, it is:

$$\begin{aligned} r_1(\mu) &= \rho(\mu + \sigma Z - \hat{\mu} - \sigma\rho[\Phi]) \\ &= \sigma\rho\left(Z - \frac{1}{\sqrt{n}}U\right) - \sigma\rho[\Phi] \\ &= \sigma\left(\sqrt{1+\frac{1}{n}} - 1\right)\rho[\Phi]. \end{aligned}$$

As $r_1(\mu)$ does not depend on the true value of the location parameter, it is $r_1(\mu) = r_1(\hat{\mu})$ and we can write

$$\eta_{bs1}(\mathbf{X}) = \hat{\mu} + \sigma\rho[\Phi] + \sigma\left(\sqrt{1+\frac{1}{n}} - 1\right)\rho[\Phi] = \hat{\mu} + \sigma\sqrt{1+\frac{1}{n}}\rho[\Phi].$$

Note that η_{bs1} above is identical to η_{bay} in Example 9, which we know eliminates residual risk. Hence it also is $\mathrm{RR}(\mu, \eta_{bs1}) = 0$.

3.3.5 Simulation study

In this section, we present simulation results for the normal and the exponential distributions. The risk measure $\rho \equiv \text{TVaR}_p$ is used, and we investigate the residual risk arising from using the capital estimator $\eta(\mathbf{X})$ as well as the alternative capital estimators derived in Sections 3.3.3 and 3.3.4. (The adjustment of Section 3.3.2 is not discussed, as it guarantees that $\text{RR}(\theta, \eta_{adj}) = 0$.)

To quantify residual risk independently of the value of the scale parameter, we report the following normalised quantity

$$\frac{\text{RR}(\theta, \eta)}{\rho(Y) - \mathbb{E}[Y]}, \tag{3.24}$$

representing residual risk as a percentage of pure risk capital. For location-scale families the value of (3.24) does not depend on the unknown parameters and the same is true when η is changed to η_{bay}, η_{bs1}, and η_{bs2}.

The risk measure TVaR_p is used with confidence levels $p \in \{0.95, 0.99, 0.995\}$ and different sample sizes $n \in \{10, 20, 50, 100\}$ are considered.

The normal distribution is considered with both location and scale parameters unknown. The MLEs are given by $\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \hat{\mu})^2$.

The exponential distribution is given by $F(y;\theta) = 1 - \exp(-y/\theta),\ y \geq 0$, where $\theta = E(Y)$ is an unknown scale parameter. The MLE is the sample mean $\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} X_j$. It can be easily shown that the TVaR_p of an exponential variable Y equals $\text{TVaR}_p(Y) = \theta(1 - \log(1-p))$.

A Monte Carlo sample size of $m = 10^7$ is used. Convergence is improved via a simple importance sampling procedure, whereby, when simulating Y, $0.9m$, samples are drawn from the conditional distribution of $Y|Y > \text{VaR}_{0.9}(Y)$ and 0.1m samples are drawn from $Y|Y \leq \text{VaR}_{0.9}(Y)$.

Residual risk for $\eta(\mathbf{X})$

Table 3.1 provides results for the normal distribution. The values obtained demonstrate the sensitivity of residual risk on the sample size. In particular,

for very small sample sizes the impact of parameter uncertainty is substantial. Thus, when $n = 10$ the residual risk is between 21% and 29% of the required capital and for a somewhat larger sample of $n = 50$, the residual risk takes values around 5% of capital. Even for a sample of $n = 100$ the residual risk does not reach zero. Consistent results are obtained for an exponential distribution, seen in Table 3.2.

From Example 7 we can see that, in the case that the scale parameter σ is known and only the location parameter μ is estimated, the normalised residual estimation risk equals to $\sqrt{1 + 1/n} - 1$, which is independent of the confidence level p. For the sample sizes $n \in \{10,\ 20,\ 50,\ 100\}$ used, this formula give normalised residuals risks of $\{0.049,\ 0.0247,\ 0.010,\ 0.005\}$ respectively. Note that these numbers are substantially lower than the corresponding figures reported in Table 3.1, implying that estimation of the scale parameter has a substantially higher impact on residual risk than estimation of the location parameter.

Residual risk for $\eta_{bay}(\mathbf{X})$

Here we demonstrate numerically the performance of the predictive distribution approach of Section 3.3.3.

For a normal distribution, with prior $\pi(\mu, \sigma) = 1/\sigma$, a standard argument (similar to Hogg et al. (2012), Example 11.3.1) shows that the predictive distribution is a student-t distribution

$$\hat{F}(y|\mathbf{X}) = t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\frac{y-\hat{\mu}}{\hat{\sigma}}\right), \tag{3.25}$$

where $\hat{\mu},\ \hat{\sigma}$ are the MLEs of $\mu,\ \sigma$, and t_{n-1} is the distribution function of a standard t variable with $n - 1$ degrees of freedom. The corresponding value

of TVaR is (McNeil et al. (2005), Example 2.19)

$$\begin{aligned}\eta_{bay}(\mathbf{X}) &= \mathrm{TVaR}_p[\hat{F}(y|\mathbf{X})] \\ &= \hat{\mu} + \hat{\sigma}\sqrt{\frac{n+1}{n-1}}\left(\frac{g_{n-1}(t_{n-1}^{-1}(p))}{1-p}\right)\left(\frac{n-1+(t_{n-1}^{-1}(p))^2}{n-2}\right),\end{aligned} \tag{3.26}$$

where g_{n-1} is the density of a standard t variable with $n-1$ degrees of freedom.

Consider now an exponential distribution. Using the prior $\pi(\theta) = \frac{1}{\theta}$, the predictive distribution is a Pareto distribution (Gerrard and Tsanakas (2011), Example 9) with

$$\hat{F}(y|\mathbf{X}) = 1 - \left(\frac{n\hat{\theta}}{y+n\hat{\theta}}\right)^n, \tag{3.27}$$

where $\hat{\theta}$ is the MLE of θ. It is easily shown that the corresponding value of TVaR is

$$\eta_{bay}(\mathbf{X}) = \mathrm{TVaR}_p[\hat{F}(y|\mathbf{X})] = n\hat{\theta}\left(\frac{n}{n-1}(1-p)^{-1/n} - 1\right). \tag{3.28}$$

Thus in both the normal and exponential cases, the respective t and Pareto predictive distributions are heavy tailed, which leads to higher estimated capital levels.

In Tables 3.3 and 3.4 we can see that using η_{bay} is a very effective method for reducing residual estimation risk. Even with only 10 data points, the residual estimation risk is nearly eliminated, reducing by more than an order of magnitude in comparison to the results for η reported in Tables 3.1 and 3.2.

Residual risk for $\eta_{bs1}(\mathbf{X})$ and $\eta_{bs2}(\mathbf{X})$

As was seen in Section 3.3.4, the bootstrap correction to the risk measure for location-scale families can be applied several times without increase in computing time. Here we report the residual estimation risk of the first

and second order bootstrap capital estimator for exponential and normal distributions.

Tables 3.5 and 3.6 report the normalised residual estimation risk for the two distributions examined. The first order bootstrap capital estimator η_{bs1} works somewhat better for the normal than for the exponential distribution. However, in both cases, the second order estimator η_{bs2} gives satisfactory results, comparable to the results with η_{bay} reported in Tables 3.3 and 3.4. It may be convenient to use a higher order bootstrap approach in comparison to a Bayesian predictive distribution capital estimator, as for more complex models predictive distributions are not always available in closed form.

3.4 Beyond location-scale families

3.4.1 Transformed location-scale families

Section 3.3 outlined methods for controlling residual estimation risk, when Y follows a location-scale family, such as the normal distribution. However, location-scale families are not always appropriate modelling choices. For example, insurance claims or operational risk losses can often be modelled via distributions that can be said to belong to a *transformed location-scale family*.

Specifically, let as before $Y \sim F(\cdot; (\mu, \sigma))$ belong to a location-scale family and consider a strictly increasing function h that is well defined on the range of Y. Then the random variable Y' follows a *transformed location-scale family*. We write $Y' \sim F'(\cdot; (\mu, \sigma)) \equiv F\big(\frac{h^{-1}(\cdot)-\mu}{\sigma}\big)$. Also, the VaR of the transformed variable is simply given by $\mathrm{VaR}_p(Y') = h(\mathrm{VaR}_p(Y))$.

The most common choice of transformation function is an exponential, $h(x) = \exp(x)$. This transformation is used in financial risk management to move from asset returns to prices. However, as risk measures are typically applied on returns rather than prices, transformed location-scale families are more relevant for insurance rather than finance applications. For example,

using an exponential transformation, a normal variable Y becomes a log-normal one Y'. The log-normal distribution is one of the most common insurance loss models used in practice. Similarly, an exponential variable transforms into a Pareto distributed one. The Pareto distribution is a popular model for tails of distributions, supported by theoretical arguments from extreme value theory, see for instance McNeil et al. (2005), Section 7.2.4. Additional examples are the Weibull distribution, which emerges through an exponential transformation of a Negative Gumbel distributed random variable, and the log-logistic or Champernowne distribution, which arises from transforming a Logistic variable and has been used for modelling insurance and operational risks, see Klugman et al. (2008), Guillen et al. (2007).

The log-normal and (single-parameter) Pareto distributions will be the focus of our examples throughout Section 3.4.

The methods for controlling residual estimation risk that were discussed in Section 3.3 are generally not applicable for transformed location scale families. In particular, an equation similar to (3.13) does not exist, so it is not possible to eliminate residual risk without knowledge of the true parameters. Hence the approach of adjusting the risk measure as in Section 3.3.2 is no longer applicable. Nonetheless, a heuristic modification of that method is presented in Section 3.4.3. A related problem concerns the application of the bootstrap corrections of Section 3.3.4; they can still be performed but at the cost of computational expense, since nested simulations cannot be avoided. Finally, the Bayesian approach of Section 3.3.3 is still applicable to the case of transformed location-scale families. However the performance of the Bayesian method is no longer directly supported by theoretical results and will be illustrated by numerical examples in Section 3.4.4.

3.4.2 Heavy tails and coherent risk measures

Before considering methods for reducing the residual risk for the log-normal and Pareto distributions, it is necessary to discuss the substantial problems

that may arise from the tail-behaviour of these distributions. In particular, we find that trying to quantify the residual estimation risk and to apply the methods of Section 3.3 can lead to distribution functions with infinite means, which do not allow for the evaluation of coherent risk measures such as TVaR. It is shown by Delbaen (2002) that for a law-invariant coherent risk measure ρ and a random variable $Y' \in \mathcal{X}$ it always is $\rho(Y') \geq E[Y']$, such that the risk measure is not well defined when the mean is not finite.

We introduce the ideas by two examples.

Example 11. Let $Y, \mathbf{X}$ follow an exponential distribution with mean θ and define the transformed variables $Y' = \exp(Y)$, $\mathbf{X}' = (\exp(X_1), \ldots, \exp(X_n))$. Then $Y', \mathbf{X}'$ follow a one-parameter Pareto with distribution function

$$F'(y; \theta) = 1 - y^{-1/\theta}, \quad y \geq 1.$$

The Pareto distribution has a finite mean if and only if the parameter $\theta < 1$; in that case the mean equals $E(Y) = \frac{1}{1-\theta}$. (More generally, the k^{th} moment is finite for $\theta < k$.) The MLE of θ is given by

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} \log(X_i')$$

and follows a Gamma distribution, $\hat{\theta} \sim \Gamma(n, \theta/n)$, with mean $E(\hat{\theta}) = n(\theta/n) = \theta$. This means that there is a non-zero probability that an outcome $\{\hat{\theta} \geq 1\}$ is observed, hence

$$\mathbb{P}(\hat{\theta} \geq 1) > 0.$$

For such outcomes of $\hat{\theta}$ the estimated distribution $F'(\cdot; \hat{\theta})$ will have an infinite mean. Therefore, if a coherent risk measure ρ is used, there are outcomes of $\hat{\theta}$ for which the capital estimator $\eta(\mathbf{X}') = \rho[F'(\cdot; \hat{\theta})]$ is not well defined. Thus, the residual risk

$$\mathrm{RR}(\theta, \eta) = \rho(Y' - \rho[F'(\cdot; \hat{\theta})])$$

will also not be well defined. This also creates a problem with calculating the estimate of residual risk $r_1(\hat{\theta})$ as defined in Section 3.3.4, which means that the derivation of bootstrap capital estimators is also not feasible. □

Similar situations to that of Example 11 arise more generally with distributions that have regularly varying tails, see Embrechts et al. (1997), p. 37. For all such distributions, the mean becomes infinite for a given range of a parameter called the *tail index*. Moreover it has been shown that estimated distributions with infinite means often appear in real-world applications, such as operational risk modelling (Nešlehová et al., 2006). We note that such problems are not specific to our definition of residual estimation risk. For instance, if one was interested in working out the bias of the capital estimator, it would be necessary to evaluate $E(\rho[F'(\cdot;\hat{\theta})])$, which again is not well defined.

A related problem can occur, when applying coherent risk measures on Bayesian predictive distributions, which can also turn out to have infinite means.

Example 12. Let $Y, \mathbf{X} \sim \mathcal{N}(\mu, \sigma^2)$ and again $Y' = \exp(Y)$, $\mathbf{X}' = (\exp(X_1), \ldots, \exp(X_n))$, such that $Y', \mathbf{X}' \sim \mathcal{LN}(\mu, \sigma^2)$. For the log-normal distribution all moments exist, regardless of the value of the parameters, such that the problems reported in Example 11 do not appear, i.e. for a risk measure such as TVaR, the quantity $\rho[F'(\cdot;\hat{\theta})]$ will always be well defined.

Consider now capital being set using the predictive distribution of Y', such that $\eta_{bay}(\mathbf{X}') = \rho[\hat{F}'(\cdot|\mathbf{X}')]$. From equation (3.25) we know that the predictive distribution of the normal variable Y is a student-t distribution with $n-1$ degrees of freedom. It follows that the predictive distribution of the log-normal variable Y' is a "log-t" distribution (see Lemma 1ii) in Gerrard and Tsanakas (2011); a more detailed discussion of this transformation is given in Section 3.4.3). This means that, under its predictive distribution $\hat{F}'(\cdot|\mathbf{X}')$, Y' is a random variable the logarithm of which follows a t

distribution. Thus the predictive distribution of Y' takes the form

$$\hat{F}'(y|\mathbf{X}') = t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\frac{\log(y)-\hat{\mu}}{\hat{\sigma}}\right),$$

where $\hat{\mu}, \hat{\sigma}^2$ are still the sample mean and variance of the normal sample $\mathbf{X}$. This implies that the expected value associated with $\hat{F}'(\cdot|\mathbf{X}')$ is:

$$\hat{E}(Y'|\mathbf{X}') = \hat{E}\left(\exp(Y)|\mathbf{X}\right).$$

However, since the t distribution has a regularly varying tail (see McNeil et al. (2005), p. 293), its moment generating function is not well defined (see Embrechts et al. (1997), p. 50), implying that $\hat{E}(\exp(Y)|\mathbf{X}) = \infty$. Since the mean associated with the predictive distribution $\hat{F}'(\cdot|\mathbf{X}')$ is infinite, any capital estimator of the form $\rho[\hat{F}'(\cdot|\mathbf{X}')]$ will also be infinite, when a coherent risk measure ρ is used. □

Coherent risk measures are seen as theoretically superior, by their full consideration of the extreme tails of distributions. However, it seems that precisely this feature becomes a drawback when dealing with heavy-tailed distributions in the presence of parameter uncertainty. Recent literature has stressed this point; in particular, Cont et al. (2010) show that coherent risk measures such as TVaR are less robust to data contamination than VaR. While our context is different to Cont et al. (2010), we adopt their suggestion of using the TTVaR risk measure as defined in equation (3.3). The risk measure TTVaR_{p_1,p_2}, by ignoring the tails of loss distributions beyond VaR_{p_2}, does not require the distribution to have a finite mean. Thus the problems encountered in Examples 11 and 12 would not appear if TTVaR was used rather than TVaR.

For this reason, for the rest of Section 3.4, the risk measure TTVaR is used, such that $\rho \equiv \text{TTVaR}_{p_1,p_2}$.

Finally, we remark that the problems of heavy-tailedness are not specific to transformed location-scale families. There certainly exist location-scale

families that are heavy tailed, and transformed location-scale families that are not. The reason that the discussion of heavy tails occurs at this place is driven by the properties of the log-normal and Pareto models that we are using as examples throughout Section 3.4.

3.4.3 Reducing residual estimation risk

Two methods for reducing residual risk for transformed location-scale families are used here, adjusting the methods of Sections 3.3.2 and 3.3.3. The bootstrapping method of Section 3.3.4 is not considered, due to the computational expense involved in nested simulations.

Heuristically adjusting the TTVaR risk measure

As discussed in Section 3.4.1, for transformed location scale families the residual risk depends on both parameters and cannot be easily eliminated using the method of Section 3.3.2. In other words, it is not straightforward to find q_1, q_2 such that for the capital estimator

$$\eta_{adj}(\mathbf{X}') = \mathrm{TTVaR}_{q_1,q_2}[F'(\cdot;\hat{\theta})] \tag{3.29}$$

it is

$$\mathrm{RR}(\theta,\eta) = \mathrm{TTVaR}_{p_1,p_2}(Y' - \mathrm{TTVaR}_{q_1,q_2}[F'(\cdot;\hat{\theta})]) = 0. \tag{3.30}$$

Instead we use a heuristic argument. Since TTVaR_{p_1,p_2} captures information about the distribution of Y' between $\mathrm{VaR}_{p_1}(Y')$ and $\mathrm{VaR}_{p_2}(Y')$, we propose deriving $q_1,\ q_2$ from solving instead:

$$\mathrm{VaR}_{p_i}(Y' - \mathrm{VaR}_{q_i}[F'(\cdot,\hat{\theta})]) = 0 \qquad \text{for } i = 1,2. \tag{3.31}$$

For the distributions that we consider in this section, (3.31) can be solved exactly. As before, let $Y' = h(Y)$, where h is a strictly increasing function

and Y is distributed according to a location-scale family. Then it is

$$\begin{aligned}
&\mathrm{VaR}_{p_i}(Y' - \mathrm{VaR}_{q_i}[F'(\cdot;\hat{\theta})]) = 0 \Leftrightarrow \\
&\mathrm{VaR}_{p_i}(h(Y) - h(\mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})])) = 0 \Leftrightarrow \\
&\mathbb{P}(h(Y) - h(\mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})]) \leq 0) = p_i \Leftrightarrow \\
&\mathbb{P}(Y - \mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})] \leq 0) = p_i \Leftrightarrow \\
&\mathrm{VaR}_{p_i}(Y - \mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})]) = 0
\end{aligned} \tag{3.32}$$

Since Y is in a location-scale family, we can solve $\mathrm{VaR}_{p_i}(Y - \mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})]) = 0$ for q_i, using the arguments of Section 3.3.2.

Application of a Bayesian predictive distribution

The predictive distribution for transformed location-scale families is derived by a straightforward transformation of the predictive distribution for the corresponding location-scale family. In particular, following Lemma 1ii) in Gerrard and Tsanakas (2011), we know that the posterior distribution of the parameter θ is the same, regardless of conditioning on the sample $\mathbf{X}$ or $\mathbf{X}'$, that is, $\pi(\theta|\mathbf{X}') = \pi(\theta|\mathbf{X})$. It follows that:

$$\begin{aligned}
\hat{F}'(y|\mathbf{X}') &= \int_{\theta\in\Theta} F'(y;\theta)\pi(\theta|\mathbf{X}')d\theta \\
&= \int_{\theta\in\Theta} F(h^{-1}(y);\theta)\pi(\theta|\mathbf{X})d\theta \\
&= \hat{F}(h(y)^{-1}|\mathbf{X}).
\end{aligned} \tag{3.33}$$

Consequently, the VaRs of the two predictive distributions are related by

$$\mathrm{VaR}_p[\hat{F}'(\cdot|\mathbf{X}')] = h(\mathrm{VaR}_p[\hat{F}(\cdot|\mathbf{X})]), \tag{3.34}$$

such that the TTVaR risk measure can be evaluated by

$$\begin{aligned}\eta_{bay}(\mathbf{X}') &= \mathrm{TTVaR}_{p_1,p_2}[\hat{F}'(\cdot|\mathbf{X}')] \\ &= \frac{1}{p_2 - p_1}\int_{p_1}^{p_2} h(\mathrm{VaR}_u[\hat{F}(\cdot|\mathbf{X})])du.\end{aligned} \tag{3.35}$$

3.4.4 Simulation study

In this section, simulation results are presented for the residual estimation risk, when the losses Y', $\mathbf{X}'$ follow the log-normal and Pareto distributions of Examples 12 and 11 respectively. For the log-normal distribution, the parameter choices $(\mu, \sigma) \in \{(4.6002, 0.0998), (4.5856, 0.1980), (4.4936, 0.4724))\}$ are used, corresponding to the same mean $E(Y') = 100$ but different coefficients of variation $\mathrm{CV}(Y') = \sqrt{\mathrm{Var}(Y')}/E(Y') \in \{0.1, 0.2, 0.5\}$. For the Pareto distribution, parameter values $\theta \in \{0.1, 0.25, 0.5\}$ are used. The Pareto becomes more heavy-tailed as θ increases; here $\theta = 0.25$ corresponds to a situation where the fourth moment (kurtosis) is not well defined, while for $\theta = 0.5$ even the second moment (variance) becomes infinite.

The risk measure TTVaR_{p_1,p_2} is used throughout, with $p_1 \in \{0.95,0.99,0.995\}$ and $p_2 = 0.997$.

Similarly to Section 3.3.5, we report the residual estimation risk as a percentage of the true pure risk capital. The capital estimators used are (a) the MLE $\eta(\mathbf{X}')$, (b) the estimator $\eta_{adj}(\mathbf{X}')$ derived by adjusting the TTVaR measure as in Section 3.4.3, and (c) the estimator $\eta_{bay}(\mathbf{X}')$ following from the Bayesian predictive distribution of Section 3.4.3.

As in Section 3.3.5, a Monte-Carlo sample of size $m = 10^7$ combined with a simple importance sampling scheme is used.

Residual risk for $\eta(\mathbf{X}') = \mathrm{TTVaR}_{p_1,p_2}[F'(\cdot;\hat{\theta})]$

The normalised residual estimation risks for the log-normal and Pareto distributions are presented in Tables 3.7 and 3.8 respectively. Similar to the results for the normal and exponential distribution in Tables 3.1 and 3.2,

the impact of parameter uncertainty on residual estimation risk is substantial. As opposed to examples drawn from location-scale families, there is now substantial dependence of the normalised residual risk on the shape parameters of the log-normal and Pareto distributions, σ and θ respectively. In particular, higher values of σ and θ imply heavier tails and produce higher residual estimation risks. Residual risk also increases with p_1, as the impact of parameter uncertainty becomes more pronounced for risk measures that focus further on the tail.

Residual risk for $\eta_{adj}(\mathbf{X}') = \mathrm{TTVaR}_{q_1,q_2}[F'(\cdot;\hat{\theta})]$

Following equation (3.32) in Section 3.4.3, we need to determine the adjusted probability levels q_1, q_2, such that for $i \in \{1,2\}$ it is

$$\mathbb{P}(Y - \mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})] \leq 0) = p_i \Leftrightarrow \mathrm{VaR}_{p_i}(Y - \mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})]) = 0, \quad (3.36)$$

where $Y = h^{-1}(Y')$ follows a location scale family, $Y \sim F(\cdot;\theta)$.

When Y' is log-normally distributed, Y is normal. For a normal distribution, it is shown in Example 5 of Gerrard and Tsanakas (2011) that

$$\mathbb{P}(Y - \mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})] \leq 0) = t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\Phi^{-1}(q_i)\right). \quad (3.37)$$

Therefore,

$$t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\Phi^{-1}(q_i)\right) = p_i \Leftrightarrow q_i = \Phi\left(\sqrt{\frac{n+1}{n-1}}t_{n-1}^{-1}(p_i)\right), \quad i \in \{1,2\}. \quad (3.38)$$

On the other hand, when Y' is Pareto distributed, Y is exponential. For an exponential distribution, it is shown in Example 4 of Gerrard and Tsanakas (2011) that

$$\mathbb{P}(Y - \mathrm{VaR}_{q_i}[F(\cdot;\hat{\theta})] \leq 0) = 1 - \left(1 - \frac{1}{n}\ln(1-q_i)\right)^{-n}. \quad (3.39)$$

Therefore,

$$1 - \left(1 - \frac{1}{n}\ln(1-q_i)\right)^{-n} = p_i \Leftrightarrow \tag{3.40}$$
$$q_i = 1 - \exp\left(-n\left((1-p_i)^{-1/n} - 1\right)\right), \quad i \in \{1,2\}.$$

The normalised residual estimation risks for the log-normal and Pareto distributions are presented in Tables 3.9 and 3.10 respectively. It can be seen that the heuristic method followed here effects a substantial reduction in residual risk, comparing with the results for MLE estimation in Tables 3.7 and 3.8. In contrast to the results for MLE, it is notable that the residual risk actually decreases with p_1. The reason is that, for fixed $p_2 = 0.997$, as p_1 approaches p_2, the TTVaR_{p_1,p_2} risk measure approaches VaR_{p_2} and we know from Gerrard and Tsanakas (2011) that for a VaR risk measure, such an adjustment to the confidence level exactly eliminates residual risk.

Residual risk for $\eta_{bay}(\mathbf{X}') = \text{TTVaR}_{p_1,p_2}[\hat{F}'(\cdot|\mathbf{X}')]$

When Y' follows a log-normal distribution, from the predictive distribution of the normal (3.25), the predictive distribution of Y' is:

$$\hat{F}'(y|\mathbf{X}') = t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\frac{\log(y) - \hat{\mu}}{\hat{\sigma}}\right). \tag{3.41}$$

As discussed in Example 12, this is a "log-t" distribution, with infinite mean. For the VaR and TTVaR measures of Y' we have, following equations (3.34) and (3.35) in Section 3.4.3,

$$\begin{aligned}\text{VaR}_p[\hat{F}'(y|\mathbf{X}')] &= \exp\left(\text{VaR}_p[\hat{F}(y|\mathbf{X})]\right) \\ &= \exp\left(\hat{\mu} + \hat{\sigma}\sqrt{\frac{n+1}{n-1}}t_{n-1}^{-1}(p)\right),\end{aligned} \tag{3.42}$$

$$\text{TTVaR}_{p_1,p_2}[\hat{F}'(y|\mathbf{X}')] = \frac{1}{p_2 - p_1}\int_{p_1}^{p_2}\exp\left(\hat{\mu} + \hat{\sigma}\sqrt{\frac{n+1}{n-1}}t_{n-1}^{-1}(u)\right)du. \tag{3.43}$$

When Y' follows a Pareto distribution, from the predictive distribution of the exponential (3.27), the predictive distribution of Y' is:

$$\hat{F}'(y|\mathbf{X}') = 1 - \left(\frac{n\hat{\theta}}{\log(y) + n\hat{\theta}}\right)^n. \tag{3.44}$$

That is a "log-Pareto" distribution, again with infinite mean. For the VaR and TTVaR measures of Y' we now have,

$$\begin{aligned}\mathrm{VaR}_p[\hat{F}'(y|\mathbf{X}')] &= \exp\left(\mathrm{VaR}_p[\hat{F}(y|\mathbf{X})]\right) \\ &= \exp\left(\hat{\theta}n((1-p)^{-1/n} - 1)\right),\end{aligned} \tag{3.45}$$

$$\mathrm{TTVaR}_{p_1,p_2}[\hat{F}'(y|\mathbf{X}')] = \frac{1}{p_2 - p_1}\int_{p_1}^{p_2} \exp\left(\hat{\theta}n((1-u)^{-1/n} - 1)\right) du. \tag{3.46}$$

The integrals in (3.43) and (3.46) can be easily solved numerically.

The normalised residual estimation risks for the log-normal and Pareto distributions are presented in Tables 3.11 and 3.12 respectively. The use of the predictive distribution is effective, producing a better improvement than that of the heuristic adjustment method in Tables 3.9 and 3.10. Again residual risk actually decreases with p_1, which can be explained by the fact that for a VaR risk measure, using a predictive distribution to set capital exactly eliminates residual risk (Gerrard and Tsanakas, 2011).

3.5 Conclusions

In the present contribution, we have introduced a method for measuring the impact of parameter uncertainty for risk measures based on a frequentist approach. For location-scale families we have shown how the dependence of this residual estimation risk on the true parameters may be eliminated. We investigated three approaches to reduce the residual estimation risk: adjusting the risk measure, using Bayesian predictive distributions and a parametric bootstrap procedure. We have seen how all these methods work well for

location-scale families. We showed that for heavy-tailed distributions, the residual estimation risk cannot be defined nor controlled properly when a law-invariant coherent risk measure is used. Hence we consider the use of the risk measure TTVaR. In particular, for transformed location-scale distributions, we investigate the effectiveness of adjusting the solvency capital requirement and the Bayesian techniques. Numerical results are presented for both location-scale families and transformed location-scale families. For location families the residual estimation risk is eliminated completely while in all the other cases we obtain a substantial reduction of the residual estimation risk.

3.6 Appendix

3.6.1 Formal results

Results relating to Section 3.3.3

We reformulate without proof, the content of Prop. 1 in Severini et al. (2002) which is used in the present section. For the sake of simplicity, details about the technical conditions are omitted, but Example 1 in Severini et al. (2002), implies that location-scale families satisfy all the necessary conditions to apply the proposition.

Proposition 3.6.1. *Severini et al. (2002). Let H($\boldsymbol{X}$) be a region such that*

$$\hat{\mathbb{P}}(Y \in H(\boldsymbol{X})|\boldsymbol{x}) = 1 - \alpha.$$

Let H satisfy the following conditions:

(i) H is invariant, ie for each $\theta = (\mu, \sigma) \in \Theta$, $y \in H(\boldsymbol{x})$ if and only if

$$y + \mu \in H(\mu + \boldsymbol{x}) \qquad \text{for location models,}$$

and

$$\sigma y \in H(\sigma \boldsymbol{x}) \qquad \text{for scale models.}$$

(ii) Let $C(\boldsymbol{x}, y) = 1$ *if* $y \in H(\boldsymbol{x})$ *and* 0 *otherwise, such that*

$$\hat{E}[C(\boldsymbol{X}, Y)|\boldsymbol{x}] = 1 - \alpha.$$

It follows that

$$E_\theta[C(\boldsymbol{X}, Y)] = 1 - \alpha.$$

Consider first a location family with parameter θ. The prior $\pi(\theta) = 1$ is used. It is known that (eg see Gerrard and Tsanakas (2011)), if $\mathbf{X} = \mathbf{Z} + b$, where $\mathbf{Z} = (Z_1, \ldots, Z_n)$ and $\mathbf{Z} \sim F$, then

$$\hat{F}(y|\mathbf{z} + b) = \hat{F}(y - b|\mathbf{z}).$$

Therefore,

$$\rho[\hat{F}(\cdot|\mathbf{x})] = \rho[\hat{F}(\cdot - b|\mathbf{z})] = \rho[\hat{F}(\cdot|\mathbf{z})] + b,$$

due to the translation invariance property of $\rho(\cdot)$.

Proposition 3.6.2 shows how using the predictive distribution eliminates residual risk for location families.

Proposition 3.6.2. *For location families, using the capital estimator* $\eta_{bay}(\boldsymbol{X}) = \rho[\hat{F}(\cdot|\boldsymbol{X})]$ *yields*

$$\rho(Y - \eta_{bay}(\boldsymbol{X})) = 0.$$

Proof. The proof follows from an application of Prop. 3.6.1. Consider the predictive region

$$H_c(\mathbf{X}) = (-\infty,\ \rho[\hat{F}(\cdot|\mathbf{X})] + c]$$

for any constant $c \in \mathbb{R}$. This region is invariant as required, indeed:

$$\begin{aligned}
&Y + b \in H_c(\mathbf{X} + b) \Leftrightarrow \\
&Y + b \ \leq \ \rho[\hat{F}(\cdot|\mathbf{X} + b)] + c \ = \ \rho[\hat{F}(\cdot|\mathbf{X})] + b + c \Leftrightarrow \\
&Y \ \leq \ \rho[\hat{F}(\cdot|\mathbf{X})] + c \Leftrightarrow \\
&Y \in H_c(\mathbf{X}).
\end{aligned}$$

It follows that

$$\hat{\mathbb{P}}(Y - \rho[\hat{F}(\cdot|\mathbf{X})] \leq c|\mathbf{x}) = \mathbb{P}(Y - \rho[\hat{F}(\cdot|\mathbf{X})] \leq c) \qquad \forall c \in \mathbb{R}.$$

As this holds for every $c \in \mathbb{R}$, it is implied that the random variable $W = Y - \rho[\hat{F}(\cdot|\mathbf{X})]$ has the same distribution under $\hat{\mathbb{P}}(\cdot|\mathbf{x})$ and $\mathbb{P}(\cdot)$. Thus if $G(w) = \mathbb{P}(W \leq w)$ and $\hat{G}(w|\mathbf{x}) = \hat{\mathbb{P}}(W \leq w|\mathbf{x})$ it is $G(w) = \hat{G}(w|\mathbf{x})$ for all w. By law invariance of ρ it then is:

$$\rho[G(\cdot)] = \rho[\hat{G}(\cdot|\mathbf{x})]$$

However, by the construction of the random variable W it is $\rho[\hat{G}(\cdot|\mathbf{x})] = 0$. Hence

$$\rho[G(\cdot)] = \rho(Y - \rho[\hat{F}(\cdot|\mathbf{X})]) = 0.$$

□

Suppose now that Y belongs to a scale family, with parameter θ. We use the prior $\pi(\theta) = 1/\theta$. It is known that (eg see Gerrard and Tsanakas (2011)), if $\mathbf{X} = b\mathbf{Z}$, where $b > 0$, $\mathbf{Z} = (Z_1, \ldots, Z_n)$ and $\mathbf{Z} \sim F$, then

$$\hat{F}(y|b\mathbf{z}) = \hat{F}(y/b|\mathbf{z}).$$

Therefore,

$$\rho[\hat{F}(\cdot|\mathbf{x})] = \rho[\hat{F}(\cdot/b|\mathbf{z})] = b\rho[\hat{F}(\cdot|\mathbf{z})],$$

due to the positive homogeneity property of $\rho(\cdot)$.

Proposition 3.6.3 shows that for scale-families the residual risk is not completely eliminated by using the capital estimator $\eta_{bay}(\mathbf{X})$. Instead, a scaled version of the residual risk goes to zero.

Proposition 3.6.3. *For scale families, using the capital estimator* $\eta_{bay}(\boldsymbol{X}) = \rho[\hat{F}(\cdot|\boldsymbol{X})]$ *yields*

$$\rho\left(\frac{Y}{\rho[\hat{F}(\cdot|\boldsymbol{X})]} - 1\right) = 0.$$

Proof. The same procedure as in the proof of Proposition 3.6.2 is followed. The predictive region is

$$H_c(\mathbf{X}) = (-\infty, c\rho[\hat{F}(\cdot|\mathbf{X})]]$$

for any constant $c \in \mathbb{R}$. This region is invariant as required in Prop 3.6.1, indeed

$$\begin{aligned}
&bY \in H_c(b\mathbf{X}) \Leftrightarrow \\
&bY \;\leq\; c\hat{\rho}[\hat{F}(\cdot|b\mathbf{X})] \;=\; cb\hat{\rho}[\hat{F}(\cdot|\mathbf{X})] \Leftrightarrow \\
&Y \;\leq\; c\rho[\hat{F}(\cdot|\mathbf{X})] \Leftrightarrow \\
&Y \in H_c(\mathbf{X}).
\end{aligned}$$

It follows that:

$$\hat{\mathbb{P}}\Big(\frac{Y}{\rho[\hat{F}(\cdot|\mathbf{X})]} \leq c|\mathbf{x}\Big) = \mathbb{P}\Big(\frac{Y}{\rho[\hat{F}(\cdot|\mathbf{X})]} \leq c\Big) \quad \forall c \in \mathbb{R}.$$

As this holds for every $c \in \mathbb{R}$, it is implied that the random variable $W = Y/\rho[\hat{F}(\cdot|\mathbf{X})]$ has the same distribution under $\hat{\mathbb{P}}(\cdot|\mathbf{x})$ and $\mathbb{P}(\cdot)$. Thus if $G(w) = \mathbb{P}(W \leq w)$ and $\hat{G}(w|\mathbf{x}) = \hat{\mathbb{P}}(W \leq w|\mathbf{x})$ it is $G(w) = \hat{G}(w|\mathbf{x})$ for all w. By law-invariance of ρ it then is:

$$\rho[G(\cdot)] = \rho[\hat{G}(\cdot|\mathbf{x})]$$

However, by the construction of the random variable W it is $\rho[\hat{G}(\cdot|\mathbf{x})] = 1$. Hence

$$\rho[G(\cdot)] = \rho\left(\frac{Y}{\rho[\hat{F}(\cdot|\mathbf{X})]}\right) = 1.$$

□

Results relating to Section 3.3.4

First, the stated equations (3.22) and (3.23) for $\eta_{bs1}(\mathbf{X})$ and $\eta_{bs2}(\mathbf{X})$ respectively are derived. Using identical notation to Section 3.3.4, the residual estimation risk $r_1(\theta)$ is:

$$\begin{aligned} r_1(\theta) &= \rho(\mu + \sigma Z - \hat{\mu} - \hat{\sigma}\rho(Z)) \\ &= \rho(\mu + \sigma Z - \mu - \sigma U - \sigma V \rho(Z)) \\ &= \sigma\rho(Z - U - V\rho(Z)), \end{aligned}$$

and thus

$$r_1(\hat{\theta}) = \hat{\sigma}\rho(Z - U - V\rho(Z)).$$

The capital estimator $\eta_{bs1}(\mathbf{X})$ is

$$\begin{aligned} \eta_{bs1}(\mathbf{X}) &= \rho[F(\cdot, \hat{\theta})] + r_1(\hat{\theta}) \\ &= \hat{\mu} + \hat{\sigma}\rho(Z) + \hat{\sigma}\rho(Z - U - V\rho(Z)) \\ &= \mu + \sigma U + \sigma V(\rho(Z) + \rho(Z - U - V\rho(Z))) \end{aligned}$$

It follows that

$$\begin{aligned} r_2(\theta) &= \rho(Y - \eta_{bs1}(\mathbf{X})) \\ &= \rho(\mu + \sigma Z - \mu - \sigma(U + V(\rho(Z) + \rho(Z - U - V\rho(Z))))) \\ &= \sigma\rho(Z - U - V(\rho(Z) - \rho(Z - U - V\rho(Z)))), \end{aligned}$$

and thus

$$r_2(\hat{\theta}) = \hat{\sigma}\rho(Z - U - V(\rho(Z) - \rho(Z - U - V\rho(Z)))).$$

Therefore

$$\begin{aligned}\eta_{bs2}(\mathbf{X}) &= \eta_{bs1}(\mathbf{X}) + r_2(\hat{\theta}) \\ &= \hat{\mu} + \hat{\sigma}\left[\rho(Z) + \rho(Z - U - V\rho(Z)) + \rho(Z - U - V(\rho(Z) - \rho(Z - U - V\rho(Z))))\right]\end{aligned}$$

Now it is shown that, for location families the first order bootstrap capital estimator bootstrap η_{bs1} eliminates exactly the residual estimation risk, that is,$\rho(Y - \eta_{bs1}(\mathbf{X})) = 0$. Let μ be the location parameter, such that we can write $Y{=}\mu + Z$ for $Z \sim F$ and $\rho(Y) \stackrel{d}{=} \mu + \rho(Z) \stackrel{d}{=} \mu + \rho[F]$. The MLE of $\hat{\mu}$ can be written as $\hat{\mu} \stackrel{d}{=} \mu + U$, where the distribution of U does not depend on μ. It is

$$\begin{aligned}r_1(\mu) &= \rho(\mu + Z - \hat{\mu} - \rho(Z)) \\ &= \rho(\mu + Z - \mu - U - \rho(Z)) \\ &= \rho(Z - U) - \rho(Z)\end{aligned}$$

As this does not depend on μ we have $r_1(\mu) = r_1(\hat{\mu})$. Thus

$$\begin{aligned}\eta_{bs1}(\mathbf{X}) &= \rho[F(\cdot, \hat{\mu})] + r_1(\hat{\mu}) \\ &\stackrel{d}{=} \mu + U + \rho(Z) + \rho(Z - U) - \rho(Z) \\ &= \mu + U + \rho(Z - U).\end{aligned}$$

The residual risk from using $\eta_{bs1}(\mathbf{X})$ then is

$$\begin{aligned}r_2(\mu) &= \rho(Y - \eta_{bs1}(\mathbf{X})) \\ &= \rho(\mu + Z - \mu - U - \rho(Z - U)) \\ &= 0.\end{aligned}$$

Table 3.1: Normalised residual estimation risk for a normally distributed risk with sample size n, risk measure TVaR_p, and the MLE capital estimator η.

	n=10	n=20	n=50	n=100
p=0.95	0.216	0.112	0.046	0.023
p=0.99	0.266	0.141	0.059	0.030
p=0.995	0.286	0.154	0.065	0.033

Table 3.2: Normalised residual estimation risk for an exponentially distributed risk with sample size n, risk measure TVaR_p, and the MLE capital estimator η.

	n=10	n=20	n=50	n=100
p=0.95	0.212	0.118	0.051	0.026
p=0.99	0.251	0.144	0.063	0.033
p=0.995	0.267	0.156	0.069	0.036

Table 3.3: Normalised residual estimation risk for a normally distributed risk with sample size n, risk measure TVaR_p, and the Bayes capital estimator η_{bay}.

	n=10	n=20	n=50	n=100
p=0.95	-0.017	-0.007	-0.003	-0.001
p=0.99	-0.013	-0.005	-0.002	-0.001
p=0.995	-0.011	-0.005	-0.002	-0.001

Table 3.4: Normalised residual estimation risk for an exponentially distributed risk with sample size n, risk measure TVaR_p, and the Bayes capital estimator η_{bay}.

	n=10	n=20	n=50	n=100
p=0.95	-0.016	-0.009	-0.003	-0.002
p=0.99	-0.012	-0.006	-0.002	-0.001
p=0.995	-0.010	-0.005	-0.002	-0.001

Table 3.5: Normalised residual estimation risk for a normally distributed risk with sample size n, risk measure TVaR_p, and the bootstrap capital estimators $\eta_{bs1},\ \eta_{bs2}$.

η_{bs1}	n=10	n=20	n=50	n=100
p=0.95	0.046	0.012	0.002	0.000
p=0.99	0.081	0.023	0.003	0.001
p=0.995	0.096	0.030	0.005	0.001
η_{bs2}	n=10	n=20	n=50	n=100
p=0.95	0.011	0.002	0.000	0.000
p=0.99	0.030	0.004	0.000	0.000
p=0.995	0.041	0.006	0.000	0.000

Table 3.6: Normalised residual estimation risk for an exponentially distributed risk with sample size n, risk measure TVaR_p, and the bootstrap capital estimators $\eta_{bs1},\ \eta_{bs2}$.

η_{bs1}	n=10	n=20	n=50	n=100
p=0.95	0.065	0.020	0.004	0.001
p=0.99	0.096	0.032	0.007	0.002
p=0.995	0.110	0.040	0.008	0.001
η_{bs2}	n=10	n=20	n=50	n=100
p=0.95	0.022	0.004	0.000	0.000
p=0.99	0.039	0.007	0.001	0.000
p=0.995	0.049	0.012	0.001	0.001

Table 3.7: Normalised residual estimation risk for a log-normally distributed risk with different values of the coefficient of variation $\mathrm{CV}(Y')$, sample size n, risk measure $\mathrm{TTVaR}_{p_1,0.997}$, and the MLE capital estimator η.

	n=10	n=20	n=50	n=100
$\mathrm{CV}(Y') = 0.1$				
$p_1 = 0.95$	0.227	0.119	0.049	0.025
$p_1 = 0.99$	0.270	0.147	0.062	0.031
$p_1 = 0.995$	0.284	0.156	0.066	0.034
$\mathrm{CV}(Y') = 0.2$				
$p_1 = 0.95$	0.244	0.131	0.055	0.028
$p_1 = 0.99$	0.289	0.161	0.070	0.036
$p_1 = 0.995$	0.304	0.171	0.075	0.039
$\mathrm{CV}(Y') = 0.5$				
$p_1 = 0.95$	0.288	0.163	0.071	0.037
$p_1 = 0.99$	0.336	0.200	0.091	0.048
$p_1 = 0.995$	0.351	0.212	0.098	0.052

Table 3.8: Normalised residual estimation risk for a Pareto distributed risk with different values of the parameter θ, sample size n, risk measure $\mathrm{TTVaR}_{p_1,0.997}$, and the MLE capital estimator η.

	n=10	n=20	n=50	n=100
$\theta = 0.1$				
$p_1 = 0.95$	0.226	0.130	0.057	0.030
$p_1 = 0.99$	0.260	0.156	0.071	0.038
$p_1 = 0.995$	0.273	0.165	0.077	0.040
$\theta = 0.25$				
$p_1 = 0.95$	0.257	0.155	0.072	0.038
$p_1 = 0.99$	0.289	0.183	0.089	0.048
$p_1 = 0.995$	0.302	0.194	0.096	0.052
$\theta = 0.5$				
$p_1 = 0.95$	0.309	0.207	0.107	0.060
$p_1 = 0.99$	0.327	0.227	0.123	0.070
$p_1 = 0.995$	0.337	0.237	0.130	0.075

Table 3.9: Normalised residual estimation risk for a log-normally distributed risk with different values of the coefficient of variation $\mathrm{CV}(Y')$, sample size n, risk measure $\mathrm{TTVaR}_{p_1,0.997}$, and the adjusted capital estimator η_{adj}.

	n=10	n=20	n=50	n=100
$\mathrm{CV}(Y') = 0.1$				
$p = 0.95$	0.068	0.028	0.010	0.005
$p_1 = 0.99$	0.024	0.008	0.002	0.001
$p_1 = 0.995$	0.005	0.002	0.001	0.000
$\mathrm{CV}(Y') = 0.2$				
$p_1 = 0.95$	0.074	0.031	0.011	0.005
$p_1 = 0.99$	0.025	0.009	0.003	0.001
$p_1 = 0.995$	0.006	0.002	0.001	0.000
$\mathrm{CV}(Y') = 0.5$				
$p_1 = 0.95$	0.089	0.038	0.014	0.007
$p_1 = 0.99$	0.029	0.011	0.003	0.002
$p_1 = 0.995$	0.008	0.002	0.001	0.000

Table 3.10: Normalised residual estimation risk for a Pareto distributed risk with different values of the parameter θ, sample size n, risk measure $\mathrm{TTVaR}_{p_1,0.997}$, and the adjusted capital estimator η_{adj}.

	n=10	n=20	n=50	n=100
$\theta = 0.1$				
$p_1 = 0.95$	0.057	0.027	0.011	0.005
$p_1 = 0.99$	0.015	0.007	0.002	0.001
$p_1 = 0.995$	0.003	0.002	0.000	0.000
$\theta = 0.25$				
$p_1 = 0.95$	0.068	0.036	0.015	0.007
$p_1 = 0.99$	0.018	0.009	0.003	0.002
$p_1 = 0.995$	0.003	0.002	0.000	0.000
$\theta = 0.5$				
$p_1 = 0.95$	0.098	0.064	0.032	0.018
$p_1 = 0.99$	0.024	0.016	0.009	0.005
$p_1 = 0.995$	0.004	0.003	0.001	0.000

Table 3.11: Normalised residual estimation risk for a log-normally distributed risk with different values of the coefficient of variation CV(Y'), sample size n, risk measure TTVaR$_{p_1,0.997}$, and the Bayes capital estimator η_{bay}.

CV(Y') = 0.1	n=10	n=20	n=50	n=100
$p_1 = 0.95$	-0.010	-0.004	-0.001	-0.001
$p_1 = 0.99$	-0.002	-0.001	0.000	0.000
$p_1 = 0.995$	0.000	0.000	0.000	0.000
CV(Y') = 0.2				
$p_1 = 0.95$	-0.013	-0.006	-0.003	-0.002
$p_1 = 0.99$	-0.003	-0.002	-0.001	-0.001
$p_1 = 0.995$	-0.001	-0.001	-0.001	-0.001
CV(Y') = 0.5				
$p_1 = 0.95$	-0.018	-0.008	-0.003	-0.001
$p_1 = 0.99$	-0.002	-0.001	0.000	0.000
$p_1 = 0.995$	0.000	0.000	0.000	0.000

Table 3.12: Normalised residual estimation risk for a Pareto distributed risk with different values of the parameter θ, sample size n, risk measure TTVaR$_{p_1,0.997}$, and the Bayes capital estimator η_{bay}.

	n=10	n=20	n=50	n=100
$\theta = 0.1$				
$p_1 = 0.95$	-0.010	-0.005	-0.002	-0.001
$p_1 = 0.99$	0.001	0.000	0.000	0.000
$p_1 = 0.995$	0.000	0.000	0.001	0.001
$\theta = 0.25$				
$p_1 = 0.95$	-0.006	-0.001	0.001	0.001
$p_1 = 0.99$	0.000	0.001	0.002	0.002
$p_1 = 0.995$	0.001	0.001	0.001	0.001
$\theta = 0.5$				
$p_1 = 0.95$	0.012	0.018	0.012	0.008
$p_1 = 0.99$	0.006	0.007	0.006	0.004
$p_1 = 0.995$	0.002	0.002	0.002	0.002

Chapter 4

Risk Measurement and Model Uncertainty

Abstract: In the present contribution we investigate the impact of model uncertainty on the calculation of risk measures, such as VaR, used to quantify solvency capital requirements. We propose to measure that impact as the extra capital that needs to be added to the position in order to eliminate the additional risk that model error incurs and we call this residual estimation risk. With such an approach we measure the effectiveness of four different methods. For a given set of candidate models the model posterior weights can be obtained via a Bayesian approach. Then we consider approaches based on: (a) worst case scenario, (b) highest model posterior, (c) averaging the capital under each model according to the model posterior weights and (d) determining the predictive distribution of the financial loss and using it to calculate the capital. It emerges that all these methods work rather well when a set of candidate models has been carefully specified, for instance via expert judgment. However, when the model set has been chosen with poor prior information the effectiveness of these approaches decreases substantially, highlighting high sensitivity to the model set specification. It is also shown that with poor prior information on the model set, averaging across models is more efficient than selecting a single model; in particular (a)

performs very poorly. Furthermore, it appears that mis-specifying the model by choosing distributions that are more heavy-tailed than the one generating the data, may reduce the capital causing a higher residual risk.

4.1 Introduction

Current financial and insurance practice, largely motivated by solvency capital requirements, places a substantial focus on the accurate quantification of the risk of financial losses via risk measures. For instance, capital requirements under the EU project Solvency II require the calculation of the Value-at-Risk with confidence level $p = 0.995$, see European Insurance and Occupational Pensions Authority (2009).

As the actual probability distribution of losses remains unknown, it needs to be estimated from samples of past available data, which are always limited in size and, sometimes, very small. Thus the limitations of available data create potential for substantial parameter and model error.

The following distinction of parameter and model error is often made in the literature and is used in this chapter. Parameter error arises from the deviation of estimated parameters from their true values, in the context of a correctly chosen probability distribution. Model error arises from incorrect specification of the loss probability distribution itself. For the potential of parameter (resp. model) error occurring, we use the term parameter (resp. model) uncertainty. For a review of the literature on risk measures and parameter uncertainty we refer respectively to Chapters 1 and 3.

In this contribution we focus on model uncertainty. In the statistical literature an early treatment of model uncertainty and model selection can be found in Jeffreys (1961), while Ellsberg (1961) is one of the first authors to investigates the impact of model uncertainty on decision making. Gilboa and Schmeidler (1989) present an axiomatic approach to model uncertainty where among a set of candidate models it is preferable to use the one that presents the worst case scenario. Also, coherent and convex risk measures,

see Artzner et al. (1999) and Föllmer and Schied (2002), can be represented as worst expected loss across a set of generalized scenarios.

Large attention has been given to model uncertainty in the context of Bayesian literature. Here a procedure called Bayesian Model Averaging (BMA) is generally proposed, where both parameters and models are considered as random variables with their own prior and posterior distributions. For a detailed treatment of Bayesian approaches we refer to Draper (1995), Hoeting et al. (1999) and Bernardo and Smith (2000). Such approaches has been investigated in the context of insurance by Klugman (1992) and Cairns (2000).

While the wide majority of risk measures used in practice and in the literature are law-invariant (or model-dependent), the impact of model uncertainty on risk measurement has been only partially investigated. Föllmer and Knispel (2011) propose a worst case approach so that each law-invariant coherent and convex risk measure is calculated across different models and then the supremum is taken. Branger and Schlag (2004) investigate different approaches to calculate risk measures under model uncertainty and their impact on hedging strategies.

The purpose of this chapter is to estimate the potential impact of model uncertainty on risk measures used to quantify solvency capital requirements and to investigate the effectiveness by which different approaches to estimating capital requirements address model uncertainty. In particular, for a specified set of candidate models $\mathcal{M} = \{M_1, \ldots, M_K\}$, we investigate four approaches:

(a) Calculating the capital according to the most conservative model;

(b) Calculating the capital according to the model with the highest posterior weight;

(c) Calculating the capital under each model and averaging the capital amounts according to the model posterior weights;

(d) Determining a weighted average of candidate model distribution: according to their posterior weights, and using that distribution to calculate the capital.

For each approach considered, its performance is quantified via the *residual estimation risk* that represents the extra capital needed to account for model uncertainty (already introduced in Chapter 3 for measuring parameter uncertainty). In order to compensate for parameter uncertainty, for every candidate model, we work with its predictive distribution. It was proved by Gerrard and Tsanakas (2011) that for a wide class of distributions and the capital calculated according to VaR, this approach eliminates the residual risk completely when there is no model uncertainty. In Chapter 3, it was proved that also for more general risk measures used to quantify solvency capital requirement this approach is very effective in reducing the effect of parameter uncertainty. Hence the residual estimation risk measured here is only due to model uncertainty.

We also distinguish two degrees of model uncertainty:

An *informative* model set $\mathcal{M}_1$ has been specified, for instance by expert judgment. The true model M may or may not belong to this set;

A *non-informative* model set $\mathcal{M}_2$ has been specified without expert judgment. Again the true model may or may not belong to this set.

To compare the different approaches we specify a Test Set $\mathcal{T}$ of models. Since we do not know in reality the true model, the models in the Test Set serve as benchmarks on which the performance of different capital estimation methods is assessed. A Test Set is not generally identical with the set of candidate models considered. Essentially, assuming that the set of benchmark models (the Test Set) for an application is identical to the set of models that a statistician specifies is a best case scenario. In reality the two sets will be different, for reasons such as insufficient expert knowledge of the statistician or computational convenience.

Finally, for each estimation method the average, the maximum and the maximum of the absolute residual risk across test models is calculated. These figures summarize the performance of each capital estimation method in relation to the Test Set.

From our study, it emerges that all the approaches considered are sensitive to the specification of the model set. In particular when this is informative, all the approaches are rather effective with the highest posterior approach being the best one. For a non-informative model set, the worst case approach performs very poorly requiring a capital that is extremely conservative, while approaches based on model averaging seem to be more effective. It is also shown that using models that are more heavy-tailed than the test model may produce the counterintuitive result of reducing the capital causing a higher residual risk.

This chapter is organized as follows. Section 4.2 introduces in more detail model uncertainty and briefly reviews risk measures. In Section 4.3 we describe the four approaches to model uncertainty considered, while Section 4.4 introduces the procedure to compare their effectiveness. Finally Sections 4.5 and 4.6 report the simulation study and the discussion of the results obtained.

4.2 Preliminaries

4.2.1 Model Uncertainty

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{X} \subseteq \mathcal{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ represent the set of all financial losses considered. The random variable $Y \in \mathcal{X}$ represents the loss of a portfolio over a given time horizon, thus the event $Y > 0$ corresponds to a loss, while $Y \leq 0$ is a gain. Under the true model M, we say that Y has distribution $F_M(\cdot; \theta_M)$ and we write $Y \sim F_M(\cdot; \theta_M)$, where F is continuous, invertible, with density function $f_M(\cdot; \theta_M)$. $\theta_M \in \Theta_M \subset \mathbb{R}^{d_M}$ is the parameter vector of model M.

For practical applications, apart from the randomness due to the stochastic nature of the random variable Y, we have to consider two other sources of uncertainty:

- Uncertainty about the parameter vector θ for model M (*parameter uncertainty*);
- Uncertainty about the model M that specifies the loss distribution (*model uncertainty*).

In these situations the probability distribution of Y is generally estimated from a sample of past available data $\mathbf{X}$. Here we assume that these data are generated from the same distribution of Y and, with slight abuse of notation, write $\mathbf{X} \sim F(\cdot;\theta)$.

4.2.2 Risk measures

A risk measure is a functional $\rho : \mathcal{X} \longrightarrow \mathbb{R}$ that assigns to every financial loss $Y \in \mathcal{X}$ a real number $\rho(Y)$. In the present contribution, $\rho(\cdot)$ represents a regulatory capital requirement. A negative outcome of ρ indicates that the financial loss Y is *acceptable*, vice versa $\rho(Y) > 0$ means that the loss is *not acceptable*.

We work with risk measures satisfying the following conditions. For every $X, Y \in \mathcal{X}$:

(1) *Law-invariance.* If $X \overset{d}{=} Y \;\Rightarrow\; \rho(X) = \rho(Y)$;

(2) *Translation invariance.* If $m \in \mathbb{R}, \;\Rightarrow\; \rho(X + m) = \rho(X) + m$;

(3) *Positive homogeneity.* If $\lambda \geq 0, \;\Rightarrow\; \rho(\lambda X) = \lambda\rho(X)$,

(4) *Monotonicity.* If $X \geq Y$ $\mathbb{P}$ a.s, $\;\Rightarrow\; \rho(X) \geq \rho(Y)$,

where $\overset{d}{=}$ denotes equality in distribution. Thanks to law-invariance, for a random variable $Y \sim F(\cdot;\theta)$, we can use both the notations $\rho(Y)$ or $\rho[F(\cdot;\theta)]$, for details we refer to Chapter 3. A risk measure $\rho(\cdot)$ that satisfies the

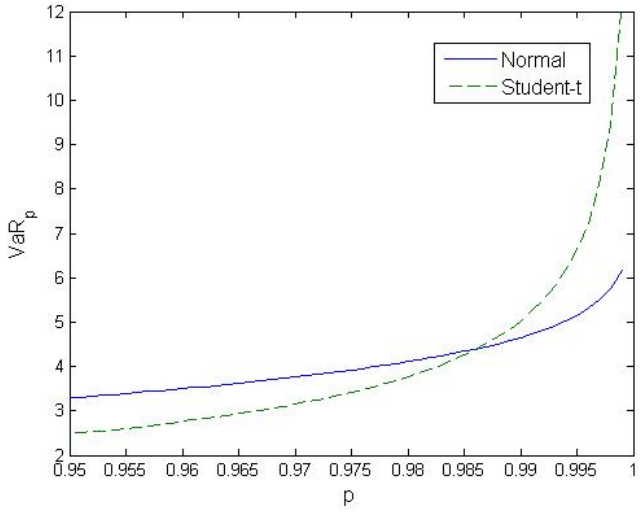

Figure 4.1: VaR_p for the Normal and t-Student distribution for $p \in [0.97, 1)$

above properties can be used to calculate solvency capital requirements. Its outcome, $\rho(Y)$, identifies the smallest capital that added to the position makes it acceptable, ie:

$$\rho(Y - \rho(Y)) = 0. \tag{4.1}$$

A standard example of risk measure that satisfies (1), (2), (3) and (4) is

$$\text{VaR}_p(Y) := \inf\{m \in \mathbb{R} : \mathbb{P}(Y \leq m) \geq p\}. \tag{4.2}$$

Law-invariant risk measures assess risk due to the stochastic nature of the financial loss Y assuming that a model for the loss probability distribution is known. Hence they do not take into account the additional risk arising from parameter and model uncertainty. As a very simple but illustrative example of model mis-specification, consider two possible models for a random variable Y: under model M_1, $Y \sim \mathcal{N}(0, 4)$; under M_2, Y follows a Student-t distribution $Y \sim t_{8/3}$. Both models have a mean of 0 and a standard deviation of 2. It is rather common in finance to choose between these two models (see Christoffersen (2011)). Fig. 4.1 gives the VaR_p for different values of p under the two models. The higher the confidence level p, the wider the gap between the two required capitals. For instance, under

M_1, $\text{VaR}_{0.995}(Y) = 5.15$ while under M_2, $\text{VaR}_{0.995}(Y) = 6.63$. Even in this basic example, model mis-specification produces a 22% difference in capital.

4.2.3 Residual estimation risk and parameter uncertainty

When the distribution of a loss $Y \in \mathcal{X}$ is unknown and estimated from a random sample $\mathbf{X}$, the capital $\rho(Y)$ is also estimated and denoted $\eta(\mathbf{X})$. Here $\eta(\mathbf{X})$ can represent any kind of estimation procedure used to calculate the capital requirement, that depends on the data. It follows that equation (4.1) does not hold anymore, instead we have:

$$\rho(Y - \eta(\mathbf{X})) \neq 0. \tag{4.3}$$

We denote the quantity on the left-hand side *residual estimation risk* and use it as a measure of model uncertainty. Note that $\eta(\mathbf{X})$ is random and not fixed. Hence, the residual risk represents the extra amount of capital required to make the position Y acceptable when assessing simultaneously the randomness arising from the stochastic nature of Y and the estimation procedure, indeed:

$$\rho(Y - \eta(\mathbf{X}) - \rho(Y - \eta(\mathbf{X}))) = 0. \tag{4.4}$$

4.3 Risk measurement approaches under parameter and model uncertainty

In this section we outline different methods proposed in the literature to deal with model uncertainty and see how they apply to the context of risk measurement. We denote by $\mathcal{M} = \{M_1, \ldots, M_K\}$ the set of candidate models considered. We assume that under each model M_k, the random variable Y has distribution $F_k(\cdot; \theta_{M_k})$, with the parameter vector $\theta_{M_k} \in \Theta_{M_k} \subseteq \mathbb{R}^{d_{M_k}}$.

The true model M may or may not belong to $\mathcal{M}$.

Following Cairns (2000), we identify three main classes of approaches to cope with parameter and model uncertainty:

[i] Fix a model $M_k \in \mathcal{M}$ and a parameter $\theta^*_{M_k} \in \Theta_{M_k}$ that, according to some criterion, fit better the data and proceed with the analysis as if they were the true ones;

[ii]According to some criterion, fix a model $M_k \in \mathcal{M}$ and use a Bayesian approach to deal with parameter uncertainty;

[iii] Use a Bayesian approach for both model and parameters.

In all the approaches considered here we deal with parameter uncertainty by using the predictive distribution of each model as explained in the next subsection. In Sections 4.3.2 and 4.3.3 we describe two approaches in class [ii], while in Sections 4.3.4 and 4.3.5 we discuss two approaches in class [iii].

4.3.1 Parameter uncertainty

In this subsection, we briefly review the Bayesian estimation method used in Chapter 3 to cope with parameter uncertainty. We assume that a model M for the loss $Y \in \mathcal{M}$ is fixed. Under M, the random variable Y has distribution $F(\cdot;\theta)$ where only the parameter $\theta \in \Theta$ is unknown and estimated from a sample of past data $\mathbf{X}$, where we assume $\mathbf{X} \sim F(\cdot;\theta)$. The key idea in the Bayesian technique, is that the unknown parameter θ is treated as a random variable. The prior $\pi(\theta)$ represents the parameter distribution when no information on the data is available. After a sample of data $\mathbf{X} = \mathbf{x}$ has been collected, a posterior distribution given the data is calculated according to the Bayes formula:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x};\theta)\cdot\pi(\theta)}{\int_{u\in\Theta} f(x;u)\cdot\pi(u)du} \tag{4.5}$$

where $f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i;\theta)$ is the likelihood of the data sample $\mathbf{x} = (x_1,\ldots,x_n)$. The predictive cumulative distribution function of Y given the

data is obtained as:

$$\hat{F}(y|\mathbf{x}) = \int_{\theta\in\Theta} F(y;\theta)\cdot\pi(\theta|\mathbf{x})d\theta. \tag{4.6}$$

For a fixed sample $\mathbf{X} = \mathbf{x}$, the estimated capital according to the Bayesian approach is:

$$\eta(\mathbf{x}) := \rho[\hat{F}(\cdot|\mathbf{x})], \tag{4.7}$$

that is the risk measure ρ applied to the predictive distribution of Y. When the sample is not fixed, then the estimated capital becomes a random variable itself:

$$\eta(\mathbf{X}) := \rho[\hat{F}(\cdot|\mathbf{X})]. \tag{4.8}$$

For location-scale families of distributions and their increasing transformations (for definitions and details we refer to Chapter 3) such approach eliminates completely the estimated residual risk, if we use the risk measure $\rho(\cdot) := \mathrm{VaR}_p(\cdot)$ and the right choice of prior (Gerrard and Tsanakas (2011)). This means that the Bayesian approach eliminates completely the risk due to parameter uncertainty independently of the unknown parameter θ. Also for other law-invariant risk measures, it was shown in Chapter 3, that this approach is very effective.

The rest of the chapter is devoted to investigate and compare different approaches to deal with model uncertainty when a Bayesian approach is used to cope with parameter uncertainty.

Assuming that the model M is unknown and instead a set $\mathcal{M}$ of candidate model has been specified, we denote $\rho[\hat{F}_{M_k}(\cdot|\mathbf{X})]$ the estimated capital according to model $M_k \in \mathcal{M}$. We highlight that in general $Y, \mathbf{X} \sim F(\cdot;\theta) \neq F_{M_k}(\cdot;\theta_{M_k})$. Hence, although the capital is estimated assuming that M_k is the correct model, it may be that Y and $\mathbf{X}$ are generated from a different distribution. Model uncertainty arises exactly from here: using a model that is not the one generating the data.

4.3.2 Worst-case approach (WC)

The first method that we consider is the Worst Case approach (WC). For each model $M_k \in \mathcal{M}$, we calculate the risk measure according to the predictive distribution $\hat{F}_{M_k}(\cdot|\mathbf{X})$, that is, $\rho[\hat{F}_{M_k}(\cdot|\mathbf{X})]$. Then we set the capital according to the most conservative model

$$\eta_{WC}(\mathbf{X}) := \max_{k \in K} \rho[\hat{F}_{M_k}(\cdot|\mathbf{X})]. \tag{4.9}$$

This method finds its root in Gilboa and Schmeidler (1989) on robust utility maximization, but its use is widespread among practitioners and academics. The idea behind it is straightforward: in order to be on the safe side we hold capital according to the worst case possible. This approach generally requires more capital than needed and is also computationally easy.

The set of models $\mathcal{M}$ considered plays a central role. Intuitively, the wider the set, the higher is the capital. It is clear that a model set that is too wide can easily lead to trivial results, such as an infinite capital. Moreover when true model M does not belong to the set $\mathcal{M}$ specified, the WC approach loses its interpretation because the true model may be more conservative than any model in $\mathcal{M}$. We will show in Section 4.5.3, how this approach may lead to unrealistic results depending on the set $\mathcal{M}$ specified.

4.3.3 Highest posterior approach (HP)

The second approach that we suggest is that of choosing the model that, according to the data, has the highest posterior weight. Details for this approach can be found in Draper (1995), Bernardo and Smith (2000) and Cairns (2000). This technique, based on a Bayesian perspective requires to specify, for each model in the set $\mathcal{M}$ a prior probability $p(M_k)$ and a prior distribution to its parameters vector $\pi(\theta_{M_k} \mid M_k)$. The posterior probability

that M_k is the correct model, given data $\mathbf{X} = \mathbf{x}$ is:

$$p(M_k|\mathbf{x}) = \frac{p(\mathbf{x}|M_k)p(M_k)}{\sum_{i=1}^{K} p(\mathbf{x}|M_i)p(M_i)}, \tag{4.10}$$

where

$$p(\mathbf{x}|M_k) = \int_{\Theta_{M_k}} f_{M_k}(\mathbf{x}\theta_{M_k})\pi(\theta_{M_k}|M_k)d\theta_{M_k} \tag{4.11}$$

is called *marginal distribution* or *prior predictive distribution* and represents the likelihood of $\mathbf{x}$ given M_k.

Formula (4.10) can be rewritten as:

$$p(M_k|\mathbf{x}) = \Big(\sum_{i=1}^{K} \frac{p(M_i)}{p(M_k)} \cdot B_{ik}\Big)^{-1}, \tag{4.12}$$

where B_{ik} is called *Bayes factor* of M_i on M_k and is defined by:

$$B_{ik} = \frac{p(\mathbf{x}|M_i)}{p(\mathbf{x}|M_k)}. \tag{4.13}$$

The Bayes factor is the ratio of the marginal distributions. Values of B_{ik} greater than 1 suggest that M_i has a higher chance than M_k to be the correct model given the data sample. If there is no prior information on the models, one can use $p(M_i) = 1/K$ for each model, and the Bayes factors encloses all the information required to calculate the model posterior distribution. Once we compute the Bayes factor, we choose the model that, given the data, is the most favorable one, that is the one that has the highest posterior.

The estimated capital is:

$$\eta_{HP}(\mathbf{X}) := \rho[\hat{F}_*(\cdot|\mathbf{X})] \tag{4.14}$$

where

$$M_* \in \mathcal{M} \;\text{ and }\; p(M_*|\mathbf{x}) \geq p(M_k|\mathbf{x}) \qquad \forall k \in \{1, \ldots, K\}.$$

The HP approach gives a rigorous and unified approach to deal with model

uncertainty. It selects the model that is the most favourable given the data and the prior knowledge. Moreover when the true model M belongs to $\mathcal{M}$, this approach recognizes it asymptotically (see for instance Bernardo and Smith (2000)) in the sense that the true model's posterior weight tends to 1. When M is not in $\mathcal{M}$, the effectiveness of this approach weakens and does not necessarily improve when the number of data increases. This is because the approach tends to focus on an incorrect model.

4.3.4 Bayesian Model Averaging 1 (BMA1)

Here we focus on a fully Bayesian approach. It takes as quantity of interest the estimated capitals under each model $\rho[\hat{F}_{M_k}(\cdot|\mathbf{X})]$ and averages them according to the posterior probability of each model M_k. The capital then is:

$$\eta_{BMA1}(\mathbf{X}) := \sum_{k \in K} \rho[\hat{F}_{M_k}(\cdot|\mathbf{X})] \cdot p(M_k|\mathbf{X}). \tag{4.15}$$

Similarly to the HP approach, this method assigns more weight to the model that is the most favourable according to the data but it has the advantage of keeping all the models in consideration. Averaging across models seems to be more reasonable than picking a single one, especially when the true model may not belong to the model set.

Moreover, if the predictive distribution for each model can be computed analytically, calculating $\rho[\hat{F}_{M_k}(\cdot|\mathbf{X})]$ is easy so that BMA1 can be simply calculated. Branger and Schlag (2004) describes a more conservative approach where the aversion to model risk is emphasized by a convex function ϕ. This approach is not pursued here.

4.3.5 Bayesian Model Averaging 2 (BMA2)

The last approach that we consider is again in a fully Bayesian perspective. Similarly to BMA1, the model is considered as a parameter itself lying in the set $\mathcal{M}$ and all the models in $\mathcal{M}$ are considered. The quantity of interest

here is the predictive distribution under each model. We first compute the predictive distribution of Y given data $\mathbf{X} = \mathbf{x}$. This is obtained computing the predictive distribution for each model M_k as in (4.6), and then averaging according to the model posteriors:

$$\hat{F}(y|\mathbf{x}) = \sum_{k=1}^{K} \hat{F}_{M_k}(y|\mathbf{x}) \cdot p(M_k|\mathbf{x}). \tag{4.16}$$

Once we have the predictive probability distribution for Y, the estimated capital is simply the risk measure applied to this distribution:

$$\eta_{BMA2}(\mathbf{X}) := \rho[\hat{F}(\cdot|\mathbf{X})]. \tag{4.17}$$

With such an approach, both parameter and model uncertainty are incorporated in the stochastic nature of Y and dealt with as if they were possible scenarios. As BMA1, when the true model belongs to model set, BMA2 chooses it asymptotically. When, on the other side, the model is not in the specified set, dealing with an average of different models seems to be more appropriate that picking a single one.

Methods 2, 3 and 4 all consider a Bayesian approach based on model posterior probability. In the following section we highlight some difficulties related to computing posterior model weights.

4.3.6 Computational issues

The use of model posteriors requires to specify a prior for each model. In general it is not clear how to select the priors and their choice may influence substantially the results. Moreover, posterior weights $p(M_k|\mathbf{X})$ are generally difficult to compute, indeed the marginal distributions $p(\mathbf{x}|M_k)$ are often not available in a closed form and require numerical calculation. Secondly, if $p(\mathbf{x}|M_k)$ is calculated using an improper prior (such as the probability matching priors used here), it will be defined only up to a constant c_{M_k}. This constant will also appear in the Bayes factor B_{ik} as c_{M_i}/c_{M_j}. Several

solutions have been proposed to overcome this issue. In what follows we will use the approach suggested in Berger and Pericchi (1996) named *Intrinsic Bayes factor*. This method consists in using part of the data to estimate the constant c_i and make the prior *proper*. The rest of the data are used to compute the Bayes factor according to this new proper prior. While selecting the correct set of training data is generally computationally demanding, it has been proved (Berger and Pericchi, 1996) that for location-scale families, or scale families, the constant c_{M_i}/c_{M_j} is always equal to 1 when using the Intrinsic Bayes Factor. In what follows, in order to focus on model uncertainty and not on computational issues, we will only work with scale distributions for which we are able to analytically compute the model posterior.

4.4 Assessing the effectiveness of risk measurement approaches under model uncertainty

We have outlined different approaches to deal with model uncertainty. The remainder of the paper is dedicated to compare such methods and understand if there is one that is always preferable to the others.

4.4.1 Model Set

We consider two different degrees of model uncertainty. In the first one we assume that a set of models $\mathcal{M}_1 = \{M_1, \ldots, M_K\}$ is well specified, for instance it has been chosen by expert judgment. We call this setting *Informative model set* and it corresponds to a best-case scenario. In our simulation study in Section 4.5, we consider an Informative model set, where each model identifies a loss distribution belonging to a scale family.

Alternatively we can assume that the model set is determined with no use of expert judgement and suggest instead using a non-informative model

set $\mathcal{M}_2$ consisting of $K = 10$ Gamma distributions with different values of the shape parameter α. Varying α we are able to reproduce a wide range of skewness.

4.4.2 Test Set

The residual estimation risk does depend on the true distribution of Y and hence it is in principle unknown. In order to compare different approaches we use a *Test Set* of distribution models $\mathcal{T} := \{T_1, \ldots, T_L\}$ that we use as benchmark. For each approach to model uncertainty, we measure its effectiveness on every distribution in the test set $\mathcal{T}$. We use the notation $Y_{T_l}, \mathbf{X}_{T_l}$ to emphasize that under the test model T_l the distribution is $F_{T_l}(\cdot; \theta_{T_l})$. In other words, for each approach, we compute:

$$\rho(Y_{T_l} - \eta(\mathbf{X}_{T_l})) \qquad \text{where} \quad Y_{T_l}, \mathbf{X}_{T_l} \sim F_{T_l}(\cdot; \theta) \quad \forall l \in \{1, \ldots, L\}. \tag{4.18}$$

Again Y_{T_l} and $\mathbf{X}_{T_l}$ are assumed to be independent. In such a way we can appreciate the effectiveness of each method across a set of different models. In order to make the results comparable, for any capital estimation method considered η, we compute the average

$$\frac{1}{L} \sum_{l=1}^{L} \rho(Y_{T_l} - \eta(\mathbf{X}_{T_l})) \tag{4.19}$$

the maximum

$$\max_{T_l} \rho(Y_{T_l} - \eta(\mathbf{X}_{T_l})) \tag{4.20}$$

and the maximum absolute value

$$\max_{T_l} |\rho(Y_{T_l} - \eta(\mathbf{X}_{T_l}))| \tag{4.21}$$

of the residual estimation risk across models in the Test Set. Depending on our preferences, we may consider more suitable the approach that gives the minimum average residual estimation risk, or the approach that minimizes

the maximum residual risk. It is also worth considering the maximum absolute value of the residual estimation risk because we may have approaches too conservative that always require an extremely high capital. Such approaches would always produce a negative residual estimation risk that is not necessarily desirable.

4.5 Simulation study

In this section the effectiveness of different approaches to model uncertainty is verified using Monte-Carlo simulations. The risk measure that we consider for residual estimation risk is VaR_p with level of confidence $p = 0.99$. We use a Monte-Carlo simulation with $m = 10^7$ simulations. As we did in Chapter 3, in order to make the results comparable, we normalize the residual risk in the following way:

$$\frac{\rho(Y - \eta(\mathbf{X}))}{\rho(Y) - E(Y)}. \tag{4.22}$$

In order to reduce the impact of the simulation error, we apply a simple importance sampling procedure. Instead of simulating $m = 10^7$ samples for $Y \sim F_{T_l}(\cdot, \theta_{T_l})$ we simulate a high proportion λm of observations from $Y > \text{VaR}_{0.9}[F_{T_l}(\cdot; \theta_{T_l})] = \beta$ and only $(1 - \lambda)m$ observations from $Y \leq \text{VaR}_{0.9}[F_{T_l}(\cdot; \theta_{T_l})]$, where $\lambda = 0.9$. In this way we have more accuracy in the estimation of the tail of the probability distribution of Y. Then, in order to compute:

$$\text{VaR}_{0.99}(Y - \eta(\mathbf{X}_{T_l})) \tag{4.23}$$

We need to find the value of z such that

$$P(Y - \eta(\mathbf{X}_{T_l}) \leq z) = 0.99. \tag{4.24}$$

Table 4.1: Distributions used for the Test set $\mathcal{T}$

Test models	$E[Y]$	$\sigma(Y)$	$\text{VaR}_{0.99}$(Y)	$\text{VaR}_{0.995}$(Y)
GM	100	20	152.30	158.98
LN	100	19.02	152.30	158.98
WB	100	24.63	152.30	157.00
IG	100	18.19	152.30	160.32

With few steps we have:

$$\begin{aligned} &\mathbb{P}(Y - \eta(\mathbf{X}_{T_l}) \leq z) - 0.99 \\ &= \mathbb{P}(Y - \eta(\mathbf{X}_{T_l}) \leq z | Y > \beta)\mathbb{P}(Y > \beta) \\ &\quad + \mathbb{P}(Y - \eta(\mathbf{X}_{T_l}) \leq z | Y \leq \beta)\mathbb{P}(Y \leq \beta) - 0.99 \\ &= \mathbb{P}(Y - \eta(\mathbf{X}_{T_l}) \leq z | Y > \beta)(0.1) \\ &+ \ \mathbb{P}(Y - \eta(\mathbf{X}_{T_l}) \leq z | Y \leq \beta)0.9 - 0.99. \end{aligned} \tag{4.25}$$

The value of z can be computed by searching for the root of (4.25). This simple approach allows us to reduce the sampling error.

4.5.1 Test Set

The Test Set considered consists of a Gamma, Log-normal, Weibull and Inverse-Gamma distribution and it is denoted $\mathcal{T} := \{GM, LN, WB, IG\}$. The shape and scale parameters of each distribution are chosen so that they all have a mean of 100 and the same $\text{VaR}_{0.99}$ of 152.30. Table 4.1, summarizes the characteristics of the distributions used. Figures 4.2 and 4.3 report respectively the probability density functions and quantile functions of the distributions specified. From now on we denote $Y_{GM}, \mathbf{X}_{GM}$ the loss and random sample generated from the Gamma distribution and similarly for the other models in the Test Set.

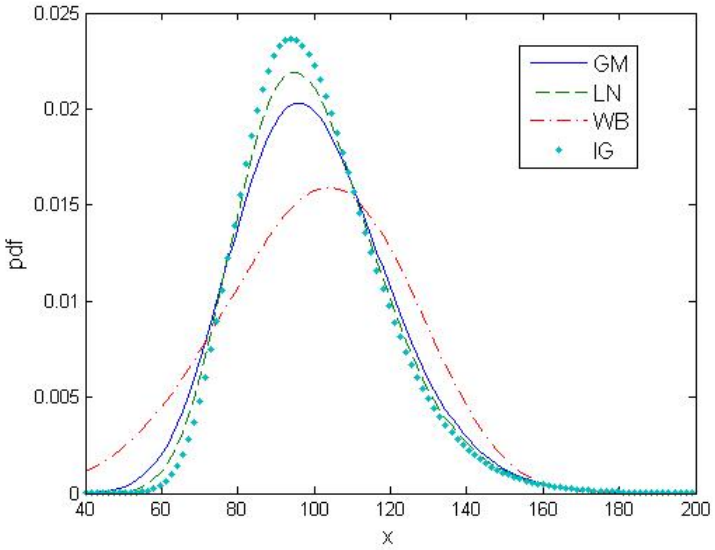

Figure 4.2: Probability density function for the models in $\mathcal{T}$

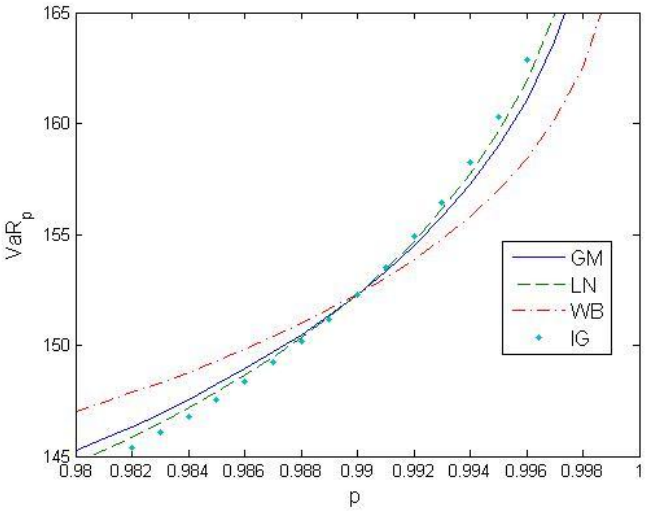

Figure 4.3: VaR_p for the models in $\mathcal{T}$

4.5.2 Model Set 1

In this section we investigate the effectiveness of approaches discussed in Section 4.3, where an informative model set $\mathcal{M}_1$ is used. In our simulation study, the fact that $\mathcal{M}_1$ has been carefully specified by expert judgment is represented by using a model set that almost overlaps with the Test Set. Hence, it consists of a Gamma, Log-normal and Inverse Gamma distributions. We denote it $\mathcal{M}_1 := \{GM, LN, IG\}$. The shape parameter are fixed and corresponds to the ones used for the Test Set $\mathcal{T}$, while the scale parameters are unknown. The purpose of this model set choice is two-folded: to verify the effectiveness of the approaches considered when the model set is carefully specified and when the true model may or may not belong to the set considered (Weibull). The capital calculated according to the Gamma model is denoted $\text{VaR}_p[\hat{F}_{GM}(\cdot; \mathbf{X}_{T_l})]$ and similarly for the other models in $\mathcal{M}_1$. In particular, if the data are generated from the Weibull distribution but the capital is calculated according to the Gamma model, we denote it $\text{VaR}_p[\hat{F}_{GM}(\cdot; \mathbf{X}_{WB})]$.

Worst-case approach

We start considering the WC approach. Using Monte-Carlo simulation, we compute:

$$\text{VaR}_p(Y_{T_l} - \eta_{WC}(\mathbf{X}_{T_l})) \qquad \forall l \in \{1, \dots, L\} \tag{4.26}$$

where

$$\eta_{WC}(\mathbf{X}_{T_l}) = \max_{k \in \{1,\dots,K\}} \text{VaR}_p[\hat{F}_{M_k}(\cdot|\mathbf{X}_{T_l})]. \tag{4.27}$$

The exact formula for $\text{VaR}_p[\hat{F}_{M_k}(\cdot|\mathbf{X}_{T_l})]$ is given in the Appendix 4.7.1.

Each row in Table 4.2 reports the residual risk when Y and $\mathbf{X}$ are respectively generated by the Gamma, Lognormal, Weibull and Inverse-Gamma distributions with a sample of size n. As we would expect, when the test model belongs to the model set, eg Gamma, Lognormal and Inverse-Gamma, the residual estimation risk is negative. This is exactly the theoretical aim

Table 4.2: Normalized residual risk using WC for $\mathcal{M}_1$ and a sample of size n

Test models	n=10	n=30	n=50	n=100	n=150
GM	-0.003	-0.001	-0.001	-0.000	0.000
LN	-0.016	-0.008	-0.005	-0.003	-0.003
WB	0.070	0.024	0.014	0.007	0.006
IG	-0.027	-0.014	-0.010	-0.008	-0.008
AVERAGE	0.006	0.0000	0.001	-0.001	-0.001
MAX	0.070	0.024	0.014	0.007	0.006
MAXABS	0.070	0.024	0.014	0.008	0.008

of the WC approach: to require enough capital to cover any scenario. This method though, fails its purpose when the test model is not in the model set, eg for the Weibull distribution.

The Weibull distribution is less heavy-tailed than any other distribution in the Test Set (as we can see from Figure 4.3), and hence, intuitively we would expect that it requires a lower capital than the others. However, it is the only one for which calculating the capital according to the WC approach produces a positive residual estimation risk. This is because mis-specifying the model by choosing distributions that are more heavy-tailed than the one generating the data, may reduce the capital. To understand better this concept, in Table 4.3 we report the average capital calculated according to each model (GM,LN,IG) when the data are generated from the distributions in the test set. For instance, the values in the third position in the last column (145.61), represents the average capital calculated according to the formula for the Inverse-Gamma model when the data are generated by a Weibull. We can see that this value is lower than the capital required if there was no model uncertainty (152.30). Thus in this case, calculating the capital using a distribution that is a more heavy tailed than the one generating the data underestimates the capital. As all the distributions considered in $\mathcal{M}_1$ are more heavy-tailed than the Weibull, they all underestimate the capital and hence the residual estimation risk is positive. Using the same reasoning, we can see that when the data come from the Gamma model, the highest capital

is given by the Gamma itself because the Lognormal and Inverse Gamma are more heavy tailed. Thus the model generating the data and the one that corresponds to the worst case are the same and hence the residual risk tends to 0 very quickly.

Note that, in general this approach does not improve its performance when the number of data increases. Here all the residual risks tend to 0 solely because all the distributions in $\mathcal{T}$ are chosen to have the same value for $\text{VaR}_{0.99}(Y)$. Hence, asymptotically all the estimated capital $\text{VaR}_{0.99}(Y_{T_l}|\mathbf{X}_{T_l})$ tend to the same value 152.30.

Table 4.3: Average of capital $E[\text{VaR}_p[\hat{F}_{M_k}(\cdot|\mathbf{X}_{T_l})]]$ for $n = 150$ with $p = 0.99$

Test models	$E[\text{VaR}_p[\hat{F}_{GM}(\cdot\|X_{T_l})]]$	$E[\text{VaR}_p[\hat{F}_{LN}(\cdot\|X_{T_l})]]$	$E[\text{VaR}_p[\hat{F}_{IG}(\cdot\|X_{T_l})]]$
GM	152.58	152.18	151.10
LN	152.61	152.56	151.92
WB	152.58	149.95	145.61
IG	152.60	152.83	152.51

Highest posterior approach

Consider now the second approach proposed. Here, for each model $M_k \in \mathcal{M}$ we compute the model posterior given the data generated by the model $T_l \in \mathcal{T}$, $p(M_k|\mathbf{X}_{T_l})$. We then select the capital calculated according to the model with the highest posterior and calculate the residual estimation risk on that. From Table 4.4 we can see that this approach works rather well. All the residual risks tend to 0. When the true distribution does not belong to the model set (Weibull), the residual risk is much worse. For instance, for $n = 10$ it is almost 10 times higher, than the other residual risks. However, as in the WC approach, it improves with the number of data points. For the Gamma, Lognormal and the Inverse Gamma the residual estimation risk tends to 0 because the posterior weight corresponding to the model generating the data tends to 1 when the sample size increases. By looking at the average, maximum and maximum absolute value of the residual estimation risks it

Table 4.4: Normalized residual risk using HP for $\mathcal{M}_1$

Test models	n=10	n=30	n=50	n=100	n=150
GM	0.007	0.002	0.001	0.000	0.000
LN	-0.006	-0.002	-0.001	0.000	0.000
WB	0.073	0.024	0.014	0.007	0.005
IG	-0.016	-0.008	-0.006	-0.004	-0.003
AVERAGE	0.014	0.004	0.002	0.001	0.000
MAX	0.073	0.024	0.014	0.007	0.005
MAXABS	0.073	0.023	0.014	0.007	0.005

emerges then that with a high number of data the HP approach appears to be slightly better than WC.

Bayesian model averaging 1

We now move to a fully Bayesian approach. Here we do not select any single model but keep all of them in consideration with their respective posterior weights. Again each capital is calculated according to the formulas in Appendix 4.7.1 and the posterior weights as in (4.12). For each model in the test set, the estimated capital is then:

$$\eta_{BMA1}(\mathbf{X}_{T_l}) = \sum_{k=1}^{K} \mathrm{VaR}_p[\hat{F}_{M_k}(\cdot|\mathbf{X}_{T_l})] \cdot p(M_k|\mathbf{X}_{T_l}). \tag{4.28}$$

The results are presented in Table 4.5.

Table 4.5: Normalized residual risk using BMA1 for $\mathcal{M}_1$

Test models	n=10	n=30	n=50	n=100	n=150
GM	0.016	0.006	0.004	0.002	0.001
LN	0.000	0.000	0.000	0.001	0.001
WB	0.089	0.027	0.015	0.007	0.005
IG	-0.014	-0.007	-0.005	-0.004	-0.003
AVERAGE	0.023	0.007	0.004	0.002	0.001
MAX	0.089	0.027	0.015	0.007	0.005
MAXABS	0.089	0.027	0.015	0.007	0.005

The residual estimation risk quickly goes to 0. It is interesting to note that the Inverse Gamma has a negative residual risk. This can be explained using a converse argument to the one used to explain the positive residual risk of the Weibull distribution. By looking at the last row in Table 4.3 we can see that the average capital calculated according to the Gamma and Lognormal is higher than the one calculated according to the Inverse Gamma (and that would give a null residual risk if there was no model uncertainty).

All the three approaches analyzed lead approximately to the same capital for data generated by the Weibull model. This is because, for the particular model set chosen, the Gamma is the model that gives the worst case scenario, but also the one with the highest posterior when the data come from the Weibull distribution. Hence all the three approaches give very similar results.

Bayesian Model Averaging 2

The last approach that we investigate is a fully Bayesian approach. The estimated capital is calculated according to:

$$\eta_{BMA2}(\mathbf{X}_l) = \mathrm{VaR}_p[\hat{F}(\cdot|\mathbf{X}_l)].$$

where $\hat{F}(\cdot|\mathbf{X}_l)$ is obtained as in (4.16). From Table 4.6 we can see that this approach presents an estimated residual risk very similar to the BMA1 approach. From a computational point of view, though it is slower. Indeed, in general, $\hat{F}(\cdot|\mathbf{X}_l)$ is not available in a closed form and the estimated capital needs to be computed numerically.

4.5.3 Model Set 2

We now pass to the second type of model uncertainty. Here the model set is unknown and instead a set of 10 Gamma distributions with different values of the shape parameter α is considered. We denote $\mathcal{M}_2 = \{GM_1, \ldots, GM_{10}\}$ the model set with parameters α respectively given by $\alpha = \{20, 21, \ldots, 29\}$. The Test Set remains unchanged. The Gamma distribution in the Test Set,

Table 4.6: Normalized residual risk using BMA2 for $\mathcal{M}_1$

Test models	n=10	n=30	n=50	n=100	n=150
GM	0.014	0.007	0.005	0.003	0.002
LN	-0.001	0.000	0.001	0.001	0.002
WB	0.090	0.027	0.015	0.007	0.006
IG	-0.015	-0.007	-0.006	-0.004	-0.003
AVERAGE	0.022	0.007	0.003	0.002	0.002
MAX	0.090	0.027	0.015	0.007	0.006
MAXABS	0.090	0.027	0.015	0.007	0.006

with our parameter choice, has shape parameter $\alpha = 25$, hence it is the only one that belongs also to the model set. The purpose of this model set choice is to test the sensitivity of these approaches when we do not have enough information to properly specify the model set.

Worst Case approach

The worst case residual risk is presented in Table 4.7. As we anticipated,

Table 4.7: Normalized residual risk using WC for $\mathcal{M}_2$

Test models	n=10	n=30	n=50	n=100	n=150
GM	-0.140	-0.135	-0.133	-0.133	-0.133
LN	-0.155	-0.139	-0.137	-0.134	-0.134
WB	-0.063	-0.109	-0.118	-0.125	-0.127
IG	-0.164	-0.143	-0.139	-0.136	-0.134
AVERAGE	-0.130	-0.131	-0.132	-0.132	-0.132
MAX	-0.063	-0.109	-0.118	-0.125	-0.127
MAXABS	0.164	0.143	0.139	0.136	0.134

using a wider model set improves our chance to cover any scenario. Here, for instance the capital for the Weibull distribution produces a negative residual risk, while under the model set $\mathcal{M}_1$ it was positive. On the other side, such an approach leads to unrealistic results when the model set is too wide. The capital required is extremely high and the residual risk that does not improve with the increasing number of data.

Highest Posterior approach

From Table 4.8, the HP approach gives results that looks much better compared to WC. When the test model does belong to the set $\mathcal{M}_2$ (Gamma), the residual estimation risk tends to 0. However, also HP is sensitive to the model set specified. For instance, under $\mathcal{M}_2$ the residual estimation risk for the Gamma test model is at least 7 times higher than under $\mathcal{M}_1$, which emphasizes a much lower performance for a non informative model set. For the other distributions, the residual risk tends to increase with the number of data. This may be interpreted by the fact that the more data we collect the more the posterior weights focus on a small portion of models that do not necessarily lead to the correct capital. In other words, with few data

Table 4.8: Normalized residual risk using HP for $\mathcal{M}_2$

Test models	n=10	n=30	n=50	n=100	n=150
GM	0.015	0.010	0.008	0.007	0.007
LN	0.026	0.038	0.043	0.050	0.053
WB	-0.018	-0.091	-0.110	-0.123	-0.127
IG	0.033	0.053	0.060	0.070	0.071
AVERAGE	0.014	0.048	0.000	0.001	0.001
MAX	0.033	0.091	0.060	0.0679	0.071
MAXABS	0.033	0.091	0.110	0.123	0.127

points there are several models that are possible candidates to be the most favourable ones. Hence the estimated capital changes frequently across simulations producing a residual risk that is moderately low. With more data points, the model posterior weights focuses on one or two models that may overestimate (Weibull) or underestimate (Inverse Gamma) the capital producing respectively a negative or a positive residual risk. In such cases, the more data points we collect, the worse the performance of HP gets.

Bayesian Model Averaging 1

The last approach that we present is the BMA1. Here again, from Table 4.9 we can see that if the test model belongs to the model set, the BMA1

approach recognizes it asymptotically and the residual risk tends to 0 slightly faster than with the HP approach. For the Lognormal, the Inverse-Gamma and Weibull its absolute value increases. As it was for the HP, the posterior weights tend to focus on a single model that does not correspond to the exact capital. However, since also the other models are kept in the average, this effect is moderated.

While looking at the average residual risk across distributions, it seems that the BMA1 is worse than the HP, but by looking at the maximum residual estimation risk, BMA1 appears preferable.

It seems then that when the model set is not accurately specified, averaging across several models is better than focussing on one or very few of them.

We do not report here results for BMA2 as they are very similar to what we obtain here and do not add much to the discussion.

Table 4.9: Normalized residual risk using BMA1 for $\mathcal{M}_2$

Test models	n=10	n=30	n=50	n=100	n=150
GM	-0.017	-0.013	-0.011	-0.007	-0.004
LN	-0.024	-0.005	0.004	0.019	0.028
WB	0.02	-0.053	-0.078	-0.106	-0.116
IG	-0.029	0.001	0.015	0.035	0.046
AVERAGE	-0.012	-0.017	-0.017	-0.015	-0.012
MAX	0.02	0.001	0.015	0.035	0.046
MAXABS	0.029	0.053	0.078	0.106	0.116

4.6 Discussion and conclusions

It appears from our study that model uncertainty is a very difficult and delicate matter. All the results presented show that there is not a straightforward and systematic approach that permits to deal with model uncertainty without requiring an in-depth analysis of the results obtained. When the model set almost overlaps with the Test Set, all the approaches considered are quite

effective. In particular, we have shown that although the WC approach does not ensure that the capital required is enough to cover also the worst case scenario, it gives the same residual risk than the HP and BMA approaches.

When the model set is non-informative and contains a wide range of models, the WC approach performs badly. This is because it requires a capital by far higher than what is actually needed by the test models. The HP and BMA approaches work better. However the posterior weights agglomerate around few models that are unable to furnish the exact capital required to eliminate the residual risk. Hence an increasing number of data corresponds to a worse performance of these approaches. This highlights that if the test model does not belong to the model set and we do not have expert judgment to properly specify a model set, then averaging across several models is better than focussing on few of them.

When the model set is informative, we find the HP approach slightly superior to the others. Otherwise, for a non informative model set, the BMA1 performs slightly better.

We do not notice any significant difference between approaches BMA1 and BMA2 apart from the fact that BMA2 is computationally more expensive.

Hence, from our analysis it emerges that all the approaches examined are very sensitive to the choice of the model set. When this substantially overlaps with the Test Set, the residual risk is significantly reduced, even if the test model do not belong to the model set.

4.7 Appendix

4.7.1 Model Set

In this section we report the calculation for the predictive density and cumulative distribution for each model M_k in $\mathcal{M}$ and the respective capital $\text{VaR}_p[\hat{F}_{M_k}(\cdot|\mathbf{X})]$. The model considered have shape parameter fixed and unknown scale parameter θ. We only use the uninformative prior $\pi(\theta) = \frac{1}{\theta}$. The parameter posterior is given by the formula:

$$\pi(\theta|\mathbf{x}) = \frac{\pi(\theta)\prod_{i=1}^{n} f(x_i;\theta)}{m(\mathbf{x})}$$

where $m(\mathbf{x}) = \int_0^\infty \pi(\theta)\prod_{i=1}^{n} f(x_i;\theta)d\theta$. The predictive density and cumulative function are calculated as:

$$\hat{f}(y|\mathbf{x}) = \int_0^\infty f(y;\theta)\cdot\pi(\theta|\mathbf{x})d\theta$$

and

$$\hat{F}(y|\mathbf{x}) = \int_0^y \hat{f}(t|\mathbf{x})dt$$

Gamma

Let $Y, X_1, \ldots, X_n \sim \Gamma(\alpha,\theta)$ with fixed shape parameter α and unknown scale parameter θ. The probability density function is

$$f(x;\theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha}x^{\alpha-1}e^{-\frac{x}{\theta}} \qquad\qquad \pi(\theta) = \frac{1}{\theta}.$$

For the parameter posterior we have

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \frac{\prod_{i=1}^{n} x_i^{\alpha-1}}{\Gamma(\alpha)^n\theta^{n\alpha+1}}e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}} \\ &= \frac{\prod_{i=1}^{n} x_i^{\alpha-1}}{\Gamma(\alpha)^n}\frac{\Gamma(n\alpha)}{(\sum_{i=1}^{n} x_i)^{n\alpha}} \cdot \Big[\frac{(\sum_{i=1}^{n} x_i)^{n\alpha}}{\Gamma(n\alpha)}\theta^{-n\alpha-1}e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}\Big].\end{aligned}$$

The quantity in the squared bracket is the density function of an Inverse-Gamma distribution with parameters $n\alpha$ and $(\sum_{i=1}^{n} x_i)$. It follows that for it to integrate to 1, the marginal distribution $m(\mathbf{x})$ is the normalizing constant:

$$m(\mathbf{x}) = \frac{\prod_{i=1}^{n} x_i^{\alpha-1}}{\Gamma(\alpha)^n} \frac{\Gamma(n\alpha)}{(\sum_{i=1}^{n} x_i)^{n\alpha}}.$$

The predictive density function is then

$$\begin{aligned}
\hat{f}(y|\mathbf{x}) &= \int_0^\infty f(y;\theta) \cdot \pi(\theta|\mathbf{x}) d\theta \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-\frac{y}{\theta}} \frac{(\sum_{i=1}^{n} x_i)^{n\alpha}}{\Gamma(n\alpha)} \theta^{-n\alpha-1} e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}} d\theta \\
&= \frac{y^{\alpha-1}(\sum_{i=1}^{n} x_i)^{n\alpha}}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^\infty \theta^{-\alpha(n+1)-1} e^{-\frac{y+\sum_{i=1}^{n} x_i}{\theta}} d\theta \\
&= \frac{y^{\alpha-1}(\sum_{i=1}^{n} x_i)^{n\alpha}\Gamma(\alpha(n+1))(y+\sum_{i=1}^{n} x_i)^{\alpha(n+1)}}{\Gamma(\alpha)\Gamma(n\alpha)(y+\sum_{i=1}^{n} x_i)^{\alpha(n+1)}\Gamma(\alpha(n+1))} \int_0^\infty \theta^{-\alpha(n+1)-1} e^{-\frac{y+\sum_{i=1}^{n} x_i}{\theta}} d\theta \\
&= \frac{\Gamma(\alpha(n+1))(\sum_{i=1}^{n} x_i)^{n\alpha}}{\Gamma(\alpha)\Gamma(n\alpha)} \frac{y^{\alpha-1}}{(y+\sum_{i=1}^{n} x_i)^{\alpha(n+1)}}.
\end{aligned}$$

and the predictive distribution is:

$$\hat{F}(y|\mathbf{x}) = \int_0^y \hat{f}(t|\mathbf{x}) dt = \frac{\Gamma(\alpha(n+1))(\sum_{i=1}^{n} x_i)^{n\alpha}}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^y \frac{t^{\alpha-1}}{(t+\sum_{i=1}^{n} x_i)^{\alpha(n+1)}} dt.$$

Considering the change of variable

$$z = \frac{t}{(t+\sum_{i=1}^{n} x_i)}$$

with

$$dz = \frac{\sum_{i=1}^{n} x_i}{(t+\sum_{i=1}^{n} x_i)^2} dt,$$

the above integral becomes

$$\begin{aligned}
\hat{F}(y|\mathbf{x}) &= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} \frac{t^{\alpha-1}(\sum_{i=1}^n x_i)^{n\alpha}}{(t+\sum_{i=1}^n x_i)^{\alpha(n+1)}} \frac{(t+\sum_{i=1}^n x_i)^2}{\sum_{i=1}^n x_i} dz \\
&= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} \frac{t^{\alpha-1}(\sum_{i=1}^n x_i)^{n\alpha-1}}{(t+\sum_{i=1}^n x_i)^{\alpha(n+1)-2}} dz \\
&= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} \frac{t^{\alpha-1}}{(t+\sum_{i=1}^n x_i)^{\alpha-1}} \cdot \frac{(\sum_{i=1}^n x_i)^{n\alpha-1}}{(t+\sum_{i=1}^n x_i)^{n\alpha-1}} dz \\
&= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} z^{\alpha-1} \cdot z^{n\alpha-1} dz = \mathrm{I}(\frac{y}{(y+\sum_{i=1}^n x_i)}; \alpha, n\alpha).
\end{aligned}$$

This corresponds to the Beta cumulative distribution function. To compute $\mathrm{VaR}_p[\hat{F}(y|\mathbf{x}, G)]$, we need to invert that function. Hence we have

$$\frac{y}{(y+\sum_{i=1}^n x_i)} = \mathrm{I}^{-1}(p, \alpha, n\alpha)$$

and

$$\mathrm{VaR}_p[\hat{F}_{GM}(|\mathbf{x})] = \big(\sum_{i=1}^n x_i\big) \frac{\mathrm{I}^{-1}(p, \alpha, n\alpha)}{1-\mathrm{I}^{-1}(p, \alpha, n\alpha)}$$

Lognormal

Consider a lognormal distribution with fixed shape parameter σ and unknown scale parameter $\gamma = e^{\mu}$. Its probability density function is

$$f(x, \gamma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp \frac{(\log(x) - \log(\gamma))^2}{2\sigma^2}.$$

The predictive cumulative function can be easily computed from the predictive cumulative function of a normal distribution with known parameter σ. Indeed if $Y \sim \mathcal{N}(\mu, \sigma^2)$ then $Y' = e^Z \sim \mathcal{LN}(\mu, \sigma^2)$ and $\hat{F}_{LN}(y|\mathbf{X}') = \hat{F}_N(\log(y)|\mathbf{X})$ where $\mathbf{X} = \log(\mathbf{X}')$ (for details see Chapter 3). The predictive distribution for the normal distribution is given $\mathcal{N}(\bar{\mathbf{X}}, \sigma^2(1+\frac{1}{n}))$. It follows that the lognormal predictive cumulative function is again Lognormal with scale parameter $e^{\frac{\sum_{i=1}^n \ln(x_i)}{n}}$ and shape parameter $\sqrt{1+\frac{1}{n}}\sigma$. The capital $\mathrm{VaR}_p[\hat{F}_{LN}(\cdot|\mathbf{X})]$ is obtained inverting that function.

Inverse-Gamma

The predictive cumulative function for the Inverse Gamma is obtained analogously to the Gamma distribution. The capital is obtained as

$$\mathrm{VaR}_p[\hat{F}_{IG}(\cdot|\mathbf{X})] = \frac{\mathrm{I}^{-1}(p, \alpha, n\alpha)}{\sum_{i=1}^{n}(1/X_i)} \tag{4.29}$$

Chapter 5

Directions for future research

The thesis investigates different challenges emerging from the use of risk measures to quantify solvency capital requirements.

Chapter 2 states sufficient conditions for a risk measure to satisfy mild notions of time-consistency with particular focus on sequential consistency. It emerges that most of the risk measures used in practice fail to satisfy these conditions. Hence, a systematic procedure to construct sequentially consistent risk measures is provided. We also propose a new approach to dynamic risk measurement, which is closer to current insurance practice. In real life, risk measures assess the risk of a financial position at a certain time horizon δ. While academic literature generally assumes that this is also the expiry date of the financial position, this is very atypical. In current practice, the risk measure is applied to the fair value of the position at the time horizon, rather than to the position itself. This situation is outside the usual framework of the risk measures literature, as it essentially corresponds to risk measurement with an argument that is shifting over time, as the fair value itself changes with newly available information. We called this setting risk measurement with rolling time horizon. This procedure not only contributes to reduce the gap between academic literature and practitioner needs, but offers some new of mathematical challenges. First one needs to characterize the new structure and properties that the risk measure has,

depending on the type of pricing measure that is used. After characterizing this new class of risk measures more appropriate time-consistency notions have to be determined. Here, indeed, all the notions of time-consistency introduced in the literature must be reinterpreted. This setting is briefly treated in the last part of Chapter 2, however it deserves more attention and it is our intent to build upon this research.

Arguably, constructing a risk measure that is consistent with future risk measurements, but ignores that these are substantially affected by model uncertainty, is of little use from a practical point of view. Hence, in Chapters 3 and 4 we discussed different approaches to reduce the residual estimation risk due to parameter and model uncertainty and verified their effectiveness via Monte-Carlo simulation studies.

It would be interesting to relate more closely the research done on dynamic risk measurement and on model uncertainty. The focus of this work would then be to construct risk measurement procedures like the ones proposed in Chapters 3 and 4, which allowing for model uncertainty in the probability space, are still able to produce assessments that are somehow consistent over time. The key point here is to define a new class of risk measures where it is possible to identify and separate the two sources of randomness: randomness due to model uncertainty and due to the stochastic nature of the process. In particular the component due to model uncertainty should decrease with time as the estimation procedure will be based on more data points becoming available and thus will be more accurate. A possible strategy would be to consider a worst case approach such as the one proposed in Chapter 4, where we calculate the risk measure according to different candidate models in a certain set $\mathcal{M}$ and then take the worst outcome. Allowing for a dynamic component in the selection of the set of models $\mathcal{M}$ used, would be a first step to include model uncertainty in dynamic risk measurements.

Another interesting link among the work done in the first and second part of this thesis relates to the notion of sequential consistency and residual estimation risk introduced respectively in Chapters 2 and 3. From Chapter

2 a risk measure is sequentially consistent if

$$\rho(Y - \rho_t(Y)) = 0. \tag{5.1}$$

Similarly, a capital estimator approach eliminates the residual estimation risk if

$$\rho(Y - \eta(\mathbf{X})) = 0. \tag{5.2}$$

Equation (5.1) seeks to eliminate the residual risk due to uncertainty in the future outcome of the risk measurement. Equation (5.2) aims at eliminating the residual risk due to the estimation procedure. It would be interesting to investigate whether approaches to deal with parameter and model uncertainty are useful for obtaining sequentially consistent risk measurement and vice versa.

Bibliography

Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance 26*(1), 1505–1518.

Acerbi, C. (2007). Coherent measures of risk in everyday market practice. *Quantitative Finance 7*(4), 359–364.

Acerbi, C. and D. Tasche (2002). On the coherence of Expected Shortfall. *Journal of Banking and Finance 26*(7), 1487–1503.

Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*(3), 203–228.

Artzner, P., F. Delbaen, J.-M. Eber, D. Heath, and H. Ku (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research 192*(1), 5–22.

Barrieu, P. and N. El Karoui (2004). Optimal derivatives design under dynamic risk measures. In G. Yin and Q. Zhang (Eds.), *Mathematics of Finance*. American Mathematical Society.

Basel Committee on Banking Supervision (2009). Revisions to the Basel II market risk framework.

Bellini, F. and M. Frittelli (2002). On the existence of minimax martingale measures. *Mathematical Finance 12*(1), 1–21.

Berger, J. and L. R. Pericchi (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association 91*(433), 109–122.

Bernardo, J. and A. Smith (2000). *Bayesian Theory*. Wiley.

Bion-Nadal, J. (2004). Conditional risk measure and robust representation of convex conditional risk measures. Working paper.

Branger, N. and C. Schlag (2004). Model risk: a conceptual framework for risk measurement and hedging. Working paper.

Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*. Springer.

Cairns, A. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics 27*(3), 313–330.

Carlier, G. and R. Dana (2003). Core of convex distortions of a probability. *Journal of Economic Theory 113*(2), 199–222.

Cheridito, P., F. Delbaen, and M. Kupper (2006). Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability 11*(3), 57–106.

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

Christoffersen, P. (2011). *Elements of Financial Risk Management* (Second ed.). Academic Press.

Cont, R., R. Deguest, and G. Scandolo (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance 10*(6), 593–606.

Delbaen, F. (2002). Coherent risk measures on general probability spaces. In K. Sandmann and P. J. Schönbucher (Eds.), *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, pp. 1–37. Springer.

Denneberg, D. (1994a). Conditioning (updating) non-additive measures. *Annals of Operations Research 52*(1), 21–42.

Denneberg, D. (1994b). *Non-additive Measure and Integral*. Kluwer.

Denuit, J. M., Dhaene, M. Goovaerts, and R. Kaas (2005). *Actuarial Theory for Depedent Risks*. Wiley.

Detlefsen, K. and G. Scandolo (2005). Conditional and dynamic convex risk measures. *Finance and Stochastics 9*(4), 539–561.

Dowd, K. and D. Blake (2006). After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance 73*(2), 193–229.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B 57*(1), 45–97.

Duffie, D. and J. Pan (1997). An overview of Value at Risk. *Journal of Derivatives 4*(3), 7–49.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics 75*(4), 643–669.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

Epstein, L. and M. Schneider (2003). Recursive multiple-priors. *Journal of Economic Theory 113*(1), 1–31.

European Insurance and Occupational Pensions Authority (2009). Directive 2009/138/ec of the European Parliament and of the Council.

Fasen, V. and A. Svejda (2010). Time consistency of multi-period distortion measures. *Statistics and Risk Modeling 29*(2), 133–153.

Föllmer, H. and T. Knispel (2011). Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. *Stochastics and Dynamics 11*(2), 333–351.

Föllmer, H. and I. Penner (2006). Convex risk measures and the dynamics of their penalty functions. *Statistics and Decisions 24*(1), 61–96.

Föllmer, H. and A. Schied (2002). Convex measures of risk and trading constraints. *Finance and Stochastics 6*(4), 429–47.

Föllmer, H. and A. Schied (2004). *Stochastic Finance: An Introduction in Discrete Time* (Second ed.). de Gruyter.

Föllmer, H. and M. Schweizer (1991). Hedging of contingent claims under incomplete information. In M. H. A. Davis and e. R. J. Elliott (Eds.), *Applied Stochastic Analysis*, pp. 389–414. Gordon and Breach.

Frittelli, M. (2002). The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance 10*(1), 39–52.

Frittelli, M. and E. Rosazza Gianin (2002). Putting order in risk measures. *Journal of Banking and Finance 26*(7), 1473–1486.

Frittelli, M. and G. Scandolo (2006). Risk measures and capital requirements for processes. *Mathematical Finance 16*(4), 589–612.

Gerber, H. (1979). *An Introduction To Mathematical Risk Theory*. Irwin.

Gerrard, R. and A. Tsanakas (2011). Failure probability under parameter uncertainty. *Risk Analysis 31*(5), 727–744.

Gibson, R. (ed.) (2000). *Model Risk, Concepts, Calibration and Pricing*. Risk Waters Group.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Goovaerts, M., F. D. Vylder, and J. Haezendonck (1984). *Insurance Premiums: Theory and Applications*. North-Holland.

Gourieroux, C. and W. Liu (2006). Sensitivity analysis of distortion risk measures. Working paper.

Guillen, M., J. Gustafsson, J. P. Nielsen, and P. Pritchard (2007). Using external data in operational risk. *Geneva Papers 32*, 178–189.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

Heyde, C., S. Kou, and X. Peng (2007). What is a good risk measure: Bridging the gaps between data, coherent risk measures, and insurance risk measures. Working paper.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*(4), 382–401.

Hogg, R. V., J. W. McKean, and A. T. Craig (2012). *Introduction to Mathematical Statistics* (Seventh ed.). Pearson.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press.

Jorion, P. (1996). Risk^2: Measuring the risk in Value at Risk. *Financial Analysts Journal 52*(6), 47–56.

Klugman, S. (1992). *Bayesian Statistics in Actuarial Science*. Kluwer.

Klugman, S. A., H. H. Panjer, and G. E. Willmot (2008). *Loss Models: From Data to Decisions* (Third ed.). Wiley.

Koopmans, T. (1960). Stationary ordinal utility and impatience. *Econometrica 28*(2), 287–309.

Kupper, M. and W. Schachermayer (2009). Representation results for law invariant time consistent functions. *Mathematics and Financial Economics 2*(3), 189–210.

Kusuoka, S. (2001). On law-invariant risk measures. *Advances in Mathematical Economics 3*, 83–95.

Markowitz, H. M. (1951). Portfolio selection. *Journal of Finance 7*(1), 77–91.

McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

McNeil, A., R. Frey, and P. Embrechts. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton University press.

Nešlehová, J., P. Embrechts, and V. Chavez-Demoulin (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk 1*(1), 3–25.

Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic Processes and Their Applications 112*(2), 185–200.

Roorda, B. and H. Schumacher (2007). Time consistency conditions for acceptability measures with an application to Tail Value at Risk. *Insurance: Mathematics and Economics 40*(2), 209–230.

Roorda, B. and H. Schumacher (2008). When can a risk measure be updated consistently? Working paper.

Roorda, B., J. M. Schumacher, and J. Engwerda (2005). Coherent acceptability measures in multiperiod models. *Mathematical Finance 15*(4), 589–612.

Rosazza Gianin, E. (2006). Risk measures via g-expectations. *Insurance: Mathematics and Economics 39*(1), 19–34.

Severini, T., R. Mukerjee, and M. Ghosh (2002). On exact probability matching property of right-invariant priors. *Biometrika 89*(4), 952–957.

Stadje, M. (2010). Extending dynamic convex risk measures from discrete time to continuous time: A convergence approach. *Insurance: Mathematics and Economics 47*(3), 391–404.

Tsanakas, A. (2004). Dynamic capital allocation with distortion risk measures. *Insurance: Mathematics and Economics 35*(2), 223–243.

Tutsch, S. (2006). *Konsistente und konsequente dynamische Risikomasse und das Problem ihrer Aktualisierung*. Ph. D. thesis, Humboldt-Universität zu Berlin.

Tutsch, S. (2008). Update rules for convex risk measures. *Quantitative Finance 8*(8), 833–843.

Černý, A. and S. Hodges (2000). The theory of good-deal pricing in incomplete markets. In M. D. P. S. R. V. T. Geman, H. (Ed.), *Mathematical Finance-Bachelier Conference*, pp. 175–202. Springer.

Wang, S. S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin 26*(1), 71–92.

Wang, S. S. and V. R. Young (1998). Risk-adjusted credibility premiums using distorted probabilities. *Scandinavian Actuarial Journal 98*(2), 143–165.

Wang, S. S., V. R. Young, and H. H. Panjer (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics 21*(2), 173–183.

Weber, S. (2004). *Measures and Models of Financial Risk*. Ph. D. thesis, Humboldt-Universität zu Berlin.

Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance 16*(2), 419–441.

Wüthrich, M. V. and R. Salzmann (2010). Cost-of-capital margin for a general insurance runoff. *ASTIN Bulletin 40*(2), 415–451.