I hereby declare that I am the sole author of this thesis. This is a true copy of this thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis compares insurance premium principles with current financial risk paradigms and uses distorted probabilities, a recent development in premium principle literature, to synthesize the current models for financial risk measures in banking and insurance. This work attempts to broaden the definition of value-at-risk beyond the percentile measures. Examples are used to show how the percentile measure fails to give consistent results, and how it can be manipulated. A new class of consistent risk measures is investigated.

... lay up treasure .. as a firm foundation for the coming age ...

– Timothy 6:19

Acknowledgements

It is my pleasure to acknowledge the guidance and support of my thesis supervisors: Dr. Mary Hardy for her endless patience, encouragement, insight and guidance and Dr. Harry Panjer for giving me an opportunity and inspiring me to succeed.

As well, I wish to thank my thesis committee: Dr. Ken Vetzal, Dr. Ken Seng Tan, Dr. Gord Willmot and Dr. Jed Frees (University of Wisconsin) for the time they dedicated to review my thesis, for their encouragement and helpful suggestions.

Special thanks go to Dr. Shaun Wang, who piqued my interest in this field and from whose work this thesis originated.

I would also like to acknowledge Dr. Don McLeish, Dr. Jiahua Chen and Dr. Thierry Duchesne for numerous stimulating discussions and Dr. Phelim Boyle for his encouragement and belief in my abilities.

My thanks also go out to The Society of Actuaries (SOA), The Casualty Actuarial Society (CAS), The Natural Sciences and Engineering Research Council of Canada (NSERC), The Department of Statistics and Actuarial Science, The Institute of Insurance and Pension Research (IIPR), Dr. Harry Panjer, Dr. Mary Hardy, Mr. and Mrs. A.J. Galloway and Dr. and Mrs. D.Q. Vining as they all provided generous financial support.

My deepest gratitude goes to all of my friends and family for their moral support, understanding and help which allowed me to stay focused and maintain a good quality of life during this process. In particular I would like to acknowledge the help of Linda Lingard, Christian Boudreau, Diego Hernandez, Kathy Lepp, Bob Preston and Sheryl Mejia.

Finally, I am forever indebted to my parents for their support, understanding, endless patience and encouragement when it was most required. You are my heroes.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation studies risk measures for capital requirements. The insufficiency of the current methods is illustrated and a new set of risk measures is proposed. Chapter 1 motivates the valuation of a capital requirement, gives definitions for risk and risk measure and uses the premium principle literature to define properties that are desirable in a risk measure. Chapter 2 discusses risk measures that are currently used for capital requirements; illustrates their inconsistencies and considers solutions that have been proposed in recent literature. Chapter 3 investigates a new set of distorted risk measures, illustrates how they can improve upon the risk measures discussed in Chapter 2. Two special cases of the new risk measure are considered in depth. Chapter 4 generalizes the two special cases from Chapter 3 and studies this larger set and attempts to determine parameters appropriate for capital requirements. Chapter 5 applies the risk measures from Chapters 3 and 4 to topical problems in insurance. Chapter 6 summarizes the results and proposes problems for future research. Portions of this work have been published in the North American Actuarial Journal (Wirch, 1999) and have been accepted for publication in Insurance: Mathematics and Economics (Wirch and Hardy, 2000).

1.1 Motivation

Over the past decade there has been an increased incidence of insurance insolvency. In the case of Confederation Life (McQueen, 1996), the most recent of the insolvencies, a large part of the blame falls on the lack of diversification of risk. As much as 60% of Confederation Life's asset portfolio was exposed to changes in real estate prices. When real estate prices fell, this exposure was too great from which to recover. Since then, modeling risk and determining appropriate levels of capital have become of increased significance for actuaries and insurance regulators.

One focus of this attention is on answering the question: how should we measure risk? The goal for financial regulators is to find a methodology for measuring risk that is simple to implement and understand, yet is able to accurately compare divisions within a firm as well as between different corporations, and which makes the risks of each company transparent to the risk holders, including shareholders, debtors, owners, policyholders, employees and potential investors.

In insurance, asset portfolios are allocated to funds in an attempt to match assets and liabilities. The premiums from an insurance liability, say a ten-year endowment insurance, are invested in assets that are often matched in duration to the estimated duration of the liabilities (10 years) with some allowance for surrenders. The premiums are calculated based on an assumed mortality table and interest rate model. For ten-year term insurance, premiums might be invested in assets of varying duration, so that they match the benefit payments required, based on an assumed mortality table, lapse table and interest rate model.

In banking, risk measures are traditionally used to analyze asset portfolios without reference to liabilities. Assets such as stocks, bonds, loans and mortgages are held by the bank using deposits from individuals and corporations. The risk from these investments is initially borne by the bank; however, the rate of interest earned on the deposits fluctuates based on the returns earned by the bank. The majority of liabilities, not including deposits, are cleared through the bank, so that the bank holds no risk. For example, bonds and T-bills, with fixed or floating rates of return, are sold through the bank by a corporation or government, and the investment risk is born by the individual or corporation investing in these instruments. Mutual funds sold by a bank are pools of individuals' investments, invested by the bank into stocks and bonds that determine the return earned on the fund, thus the full risk is born by the investor. Thus the traditional risks that banks were concerned with relate predominantly to their assets, and risk measures relating to fund performance and asset management are of key importance.

Over the past decade, corporations have been investing more often in derivative products to hedge risk. Banks assuming these derivative products may not be able to find an investor to take the opposite position in order to clear the risk, and may hold the risk themselves, hedged by selling a set of assets/liabilities that replicate the opposite position as best they can. In this case, the derivative may not be completely matched, and the risk of both positions should be evaluated. To hedge the risk in the derivative, the bank might have to assume an interest rate model, a credit risk model for the corporation, and a stock return model. Based on these assumptions, they then invest in a hedged position that matches the risk in the derivative. The financial instruments used in this example may be a combination of assets and liabilities that are directly linked to each other. Using the traditional approach to managing risk by focusing on managing the risk in the assets by themselves does not take this into account and may increase the risk overall.

Similarly, over the past decade, insurers have used derivatives to hedge some of the risks, especially interest rate risk, inherent in their liabilities, and have sold products containing options and guarantees, which complicate their assets and liabilities.

In light of these changes, and the growing complexity of the assets and liabilities in banking and insurance, the risk measures that have been traditionally used must evolve, and overall risk measures that encompass the whole company or financial conglomerate will have to be implemented.

1.2 Risk

Before a risk measurement system can be introduced, it is imperative to know the basic unit to which it will be applied. A risk may be defined as an exposure to events that may cause economic loss; the risk may be one bond, a portfolio of assets and liabilities, or an entire firm. The exposure events are a subset of the possible outcomes of the world and may include variables such as economic indicators, prices of goods, services and financial instruments as well as sociological and environmental indicators.

In this thesis, it is assumed that all risks have been specified, and that the set of events that will cause economic loss are known.

Let $X(\omega)$ be a random variable defined on the probability space (Ω, P, F) and let X_0 be its non-random initial value.

Define the set of events, L, as the set of losses, such that,

$$L = \{\omega \| X(\omega) < X_0, \omega \in \Omega\}.$$

Similarly define the set of gains, G, as,

$$G = \{\omega \| X(\omega) \geq X_0, \omega \in \Omega\}.$$

What is unknown is which event will occur; however, based on this information, a distribution for the change in portfolio value can be determined. The distribution for the change in portfolio value can be referred to as the gain distribution, where positive values pertain to gains and negative values pertain to losses. The gain distribution is often used in economics or finance. The negative of this distribution is called the loss distribution, or the risk distribution, where positive values pertain to losses and negative values pertain to gains. Actuaries predominantly use the loss distribution. It is important to identify whether we are analyzing the loss distribution or the gain distribution, and unless otherwise stated, we will use the loss distribution.

The risks considered in this paper are meant to be general. The main applications are to the investment portfolios of insurance companies, banks, securities firms, trading houses and other investment companies. The main difference between insurance portfolios and other investment firm portfolios is that insurance companies tend to invest in assets that match their liabilities, or hedge a liability risk. They are not generally very concerned about day-to-day fluctuations in the value of their asset portfolio, but focus on much longer time scales, and only have to prove solvency to regulators once or twice a year. Investment firms tend to engage in more speculative investments, and have more onerous regulatory requirements; thus, they are much more concerned with day to day changes. With the increased demutualization of insurance companies and the progressive amalgamation of the insurance and banking industries, the regulatory requirements for the insurance industry may veer towards that of the banking industry. We consider applications in both industries.

1.3 Risk Measure

The importance of a risk measure is in its ability to differentiate between different types of risk, its ability to accurately and consistently compare the severity of different risk portfolios, and its ability to be easily understood and applied. Risk measures are usually described in terms of positive numbers which relate to the magnitude of a potential loss, or amount that should be held to cover a risk. The application of a risk measurement technique should be general. Risk values can be used in areas, such as: in the evaluation of investment risk, in the identification of the optimal capital allocation, in the development and evaluation of portfolio strategies, in the measurement of the quality of a portfolio, or in the evaluation of portfolio managers. Unfortunately, the same risk measurement technique may not be appropriate for all the above applications. For example, a risk measure that rewards conservatism may not be appropriate to evaluate portfolio managers of an aggressive fund, but it may be useful for an investor evaluating the fund.

1.4 Review of Premium Principles

In this section, we draw on the literature of premium principles. Insurance premium principles result from the assumption that random claims can be funded by series of fixed payments, where the actual claims experience is considered to be the realization of a 'known' stochastic model (Goovaerts, DeVylder and Haezendonck, 1984). The insurance risk can be described as a non-negative real-valued random variable, X, which represents the claims, or random losses that may occur, and a premium principle is a rule, ρ that assigns a non-negative number to the insurance risk X, which represents the initial capital that needs to be held to cover the risk.

$$\rho : X \to [0, \infty]$$

A set of axioms has been proposed in Goovaerts et al. (1984) to define useful premium principles; Wang (1995c) lists other characteristics and properties that are often desirable for a premium principle. Similar sets of axioms and lists of characteristics can be found in van Heerwaarden (1991), Kaas et al. (1994) and Gerber (1979). We are interested in extending these criteria to general risk measures and identifying whether these characteristics are relevant for investment portfolios.

Premium principles in the insurance industry are formulae which attempt to calculate the appropriate price to transfer a risk. Premiums are often set with the knowledge that there will be a large pool of independent diversifiable risks with similar characteristics. The central limit theorem can be used to get an accurate estimate of the average cost and standard error for the average risk. This is different from risk measures that attempt to measure how much capital should be placed into safe and fairly liquid investments in order to cover potential losses; there is a limited amount of pooling of small risks in an asset portfolio (risk pooling may occur more often in the case of credit risk). However, premium principles and risk measures share many of the same characteristics and many theoretical properties that define a good premium principle can also be applied to define a good risk measure. As well, an initial single premium may be considered the time-zero risk measure from the point of view of the company, though economic pricing issues will also be important.

General risk random variables can take values in the extended real numbers. Artzner et al. (1997) state that the size of possible gains should be irrelevant for a consistent risk measure and therefore a loss distribution left-censored at zero should be used to value the risk for the full loss distribution. This idea is simple when discussed in terms of a capital requirement. If you have a gain of $\$G$, no capital will be required since there is no loss. However, if you have a loss of $\$L$, you will need capital of $\$L$ to pay off the loss. Thus the capital requirement for this risk is \$0 with probability Pr(gain) and the capital requirement is positive with Pr(loss). Even if the amount of the gain was much greater than the amount of the loss, the capital requirement would stay the same, thus we use a loss distribution left-censored at zero to calculate the risk in a portfolio. Using the censored loss distribution, it is possible to directly apply insurance premium principles to measure more general risks.

Denote the risk measure for loss X as $\rho(X) \in R$. Goovaerts et al. (1984) list a set of properties that should hold for a useful premium principle, $\rho(X)$:

A1. No unjustified premium (certain gain):

$$\rho(X) \leq \max(X) \tag{1.1}$$

A2. Non-negative Risk Loading:

$$\mathrm{E}[X] \leq \rho(X) \tag{1.2}$$

A3. Scale Invariant:

$$\rho(aX) = a\rho(X), \qquad \text{for constant} \quad a \geq 0 \tag{1.3}$$

A4. Translation invariant:

$$\rho(X + b) = \rho(X) + b, \qquad \text{for constant} \quad b \geq 0 \tag{1.4}$$

A5. Subadditive:

$$\rho(X+Y) \leq \rho(X) + \rho(Y), \text{ where } Y \text{ is an arbitrary loss random variable.} \tag{1.5}$$

Property A3 ensures that a change in currency does not affect the risk measure. Property A4 ensures that a degenerate risk (specified loss with probability 1) has a premium equal to its certain loss. Often properties A3 and A4 are combined to give a linearity property $\rho(aX+b) = a\rho(X) + b$. Property A5 ensures that there is no incentive to split the risk into smaller risks.

At this point, it is illustrative to show that the translation invariant property does not hold generally for the loss distribution, but only applies to censored loss distributions. Consider portfolio X which has loss distribution $Uniform[-1,1]$. The distribution of the loss censored at zero is given by $X^+ = \max(0, X)$

$$\Pr(X^+ = 0) = \tfrac{1}{2}$$
$$\Pr(X^+ \leq u) = \tfrac{1}{2} + \tfrac{u}{2}, \quad \text{for} \quad 0 < u \leq 1.$$

In other words, X^+ is 0 with probability $\frac{1}{2}$, or $Uniform[0,1]$ with probability $\frac{1}{2}$. If we add a degenerate loss $Y = \frac{1}{2}$ to the uncensored risk X, the loss distribution is,

$$X + Y \quad \sim \quad Uniform\left[-\frac{1}{2}, \frac{3}{2}\right],$$

and the censored loss distribution of $X+Y$ is

$$[X+Y]^+ = \max(0, X+Y) \;\sim\; \begin{cases} \Pr([X+Y]^+ = 0) = \frac{1}{4} \\ \Pr([X+Y]^+ \le u) = \frac{1}{4} + \frac{u}{2}, \quad \text{for } 0 < u \le \frac{3}{2}. \end{cases}$$

In other words, $[X+Y]^+$ is 0 with probability $\frac{1}{4}$, or $Uniform[0, \frac{3}{2}]$ with probability $\frac{3}{4}$, which does not have the translation invariant property. However, if $Y = \frac{1}{2}$ is added to the censored risk X^+ the loss distribution is non-negative,

$$X^+ + Y = \max(0, X) + Y \;\sim\; \begin{cases} \Pr(X^+ + Y = \frac{1}{2}) = \frac{1}{2} \\ \Pr(X^+ + Y \le u) = \frac{1}{2} + \frac{u - \frac{1}{2}}{2} \quad \text{for } \frac{1}{2} < u \le \frac{3}{2}. \end{cases}$$

which does follow the translation invariant property.

In adding a risk to a portfolio, it seems intuitive that the risk would be added to the portfolio before censoring the portfolio loss distribution. This will not lead to translation invariance. Thus there is still some debate over whether to use the whole loss distribution or the censored loss distribution. Using the censored distribution also leads to discrepancies between time zero pricing of risks and risk measures. Although these two values serve very different purposes, there should be some reconciliation between the two.

Other properties listed in Goovaerts et al. (1994) and Wang (1995c), that are often useful for premium calculations are:

Let X and Y be random variables defined on Ω

B1. Iterativity (Goovaerts, DeVylder and Haezendonck, 1984):

$$\rho(X) = \rho(\rho(X|Y)), \tag{1.6}$$

where $X|Y$ is a conditional distribution, and Y is an observable random variable.

B2. Multiplicativity: If X, Y are independent, then

$$\rho(XY) = \rho(X)\rho(Y). \tag{1.7}$$

B3. Comonotonicity(Wang, 1996c): If X and Y are comonotonic, then for any pair of outcomes, t_1 and t_2,

$$[X(t_1) - X(t_2)][Y(t_1) - Y(t_2)] \geq 0 \tag{1.8}$$

$$\text{and} \quad \rho(X+Y) = \rho(X) + \rho(Y). \tag{1.9}$$

B4. Layer Additivity (Wang, 1996b): Given a partition of the domain of X,

$$\{(x_i, x_{i+1}], i = 0, 1, 2, ...\}, \quad 0 = x_0 < x_1 < x_2 < ..., \tag{1.10}$$

and indicator function I such that,

$$I_{(x_i,x_{i+1}]} = \begin{cases} 0 & \text{for all} \quad x < x_i \\ x - x_i & \text{for all} \quad x_i \le x < x_{i+1} \\ x_{i+1} - x_i & \text{for all} \quad x_{i+1} \le x \end{cases} \tag{1.11}$$

so that, $X = I_{(x_0,x_1]} + I_{(x_1,x_2]} + \ldots,$

then

$$\rho(X) = \sum_{i=0}^{\infty} \rho(I_{(x_i,x_{i+1}]}). \tag{1.12}$$

B5. Decreasing Absolute Risk Load (Wang, 1996b):

For $y < x$ and constant h,

$$\rho(I_{(x,x+h]}) \le \rho(I_{(y,y+h]}). \tag{1.13}$$

B6. Increasing Relative Risk Load (Wang, 1996b): For any fixed x,

$$\phi(x) = \lim_{h \to 0} \frac{\rho(I_{(x,x+h]})}{E[I_{(x,x+h]}]}, \tag{1.14}$$

is an increasing function in x.

Note that, as a special case,

$$\Phi(x) = \frac{\rho(I_{(0,x]})}{E[I_{(0,x]}]} \tag{1.15}$$

is also an increasing function in x.

Iterativity suggests a possible method for obtaining marginal risk measures. Multiplicativity parallels the property that $E[XY] = E[X]E[Y]$ for independent risks X, Y. Layer Additivity has applications in stop loss insurance and reinsurance, and is implied for comonotonic risk measures. Decreasing Absolute Risk Load provides that the absolute risk loading decreases at upper layers. Increasing Relative Risk Load ensures that higher levels have greater risk loading relative to the mean loss for that layer.

There are many properties of premium principles which help with the ordering of risks. $X \prec Y$ denotes that X is less risky than Y, and usually pertains to a specific type of ordering such as those listed below (C1-C4). For example a premium principle ρ satisfies a specified risk order if $X \prec Y$ implies that $\rho(X) < \rho(Y)$. Given that $F_X(x)$ is the cumulative distribution function (cdf) of X then $S_X(x) = 1 - F_X(x)$ is the decumulative distribution function (ddf) of X, some ordering properties are defined below:

C1. First Order Stochastic Dominance (Goovaerts, DeVylder and Haezendonck, 1984): If $S_X(t) \leq S_Y(t)$ for all $t \geq 0$ then $\rho(X) \leq \rho(Y)$. (Note: There are many other equivalent conditions. (Wang, 1998))

C2. Ordering of dangerousness (Goovaerts, DeVylder and Haezendonck, 1984): If $E[X] < E[Y] < \infty$ and there is a constant β such that $F_X(t) \leq F_Y(t)$ for $t < \beta$ and $F_X(t) \geq F_Y(t)$ for $t \geq \beta$ then $\rho(X) < \rho(Y)$.

C3. Second Order Stochastic Dominance (Wang, 1996b): If

$$\int_x^\infty S_X(t)dt \leq \int_x^\infty S_Y(t)dt,$$

for all $x \geq 0$, with strict inequality for some $x \in (0,1)$ then $\rho(X) < \rho(Y)$. (Note: There are many other equivalent conditions such as net stop-loss ordering. (Wang, 1998))

C4. Consistent partial ordering (Goovaerts, DeVylder and Haezendonck, 1984): If $A = \{\beta || \frac{\partial}{\partial t}\frac{E(e^{tX})}{E(e^{tY})^\beta} > 0\}$, then for any $\beta \in A$, $\rho(X) \geq \beta\rho(Y)$.

The definition for the ordering of dangerousness, is referred to as the once crossing rule, and relates to second order stochastic dominance by the following proposition from Müller (1996):

Proposition 1.4.1 *X precedes Y (denoted by $X \prec_{SSD} Y$) in second stochastic order if, and only if, there exists a sequence of decumulative distribution functions (ddf) $\{S_1, S_2, ...\}$ such that,*

- $S_1 = S_X$
- *The means for this sequence of ddfs are non-decreasing and converge to $E(Y)$.*
- *S_i and S_{i+1} cross once: there exists t_i such that*

$$S_i(t) \geq S_{i+1} \quad for \;\; t < t_i,$$

$$S_i(t) \leq S_{i+1} \quad for \;\; t \geq t_i.$$

- $S_Y(t) = \lim_{i\to\infty} S_i(t)$.

Thus, it is sufficient to show that

1. $E[X] \le E[Y]$, and
2. There exists a once crossing point t_0 such that

$$\begin{aligned} S_X(t) &\ge S_Y(t) \quad \text{for } t < t_0 \\ S_X(t) &\le S_Y(t) \quad \text{for } t \ge t_0, \end{aligned}$$

for $X \prec_{SSD} Y$.

Second order stochastic dominance (SSD) has become a common standard for relating risks, and is equivalent to net stop-loss ordering (Wang, 1996b).

A list of traditional premium principles is given below. Based on a loss distribution for loss random variable X, an equation for the premium, $\rho(X)$, is given for each premium principle (Goovaerts, DeVylder and Haezendonck, 1984):

D1. Expected Value principle: $\rho(X) = (1+a)E[X], \quad a \ge 0.$

D2. Maximum loss principle: $\rho(X) = aE[X] + (1-a)\max(X), \quad 0 \le a \le 1.$

D3. Generalized Percentile Principle: $\rho(X) = aE[X] + (1-a)r_\epsilon$,
$r_\epsilon = \min\{r | F_X(r) \ge 1-\epsilon\}, \quad 0 \le a \le 1.$

D4. Variance principle: $\rho(X) = E[X] + a\sigma_X^2, \quad a \ge 0.$

D5. Standard Deviation Principle: $\rho(X) = E[X] + a\sigma_X, \quad a \ge 0.$

D6. Semi-Variance Principle: $\rho(X) = E[X] + a\sigma_+^2(X)$,

$\sigma_+^2(X) = \int_{E[X]}^{\infty}(x - E[X])^2 dF_X(x), \quad 0 < a \leq 1.$

D7. Exponential Principle: $\rho(X) = \frac{1}{a}\log\{E(e^{aX})\}, \quad a > 0.$

D8. Mean value principle: $u(\rho(X)) = E[u(X)]$, for any strictly increasing, concave function u.

D9. Zero-Utility Principle: $E[u(W + \rho(X) - X)] = E[u(W)]$, for any strictly increasing, concave function u, where W is the initial wealth of the insurer.

D10. Swiss Principle: $E[u(X - a\rho(X))] = u((1-a)\rho(X))$, for any strictly increasing, concave function u, $0 < a \leq 1$.

D11. Orlicz principle: $E[\phi(\frac{X}{\rho(x)})] = \phi(1)$, for any continuous increasing, convex ϕ.

D12. Dutch Principle: $\rho(X) = E[X] + aE[(X - \alpha E[X])_+], \alpha \geq 1, \quad a \geq 1.$

D13. Esscher Principle: $\rho(X) = E[Xe^{aX}]/E[e^{aX}], \quad a \geq 0.$

Each of the above principles either does not follow all of properties A1-A5, or have a property that would limit its range of application as a risk measure. D1 only depends upon the expected value which is not translation invariant, and gives an unjustified premium for degenerate risks. If the loss distribution is unbounded, D2 is either infinite or gives the expected loss. D3 will be shown to allow super-additivity (see Example 2.2.3). D4, D6, D7, D8, D9, D10, D11 and D13 do not allow both proportional and translation invariance. D4 and D5 do not exist if the second moment of the distribution does not exist. D12 has a maximum relative risk loading of 100%.

Although the purpose of a financial risk measure is not the same as the purpose of an insurance premium principle, there are many similarities in the risks involved. In both cases, it is important to realize that the size of possible positive outcomes (the net of premium income less claims in the case of (re)insurance, or gains in the case of an asset portfolio) is often irrelevant; however, the probability of obtaining a positive outcome is relevant. Since the effects of a negative outcome are the main concern, financial risk measures often consider all gains to be zero losses. This makes the financial risk similar to the risk of the insurance provider who does not participate in the gains of their clients, beyond the set premium.

The set of properties A1-A5 for a useful premium principle has recently been applied to financial risk measures. Artzner (1999) adapted these principles to define a coherent risk measure.

Definition 1.4.1 *Consider two arbitrary risks X and Y. A risk measure ρ is called a* **coherent** *risk measure if it satisfies the following characteristics:*

A1, A2. *A risk measure should be bounded above by the maximal loss, and bounded below by the expected value of the loss: $E[X] \leq \rho(X) \leq max(X)$.*

A3. *A risk measure should be scale invariant: $\rho(aX) = a\ \rho(X), \quad a \geq 0$.*

A4. *A risk measure should be scalar additive(translativity):*

$$\rho(X + b) = \rho(X) + b, \quad b \geq 0,$$

and a degenerate risk should have a risk measure equal to its certain loss:
$If \quad Pr(X = b) = 1, \quad then \quad \rho(X) = b, \quad b \geq 0.$

A5. *A risk measure should be sub-additive, so that there is no motivation to divide the risk:* $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

The characteristics of a premium principle determine its usefulness. Many authors consider a set of rules, analyze the traditional principles based on these rules, and often develop new principles which follow these rules. Deprez and Gerber (1985) consider convex premium principles that are invariant under translation, and show that the principles are sub-additive if and only if the premium principle is also proportional, ie. $\rho(aX) = a\ \rho(X), \quad a \geq 0$.

Artzner et al. use their definition of a coherent risk measure to identify weaknesses in current risk measures and propose a new risk measure, using a conditional expectation, which will be discussed later in Section 2.5.

Wang (1995a) uses a similar set of properties for risk premiums, and proposes a new family of premium principles using distorted probabilities. These will also be discussed later in this paper. Wang, Young and Panjer (1997) use this same set of properties with an extra criterion which results in a unique distorted premium principle.

Chapter 2

Problems with Current Risk Measures

In this chapter, current capital requirement risk measures are investigated using the coherency properties. Examples are used to illustrate where these risk measures fail to give consistent results. Possible solutions that have been proposed in recent literature are studied and applied to four two-parameter distributions.

2.1 Short Review of Value-at-Risk

The phrase 'Value-at-Risk' (VaR) has become synonymous with the percentile risk measure (Duffie and Pan, 1997; Morgan, 1995; Hull, 1997), which is identical to the percentile premium principle, D3 on page 16 with $a = 0$. The typical VaR measure uses a loss distribution for daily changes to the risk and the 95th or 99th percentile. Generally, for a portfolio P, and an associated n-day loss random variable L_n, the

α percentile-VaR, or $V_\alpha(P)$, can be determined by solving the following equation (Duffie and Pan, 1997):

$$Pr(L_n \leq V_\alpha(P)) = \alpha. \tag{2.1}$$

In this thesis, the percentile definition of the value-at-risk will always be referred to as the percentile-VaR.

The percentile-VaR risk measure has gained a lot of attention over the past decade. Increased activity of banks in the derivatives markets and the rate at which these markets are expanding, becoming global and more complex, has stimulated concern over the risk management practices used in banks and other financial institutions. The Basel Capital Accord (Basel Committee, 1999), first drafted in 1988, proposed a set of international capital requirements for banks and other financial investment firms based on the inherent volatility of their individual assets, and accounting mostly for credit risk. The requirements were determined separately for each asset, using the risk-based capital approach of multiplying the nominal value of each type of asset by a capital charge or default rate. Then the capital requirements for each asset type were added together to obtain the capital requirement of the portfolio. This method did not permit reduced capital requirements for hedged portfolios. In the 1993 revision of the Accord, a standard model for the evaluation of the capital requirement computed the capital requirement, using the percentile-VaR measure applied to modeled or simulated distributions for each of four risks: interest rate risk, exchange rate risk, commodity risk and equity risk, and summed across the four categories to obtain the final capital requirement. Unfortunately, this method did not permit a reduction in capital requirements for diversified portfolios, but allowed some reduction for hedged portfolios. To account for the sub-additivity of diversified or hedged risks, the 1995 revisions allowed the use of computer models to calculate the daily risk of a portfolio using the percentile-VaR, as long as the models complied with Commission standards. The capital requirement, which is supposed to cover a ten-day period, is calculated at three times the percentile VaR measure. The ten-day period is considered to be a reasonable period over which it is possible to make a significant change in a corporate portfolio. This percentile-VaR measure was intended to be an appropriate amount to cover most losses that could occur in a portfolio of assets due to adverse changes in any of these four risks before the investment strategy could be revised and implemented. The Basel rules came into effect in 1998. A revised capital adequacy framework has been proposed in 1999 by the Basel Committee, which promotes the development of internal capital assessment processes by bank management and a more comprehensive approach to addressing risk including operational, liquidity, legal and reputational risk as well as the current focus on credit and interest rate risk (Basel Committee 1999).

In Europe, capital adequacy requirements have moved toward a common rule. The European Union's Investment Services Directive allows firms based in one EU country to do business in any other EU country. As well, the Capital Adequacy Directive(CAD), published in 1993, provides Europe-wide capital requirements which were similar to those of the 1993 Basel Accord revisions. More recent revisions allow the use of in-house models for risk calculations and also use percentile-VaR measures. The CAD guidelines were put into effect in 1996 for all banks and security firms.

Risk-based capital (RBC) methods are currently used in Canada (Minimum Continuing Capital and Surplus Requirements) and the United States to calculate minimum capital requirements for insurers. Capital margins are calculated independently for each type of asset within each type of risk based on ad hoc capital charges which are multiplied by the nominal value of assets. The four main types of risk are: C1 - Asset Risk, C2 - Underwriting Risk, C3 - Interest Rate Risk, and C4 - Other Risks. The margins are combined using formulae specific to the type of insurer that have been derived to allow for some correlation between risk types. The capital charges for life insurers are slightly different from the capital charges for casualty insurers, due to the nature of their assets and liabilities; however both suggest dynamic financial analysis or cashflow testing to validate the adequacy of the capital requirement.

Percentile-VaR is predominantly used in determining the capital requirements for C1 or asset risk, for which a loss distribution can be modeled objectively; it may be significantly different between insurers, depending on their investment strategy and management philosophy. In contrast, the RBC capital charges are usually based on credit rating of the assets, and are determined independently of any specific asset portfolio by the regulators.

The usual percentile-VaR definition for the value-at-risk reduces the information in the loss distribution to one number, or possibly a few numbers, and the user loses much of the information needed to fully understand the risk of the portfolio.

Using this measure for one portfolio will give very simple, easily interpreted results: For a risk X, $V_{.95}(X)$ is the value that represents a division in the loss distribution of X. There is a 95% chance that the loss will be less than this amount, and a 5% chance that the loss will be larger than this amount (subject to an accurate assessment of the loss distribution). Thus, there is a 95% probability that the percentile-VaR is sufficient, subject to model error.

The methodologies for calculating VaR take into consideration market factors, such as domestic and foreign interest rate structures, exchange rates, stock prices and inflation rates, which are examples of risk factors that may have an impact on the financial risks of a portfolio. There are many methods which use these market factors to determine the distribution of L_n. These methods can be classified into three model types: the historical model, the analytic model and the simulation model. Each of these methods determines a distribution for changes to the portfolio value. Below we outline how the percentile-VaR measure can be used with each method.

Historical Model: In the historical approach, using a one day holding period, previous one day fluctuations (typically using daily data for the last 10 years) in market factors are used to model possible fluctuations to current market factor values. Alternative profit/loss realizations are valued based on these fluctuations and a distribution for profit/loss can be obtained from these realizations to produce a confidence interval. The algorithm is:

Let $F_d = (f_{(1,d)}, f_{(2,d)}, ..., f_{(k,d)})$ the vector of observed risk factor values on day d, where $d = 0, -1, ..., -n$, $n =$ number of days used;

$P(F) =$ the value of a portfolio, using the factor values F;

$F_0 =$ today's risk factor values;

$\Delta F_d = F_d - F_{d-1}$ are one-day factor changes;

$P_0 = P(F_0)$ Initial value of the portfolio;

$P_d = P(F_0 + \Delta F_d)$;

Order the P_d's to get $\{P_{(d)}\}_{d=1}^{n}$, where $P_{(d+1)} \geq P_{(d)}$;

Assign equal probability to each observed factor change;

Then, $V_\alpha(P_0) = P_0 - P_{(\alpha n)}$.

Simulation Model: An alternative method, using simulation, requires a distribution for changes in each market factor, including correlations between factors. Normal and lognormal distributions are often used, with correlations derived from historical data. Given distributions for each of the risk factors, Monte Carlo simulation is used to obtain simulated changes in the market factors, which are used to obtain a profit/loss distribution and confidence intervals in the same way as in the historical method, using the algorithm:

Obtain the joint density function of the risk factors $F = (f_1, f_2, ..., f_k)$;

Let n be number of simulations;

Simulate n vectors $\{F_j\}_{j=1}^{n}$;

Calculate $\{P_j\}_{j=1}^{n}$, the value of the portfolio for factor values $\{F_j\}_{j=1}^{n}$;

Order the P_j's to get $\{P_{(j)}\}_{j=1}^{n}$, where $P_{(j+1)} \geq P_{(j)}$;

Let $P_0 =$ the initial value of the portfolio.

Then, $V_\alpha(P_0) = P_0 - P_{(\alpha n)}$.

Analytic/Variance-Covariance Model: A more restrictive approach, the analytic method, decomposes the portfolio into elemental instruments each of which is exposed to only one market factor. A set of distributions for changes in the market factors is used to calculate the VaR and the portfolio variance. Since the portfolio is the sum of the elemental instruments, if the market factors have a multivariate normal distribution, then the portfolio is also normally distributed. The algorithm for this model is as follows:

> Let the random variables $P_{f_1}, P_{f_2}, ..., P_{f_k}$ be the decomposition of Portfolio P into component securities.
>
> Then,
>
> $$P = \sum_{i=1}^{k} P_{f_i}.$$
>
> Assuming that the component securities are related through a known covariance structure, using the multivariate normal distribution, the portfolio distribution can be calculated, and $V_\alpha(P_0)$ obtained from the distribution.

Even though these models seem intuitively reasonable and they are easy to explain, their tractability is based on the assumption that the percentile-VaR measure is sub-additive. Unfortunately, it is possible to show that percentile-VaR, as with all percentile measures, can be super-additive (see Example 2.2.3). There are simple examples that reveal the inconsistencies of the percentile-VaR, some of these are illustrated in the next section.

2.2 The Inconsistencies of VaR

Using the five basic properties for coherent premium principles, it is easy to show that percentile-VaR satisfies properties A1 and A4.

For A1: Let L be a loss random variable censored from below at zero, where $\max(L) \geq 0$ is the maximum loss. For $0 \leq \alpha \leq 1$, we have that $V_\alpha(L)$ is a increasing function of α, and thus obtains its maximum when $\alpha = 1$, thus

$$V_\alpha(L) \leq V_1(L) = \max(L). \tag{2.2}$$

For A4: Given some arbitrary α, and some constant $b > 0$, using the same loss random variable L, we know that

$$Pr(L \leq V_\alpha(L)) = \alpha. \tag{2.3}$$

Now, define a new loss random variable $Y = L + b$, then we want to determine $x = V_\alpha(Y)$ so that

$$Pr(Y \leq x) = \alpha. \tag{2.4}$$

Then

$$
\begin{aligned}
Pr(Y \leq x) = Pr(L + b \leq x) &= Pr(L \leq x - b) = \alpha, && (2.5)\\
\text{which implies that} \quad x - b &= V_\alpha(L) && (2.6)\\
\text{and} \quad V_\alpha(L + b) &= V_\alpha(L) + b. && (2.7)
\end{aligned}
$$

Note that if a degenerate risk is added to the uncensored instead of the censored distribution and VaR is calculated, then VaR is not necessarily scalar additive. To illustrate this point, consider the portfolio with loss distribution Q that always produces a gain of $b+1$, $Q = -(b+1)$ for some constant $b > 0$. This portfolio has a negative loss with probability 1 and $V_\alpha(Q) = 0$. As well, adding a degenerate risk of b to the uncensored loss produces a portfolio with loss random variable $Q + b = -1$ always produces a gain of 1, and $V_\alpha(Q + b)$. Thus, for any arbitrary α, $V_\alpha(Q) = 0$, and

$$V_\alpha(Q + b) = 0 \neq V_\alpha(Q) + b. \tag{2.8}$$

It is simple to show that percentile-VaR is not necessarily mean value exceeding (A2). Choosing $\alpha = 0$, V_0 is the minimum value of the loss distribution, which is less than $E[L]$ for any non-degenerate loss.

The failure of percentile-VaR to satisfy subadditivity or proportionality (properties A3 and A5) is illustrated in Examples 2.2.2 and 2.2.3 later in this chapter.

The following three examples (Wirch, 1999) will be used to clarify how a percentile measure falls short of what is desirable in a risk measure. Example 2.2.1 shows how percentile-VaR is unable to differentiate between a risk averse and a risk taking portfolio. Example 2.2.2 shows that percentile-VaR is not proportional and may inadequately order portfolios. Example 2.2.3 identifies a portfolio for which percentile-VaR is super-additive. In these examples we use the loss distributions censored from below at 0.

Example 2.2.1 Let S_0 be the initial price of a stock and let S_n be the price of the stock n days later, where $\log(S_n) \sim Normal(\log(S_0), \sigma^2)$, and where σ is assumed to be a known constant.

Compare the following two portfolios:

Portfolio X:

$$X_n = S_n \tag{2.9}$$

$$E[X_n] = S_0 e^{\frac{\sigma^2}{2}} \tag{2.10}$$

$$\text{so,} \quad L_n(X) = \begin{cases} 0, & S_n \geq S_0 \\ S_0 - S_n, & S_n < S_0 \end{cases} \tag{2.11}$$

Portfolio Y:

$$Y_0 = S_0 \tag{2.12}$$

and for $\sigma^2 < -2\log .95$

$$Y_n = \begin{cases} 0, & \frac{1}{\sigma}\log(\frac{S_n}{S_0}) \le \Phi^{-1}(0.05) \\ \frac{1}{0.95}S_0 e^{\frac{\sigma^2}{2}}, & \frac{1}{\sigma}\log(\frac{S_n}{S_0}) > \Phi^{-1}(0.05) \end{cases} \tag{2.13}$$

$$E[Y_n] = S_0 e^{\frac{\sigma^2}{2}} \tag{2.14}$$

$$\text{so,}\quad L_n(Y) = \begin{cases} S_0, & \frac{1}{\sigma}\log(\frac{S_n}{S_0}) \le \Phi^{-1}(0.05) \\ 0, & \frac{1}{\sigma}\log(\frac{S_n}{S_0}) > \Phi^{-1}(0.05) \end{cases} \tag{2.15}$$

Both portfolios have the same initial value, and the same expected value at time n. The maximal loss of each portfolio is S_0; however, using the 95th percentile, the VaR of portfolio Y is zero, and for portfolio X the percentile-VaR is positive and equal to $S_0(1 - e^{-1.645\sigma})$. The risk in Portfolio Y is isolated to a specific range of outcomes for the stock price. This range of outcomes has a probability of less than 5%, resulting in a percentile-VaR of zero. In portfolio X, the risk of a loss is distributed over a much larger range and the probability of losing everything is negligible. If we rely on percentile-VaR to compare these two portfolios, we would choose Portfolio Y as the least risky; however, if we consider a person who is looking for financial security, a 5% probability of losing everything is excessive.

Example 2.2.2 Let Z_0 be the value of a risk factor today, and let Z_n be the random risk factor n days from now. Assume Z_n has a $Normal(0,1)$ distribution. Compare the loss distributions for the following two portfolios:

Portfolio X:

$$L_n(X) = \begin{cases} 0, & |Z_n| \leq \Phi^{-1}(0.975) \\ 10, & |Z_n| > \Phi^{-1}(0.975) \end{cases} \tag{2.16}$$

Portfolio Y:

$$L_n(Y) = \begin{cases} 0, & |Z_n| \leq \Phi^{-1}(0.975) \\ 100, & |Z_n| > \Phi^{-1}(0.975) \end{cases} \tag{2.17}$$

Both portfolios have losses over the same risk factor values, and both portfolios have losses only in a region having less than a 5% probability of occurring. So both portfolios have a $V_{.95}$ of zero. However, in the region where there is a loss, the loss for portfolio Y is ten times that of portfolio X. If we rely on VaR as a percentile measure, to compare these two portfolios, we would be indifferent between them even though it is clear to see that Portfolio X is the investment with less risk.

Example 2.2.3 Assume that our only risk factor is the price of a stock in n days, S_n, which has a $Lognormal(\mu, \sigma^2)$ distribution, with $\mu = 0.05$ and $\sigma = 0.1$ (see figure 2.1). These parameters are appropriate for a 6 month ($n = 180$) duration. Consider the following two portfolios:

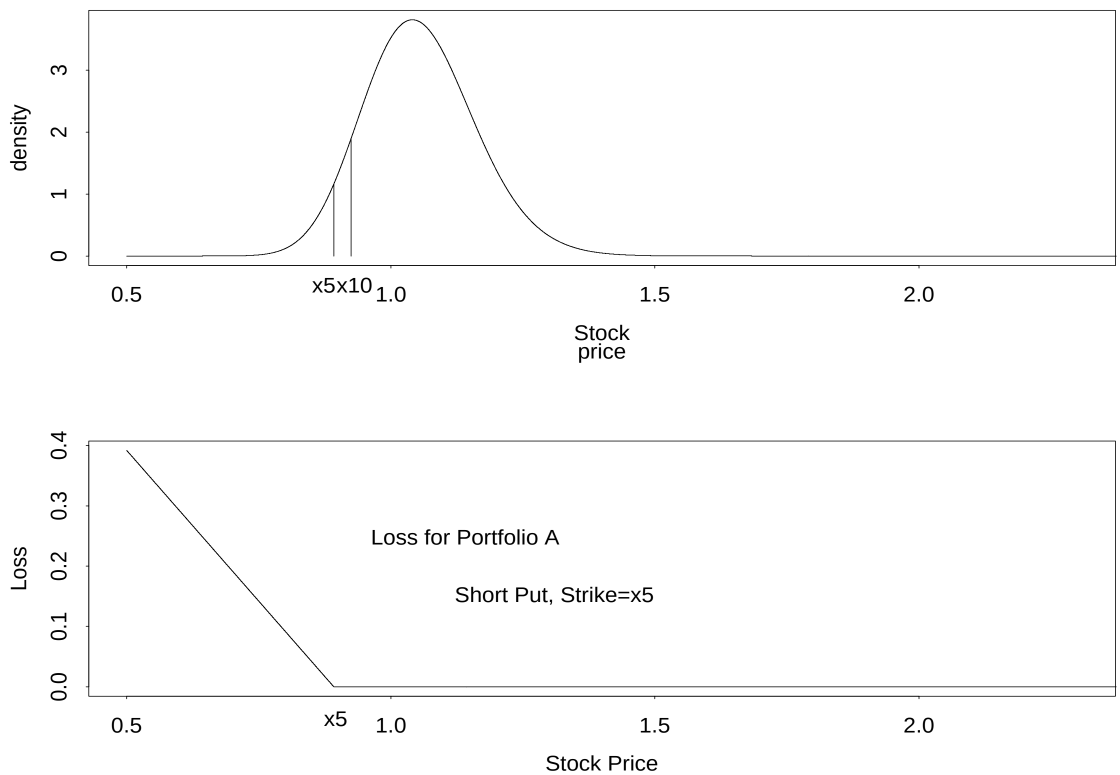

Figure 2.1: Loss Function for Portfolio A

Portfolio A is an n-day short put option with strike price $F_{S_n}^{-1}(0.05) = x_{.05}$ (see Figure 2.1 for the loss function of Portfolio A).

Portfolio B is a reverse butterfly spread, consisting of ten short call options on the same stock, five with a strike price of $x_{.05}$ and the other five with a strike price of $F_{S_n}^{-1}(0.10) = x_{.10}$, and ten long call options on the stock, with a strike price of $\frac{x_{.10}+x_{.05}}{2}$. (See Figure 2.2 for loss function.)

The probability that the reverse butterfly spread will produce a loss is 5%, and the probability that the short put will produce a loss is 5%. Note that if portfolio

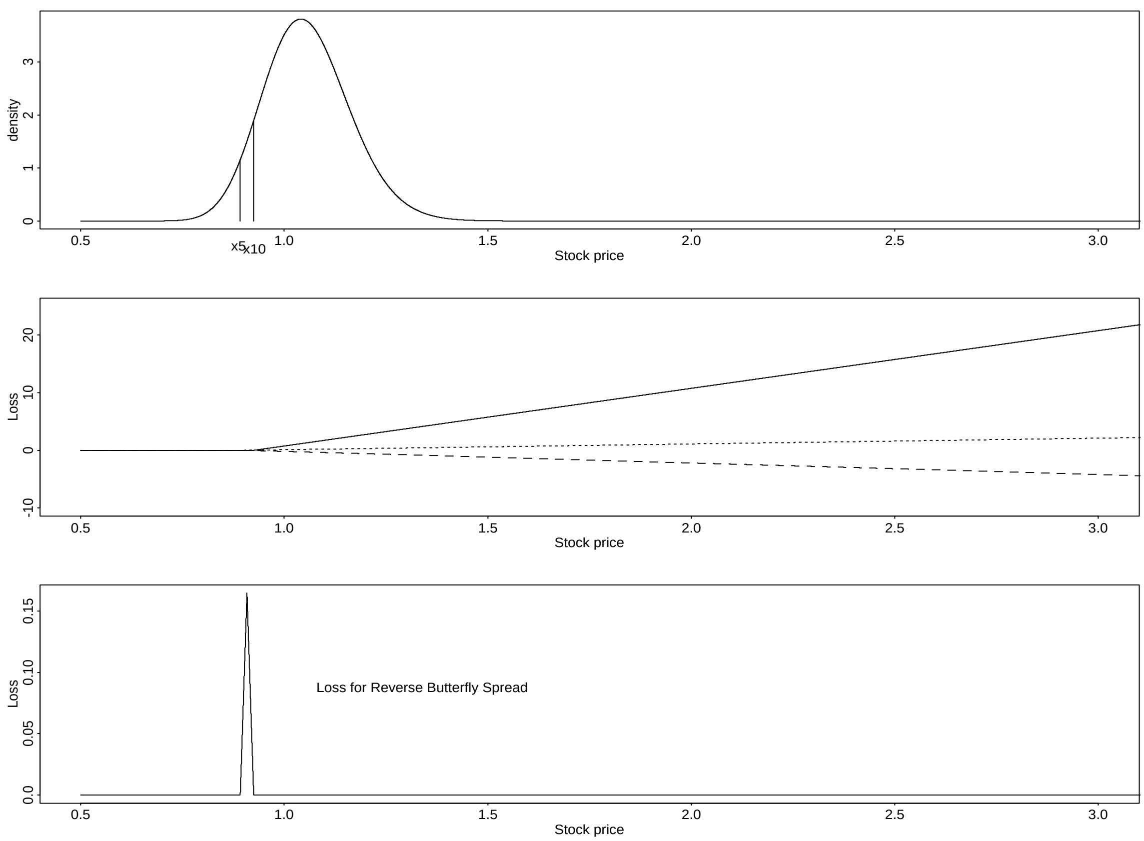

Figure 2.2: Loss Function for Portfolio B

A has a loss, then portfolio B will not have a loss and vice versa; the support of the loss distributions is non-overlapping. Now consider the combined portfolio, $A + B$. The cost of entering into these contracts is not taken into account because they are sunk costs and do not affect the change in value of the portfolio between time 0 and time n. The regions of the distribution of S_n, for which A and B have losses are disjoint. The combined portfolio now has a 10% probability of having a loss.

This implies that the percentile-VaR value for the combined portfolio is positive,

$$V_{.95}(A+B) > 0, \tag{2.18}$$

and greater than the sum of the percentile-VaR measures for the individual portfolios,

$$V_{.95}(A+B) \ > \ V_{.95}(A) \ + \ V_{.95}(B). \tag{2.19}$$

This example has shown that the percentile-VaR can be superadditive, and thus the percentile definition of value-at-risk fails to adhere to the subadditivity property(A5) of a coherent measure.

The potential hazard of percentile-VaR is not that it produces useless results all the time, but that in using derivatives, it is possible to manipulate percentile-VaR by isolating small segments of the joint risk distribution, and concentrating the losses of a portfolio on one of these small-probability segments. In removing the risk(losses) from other outcomes it is necessary to trade for additional risk on the same small-probability segment. By increasing the size of the loss on that segment, we can reduce the probability of the occurrence of a loss. In this way it is possible to set any portfolio's percentile-VaR to zero or any other desired number. This example is revisited in Section 3.3. Considering the above characteristic of percentile-VaR, it is important to ensure that the risk measures used satisfy the coherency requirements.

Often in insurance, liabilities and assets are valued separately and truncated data is often the most readily available for losses. Adding the risk measures or using the truncated data risk measure, will overestimate the total risk when using a sub-additive risk measure. However, this may still be useful to obtain an upper bound on the estimate.

2.3 Improving the Current Risk Measures

Risk functions that utilize more information from the loss distribution tend to be more difficult to implement or more computer intensive. It is very important to consider the types of risks that these functions can identify, and how they would improve upon the simpler models that lose much of the loss distribution information.

We have assumed that the distribution for the loss random variable holds as much information as the user is capable of knowing. Sometimes a few characteristic parameters fully define this distribution. However, loss distributions for complex portfolios rarely follow a simple distribution, and managers usually prefer to see summary characteristics (statistics) of the distribution, rather than the distribution itself.

There are three tools from statistics and actuarial science that will be discussed, which will help to produce a model that may overcome many of the limitations of percentile-VaR: quantile distribution theory (Section 2.4), conditional tail expectation (Section 2.5) and distortion functions (Chapter 3).

2.4 Quantile Distribution Theory

(Embrechts, Kluppelberg and Mikosch, 1997) Right-tail risk analysis has proved to be of considerable importance when comparing risks. Extremely large losses that occur with very small probability tend to be overlooked by many financial risk tools; however, there exists extensive literature on right-tail losses relating to insurance risk (Wang, 1998).

Percentile-VaR is an extreme value statistic (Bassi, Embrechts and Kafetzaki, 1997). As such, if it is to be used, it is important to understand the properties of this type of statistic in order to improve our understanding of our percentile-VaR results.

A definition of a p-percentile, x_p of the distribution $F(x)$ is:

$$x_p = F^{-1}(p) = inf\{x \in \mathcal{R}; F(x) \geq p\}. \tag{2.20}$$

Based on an independent identically distributed ($i.i.d.$) random sample of n data points, $X_1, ..., X_n$, the empirical distribution of the random variable X is defined as:

$$F_n(x) = \frac{\#\{i\ :\ 1 \leq i \leq n\ and\ X_i \leq x\}}{n}, x \in \mathcal{R}. \tag{2.21}$$

Defining the order statistics for this distribution as

$$X_{(1)} = \min(X_1, ..., X_n) \leq X_{(2)} \leq ... \leq X_{(n)} = \max(X_1, ..., X_n), \tag{2.22}$$

then x_p can be estimated by

$$\hat{x}_{p,n} = F_n^{-1}(p) = X_{(k)}, \quad \frac{k}{n} \leq p \leq \frac{k+1}{n}. \tag{2.23}$$

By the Central Limit Theorem, it is possible to show that

$$\hat{x}_{p,n} \quad \sim \quad AN\left(x_p, \frac{p(1-p)}{nf^2(x_p)}\right) \tag{2.24}$$

where AN stands for asymptotically normal. Using this, we can obtain approximate confidence intervals for the estimated percentile.

As well, if $X_1, ..., X_n$ are $i.i.d.$, the binomial model for an order statistic can be used to produce percentile confidence intervals,

$$Pr(X_{j,n} \leq x_p < X_{i,n}) = \sum_{r=i}^{j-1} \binom{n}{r} p^{n-r}(1-p)^r \text{ for } i < j. \tag{2.25}$$

The resulting confidence intervals can help to identify the accuracy of a percentile-VaR value.

2.5 Conditional Tail Expectation

One measure of right-tail risk, the conditional tail expectation, is similar to the mean excess loss (MEL) (Bassi, Embrechts and Kafetzaki, 1997; Bowers et al., 1997; Klugman, Panjer and Willmot, 1998),

$$e(x) \;=\; E[\, X - x \mid X > x \,] \;=\; E[\, X \mid X > x \,] - x, \tag{2.26}$$

which is the same as the Mean Residual Lifetime, or life expectancy, e_x, used by life actuaries. It is also the same as the expected loss given a loss occurs for the reinsurer for a stop loss contract with attachment point x.

The Conditional Tail Expectation (CTE), conditioned at one tail value x, is the expected loss taken over all losses in excess of x, where x is the lower bound of the tail region being considered, referred to as the tail boundary value. Mathematically,

$$CTE\,(x) \;=\; E[\,X \mid X \;>\; x\,] \;=\; e(x) \;+\; x. \tag{2.27}$$

The CTE evaluated at a specific tail boundary value does not hold much information on its own; it is simply a conditional mean. However, as a function of the tail boundary value, its shape can well describe the risk implied by the loss distribution. A variation of the CTE is to define the tail boundary by a percentile, V_α. This CTE_α measure is also referred to as tail-VaR (Artzner, 1999), and is defined as the expected value of the loss given that the loss falls in the upper $(1-\alpha)$ tail of the distribution. In defining CTE_α, it is important to identify the case when V_α falls in a probability mass, where

$$V_\alpha = V_{\alpha+\xi}, \quad \text{for some } \xi > 0.$$

In this case, we define the conditional tail expectation function as:

$$CTE_\alpha \;=\; \frac{(1-\beta')E[\,X \mid X \;>\; V_\alpha\,] + (\beta'-\alpha)V_\alpha}{1-\alpha}. \tag{2.28}$$

$$\text{where} \quad \beta' = \max\{\beta : V_\alpha = V_\beta\}. \tag{2.29}$$

Intuitively, for a given loss distribution, the $CTE_{.95}$ would be enough on average to cover a 1 in 20 event.

2.6 Two-Parameter Distributions

In order to illustrate the characteristics of the CTE, we apply the CTE to four two-parameter distributions. The Pareto distribution was chosen for its heavy-tail and finite number of moments. The second distribution chosen is the lognormal distribution, which has tail that is not as heavy as the Pareto, and has many applications in finance. It is often used for stochastic stock price models. The gamma distribution is chosen because it is a generalization of the exponential distribution which has a constant failure rate, and the gamma distribution has a moderate sized tail. Lastly the normal distribution is chosen for its simplicity and light tail, and it is used as a standard approximation tool when discussing confidence intervals and value-at-risk techniques.

To set parameters for these distributions we used two methods. In our first comparison we matched the means, so that each distribution had a mean of 3.0, and we matched the 95th percentile, so that each distribution had a 95th percentile of 10.415. The pdfs of these distributions are illustrated in Figure 2.3 and their ddfs are illustrated in Figure 2.4. In our second comparison, we set parameters for the distributions to match the first two moments, using a mean of 3.0 and a variance of 45.0. The pdfs of these uncensored distributions are illustrated in Figure 2.5 and their ddfs are illustrated in Figure 2.6.

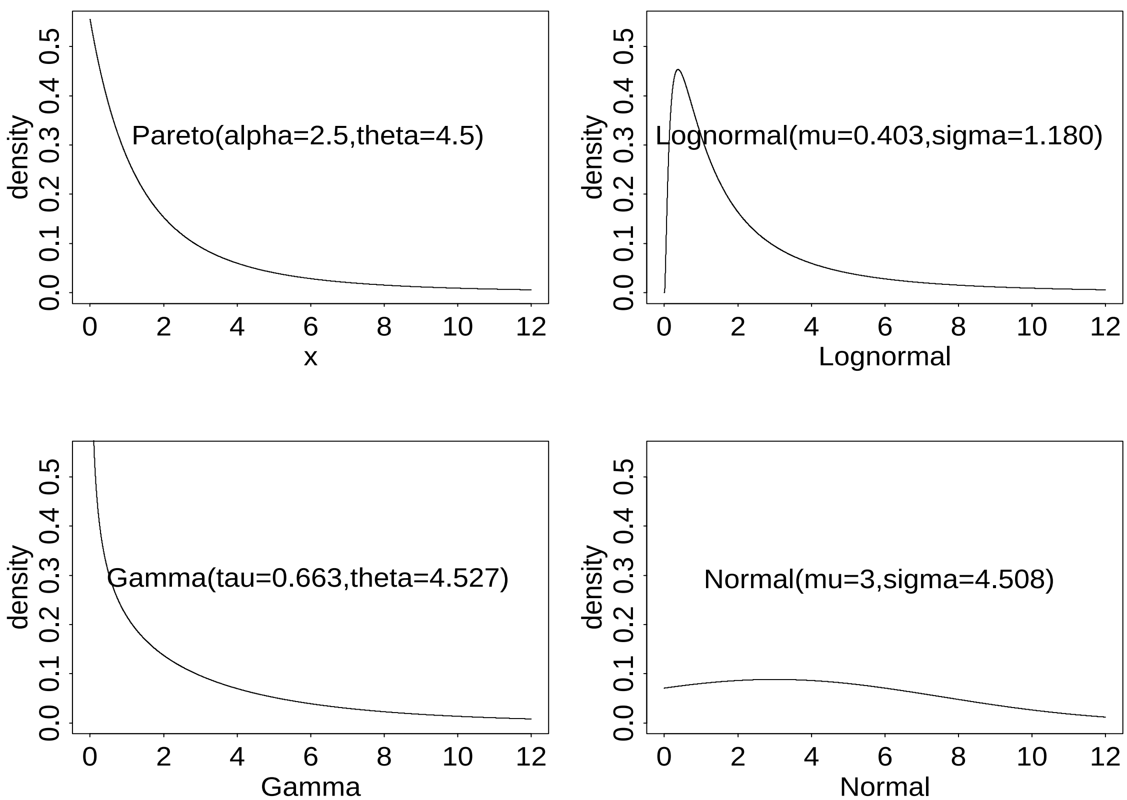

Figure 2.3: Two Parameter PDFs (Mean= 3 and 95%ile= 10.415)

For our first comparison, using the distributions with equated mean and 95th percentile, we illustrate how the CTE measures the risk for each loss distribution by plotting the CTE(x) against x, the truncation value in Figure 2.7. This figure illustrates the direct relation between the CTE and the tail of the distribution. The heavier tailed Pareto and lognormal distributions are more steeply sloped and have higher CTEs further out in the tail. The Pareto distribution has a constant mean excess lifetime, which is linearly related to the CTE (see Equation 2.27), thus

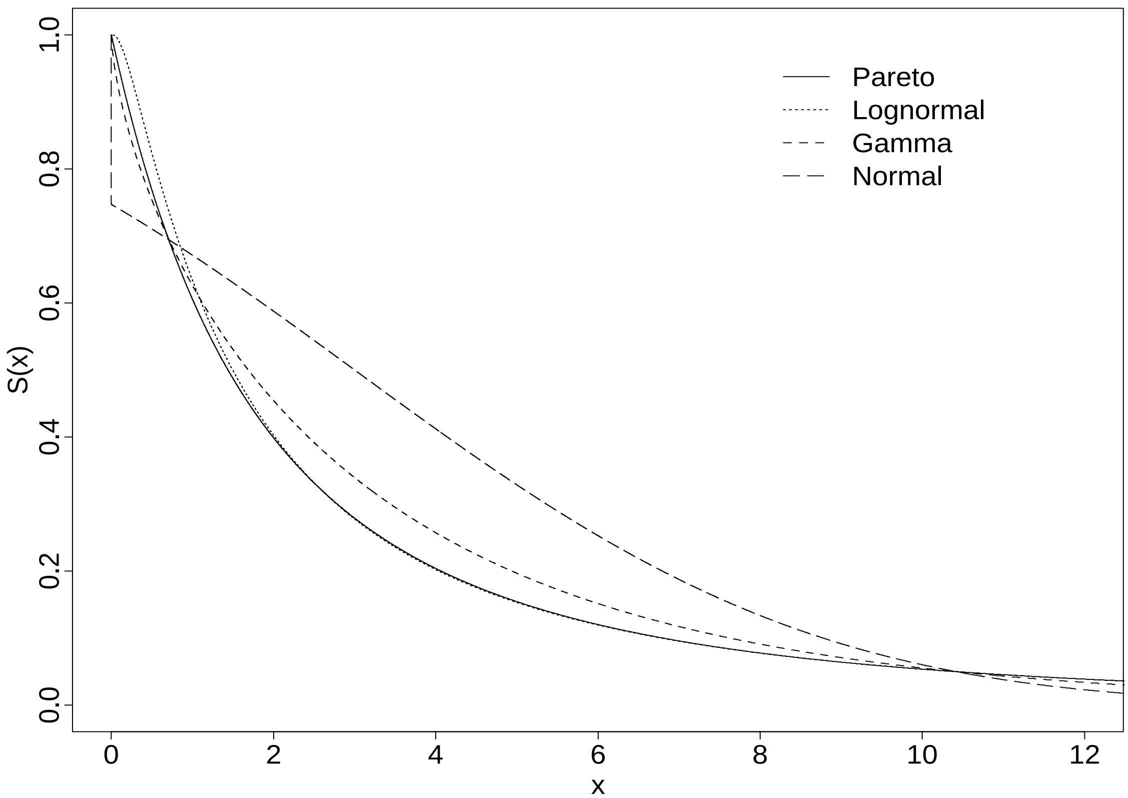

Figure 2.4: Two Parameter $S(x)$ (Mean= 3 and 95%ile= 10.415)

the CTE for the Pareto is a straight line. At the 95th percentile, $x = 10.41504$, there is a significant difference between the CTE values, where the heavier tailed distributions have significantly larger CTE values. From this graph, one can see the expected severity of the excess risk caused by a heavier tailed distribution.

Plotting the CTE against the truncation value compares the distributions based on the same truncation value, however for the heavy tailed distributions, this figure does not illustrate the limit of the CTE as the truncation value tends to infinity, or the maximum of the distribution. In order to obtain a more com-

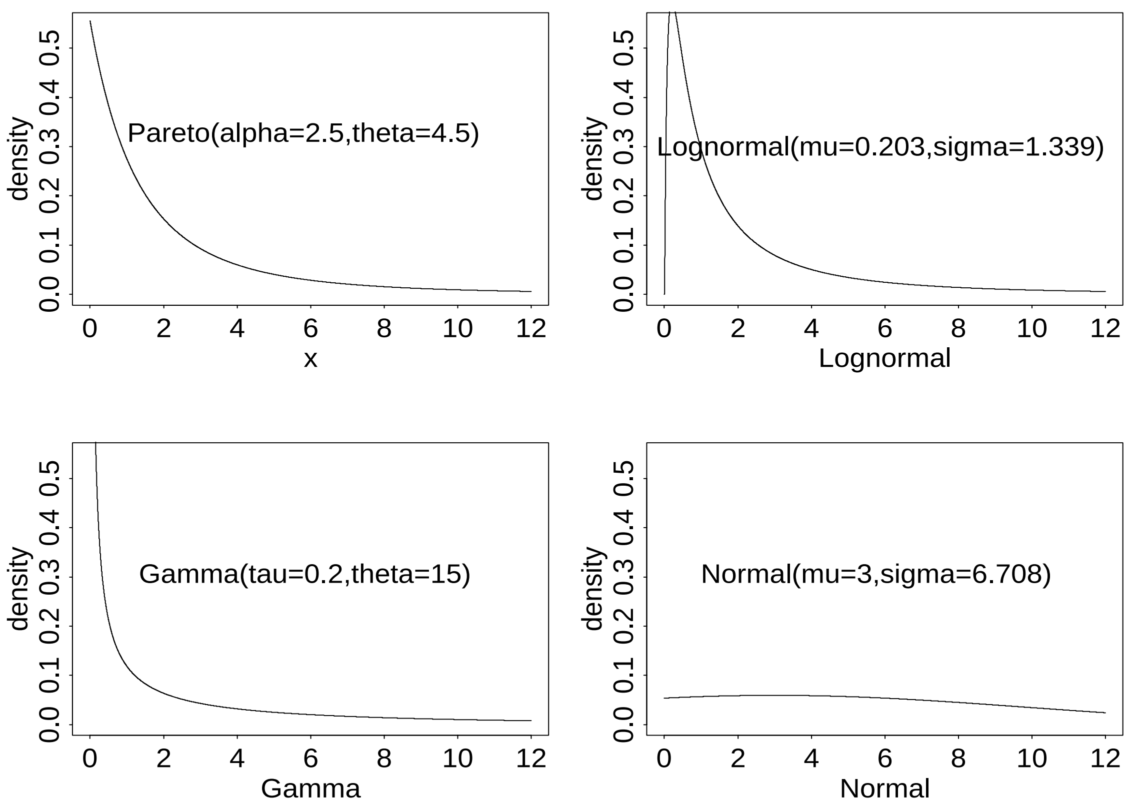

Figure 2.5: Two Parameter PDFs (Mean= 3 and Variance= 45)

prehensive view of the risk, Figure 2.8 illustrates the CTE over the full tail of the distribution by plotting the CTE(x) against the tail boundary percentiles of the distribution, $F(x)$. In this figure we see a more significant relation between the CTE and the heaviness of the tail of the distribution, especially far out in the tail. In both figures, the CTE at the 95th percentile, $V_{.95} = 10.41504$, can be compared directly.

When analyzing the empirical distribution of any portfolio, a common approach is to calculate the first few moments of the empirical distribution, and fit

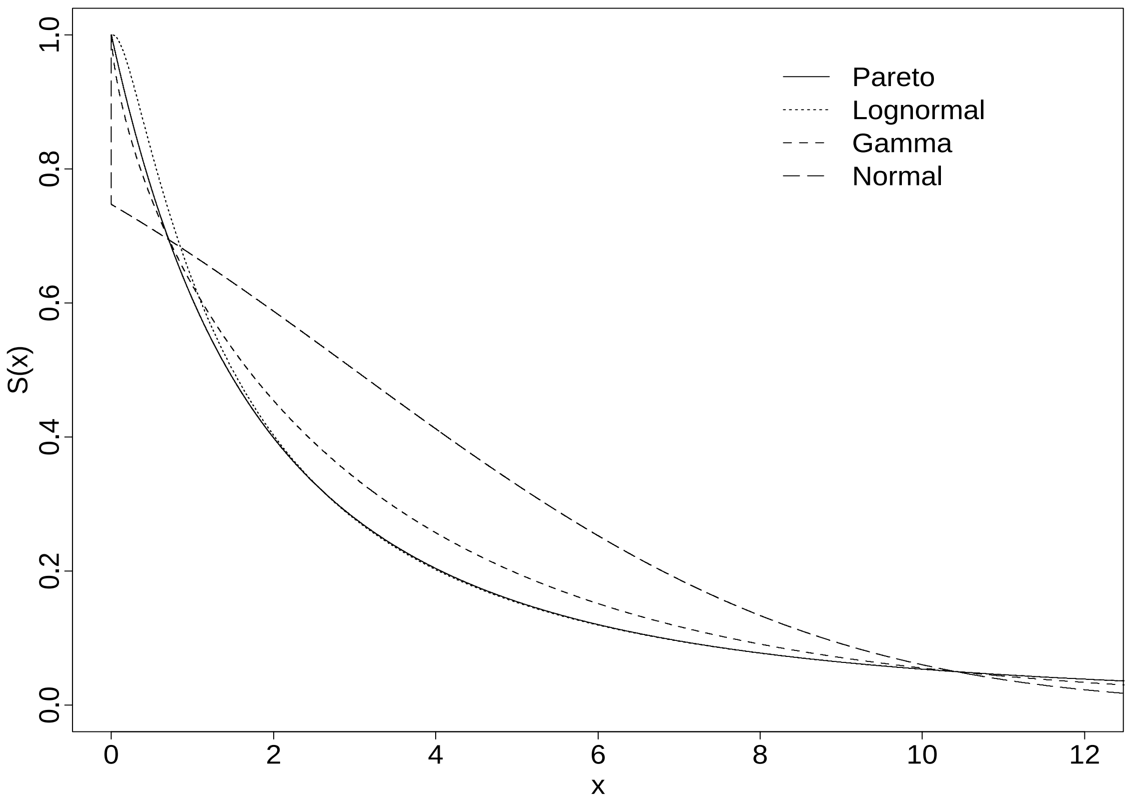

Figure 2.6: Two Parameter $S(x)$ (Mean= 3 and Variance= 45)

the empirical distribution to some known distributions by equating the moments. For a second comparison of the two-parameter distributions, we parameterize the distribution to have the same first and second moments about the mean. In Figure 2.5, the four probability density functions are illustrated, each distribution has a mean of 3 and a variance of 45.

Again, we illustrate how the CTE measures the risk in the distribution by plotting the CTE(x) against x, the truncation value, for each distribution in Figure 2.9. From Figures 2.7 and 2.9 other than the normal distribution, the risk

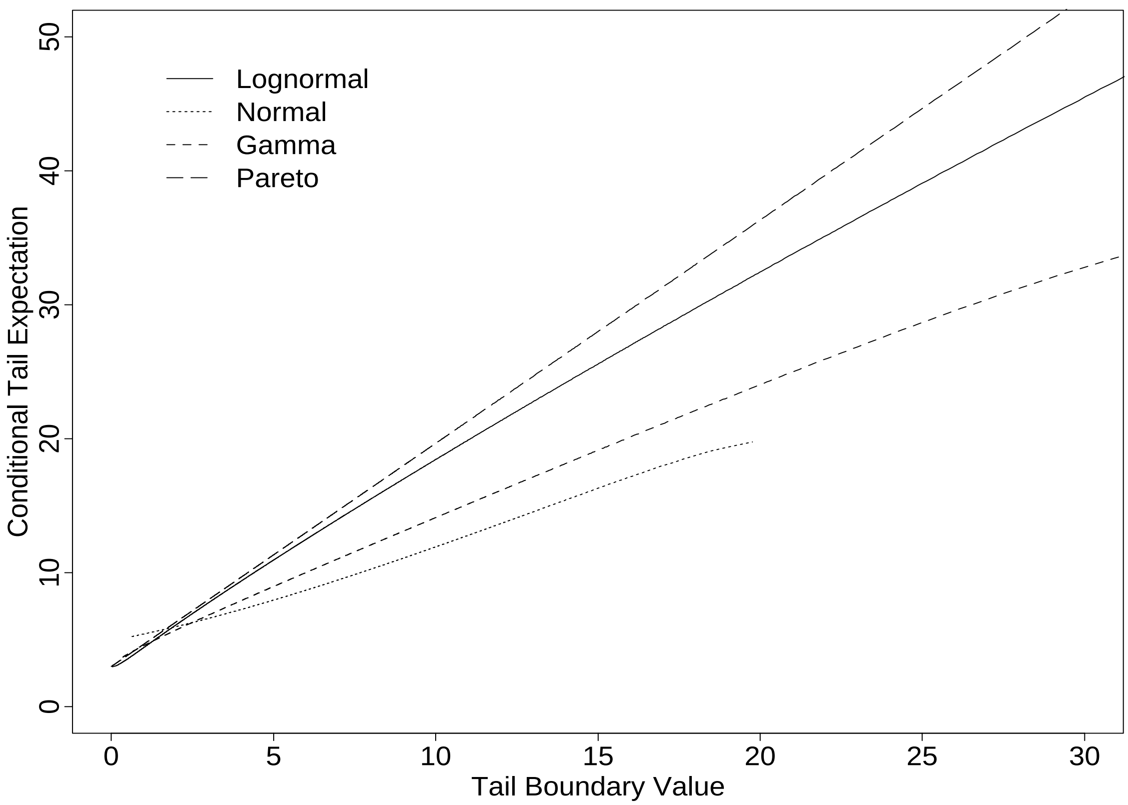

Figure 2.7: CTE vs. Tail Boundary Value (Mean= 3 and 95%ile= 10.415)

distributions with equated moments are more closely fitted than the distributions with equated 95th percentile. To see this, compare the range of the CTEs at a tail boundary value of 10, the lines in Figure 2.7 have a range from 14 to 20 and are diverging, whereas the lines in Figure 2.9 have a range from 18 to 22 and do not start to diverge until the tail boundary value is 12. This suggests that comparing the risk in a distributions based on a percentile is not as informative as using the moment approach. This is supported by the fact that a percentile is one number obtained from one point on a distribution, whereas a moment of the distribution is

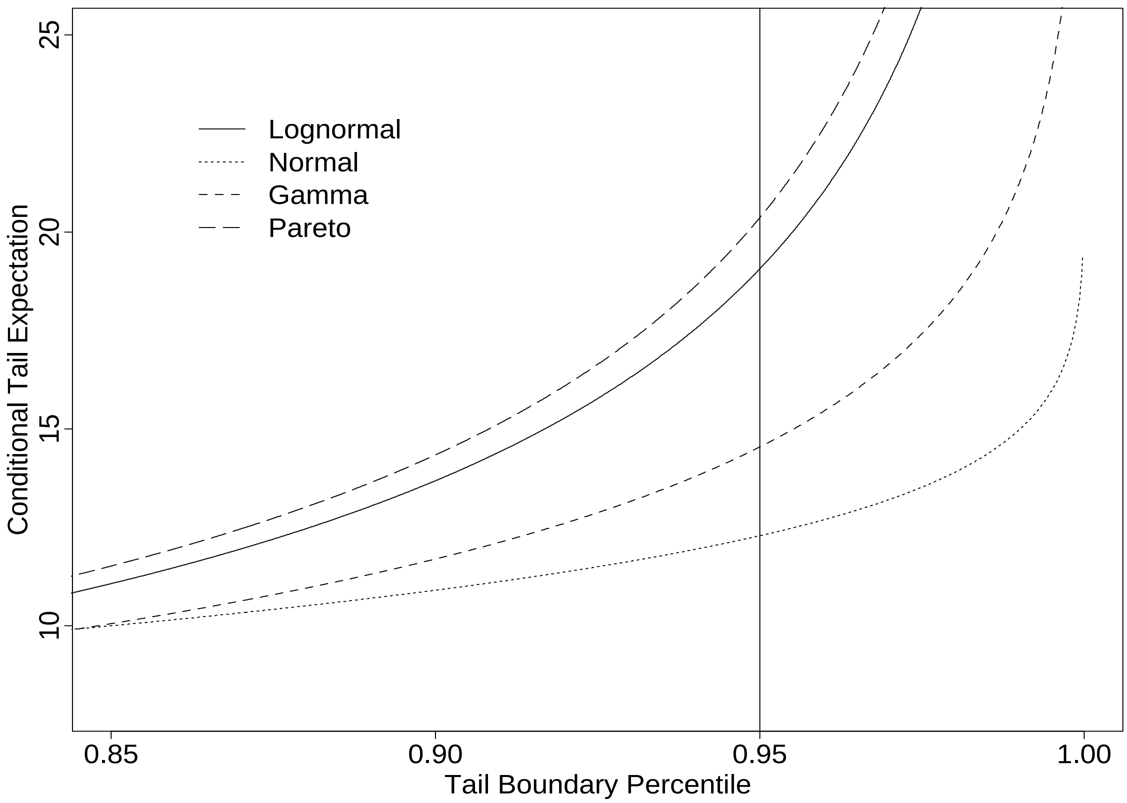

Figure 2.8: CTE vs. Tail Boundary Percentile (Mean= 3 and 95%ile= 10.415)

one number but the whole distribution is used to calculate it.

In Figure 2.10, the ordering of the CTEs in the far right tail is consistent with our intuition based on second order stochastic dominance. However, the CTE does not consistently order the risks independently of the percentile chosen; the gamma and the lognormal distributions seem to be more risky than the Pareto unless a percentile greater than 98% is used. In the extreme tail of the distributions, the Pareto distribution, which has the fattest tail, has the highest CTE values. In Chapter 4, the CTE is shown to rank distributions inconsistently with second order

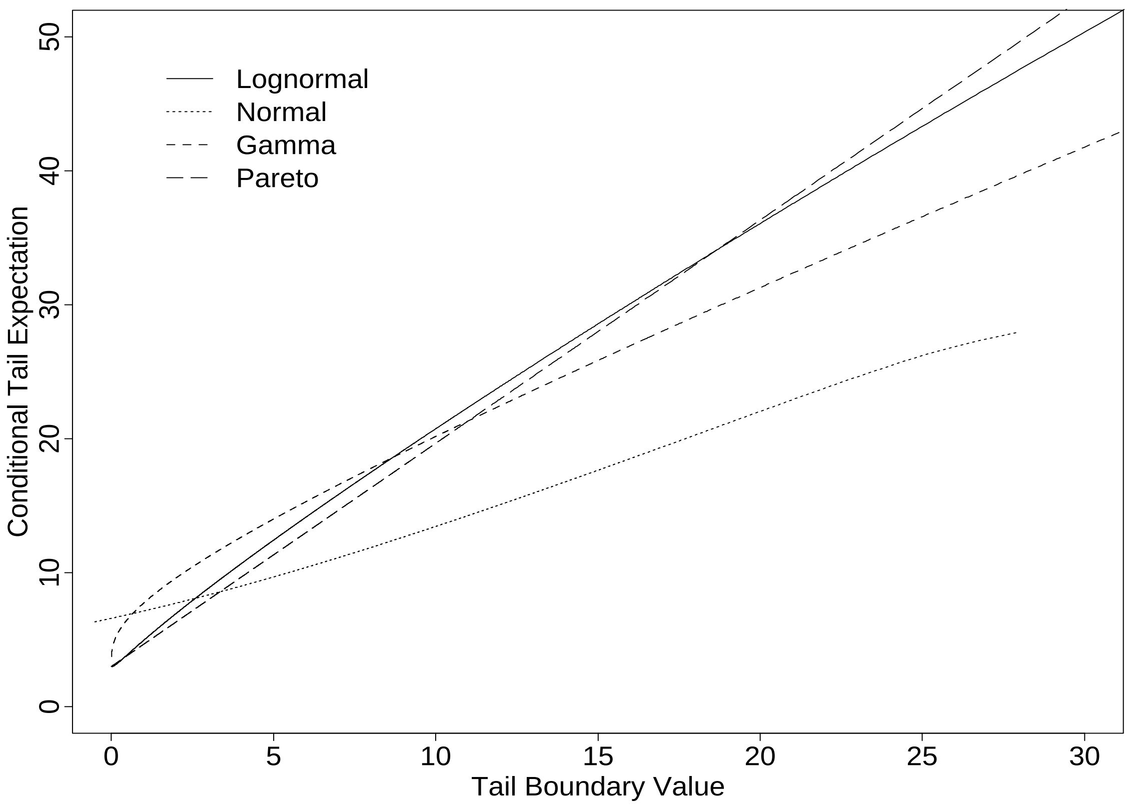

Figure 2.9: CTE vs. Tail Boundary Value (Mean= 3 and Variance= 45)

stochastic dominance. The extreme tail is illustrated by plotting the CTE against the tail boundary percentile in Figures 2.10 and 2.11. Comparing the figures, using tail boundary value and tail boundary percentile for this illustration, is not as transparent, since the 95th percentiles are not equal.

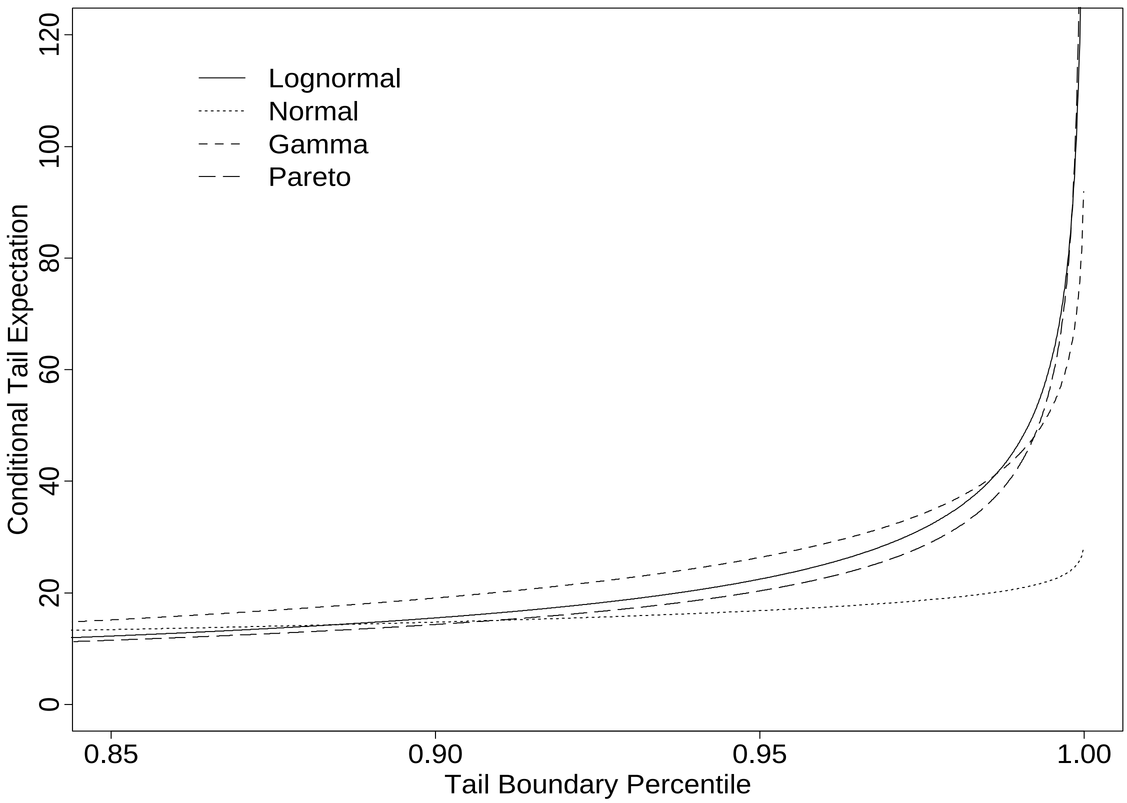

Figure 2.10: CTE vs. Tail Boundary Percentile (Mean= 3 and Variance= 45)

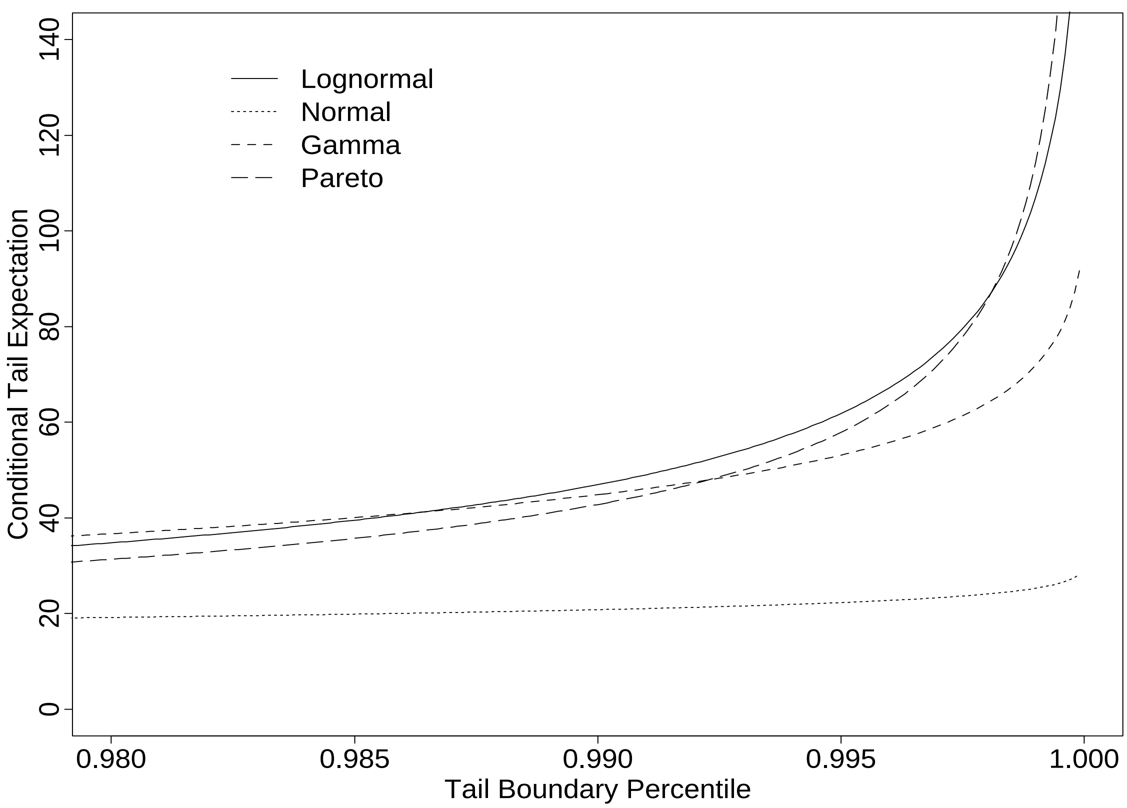

Figure 2.11: CTE vs. Extreme Tail Boundary Percentile (Mean= 3 and Variance= 45)

Chapter 3

Synthesis of Transformations

In this chapter we consider distorted risk measures, and show that VaR and the CTE risk measures are special cases. We specify the coherent set of distorted risk measures and discuss their properties. Two of these measures are considered in more depth and applied to the examples from the Chapter 2.

Definition 3.0.1 *A distortion function* $g : [0,1] \to [0,1]$ *is a non-decreasing function with* $g(0) = 0$ *and* $g(1) = 1$.

Function g is a concave distortion function if for all $a, b \in [0,1]$ such that $a < b$, and given any arbitrary $\epsilon \in (0,1)$,

$$g(a(1-\epsilon) + b\epsilon) \geq g(a)(1-\epsilon) + g(b)\epsilon. \tag{3.1}$$

A distorted probability μ (Wang and Young, 1998) is defined on a σ-algebra Ω as $\mu(A) = g[P(X \in A)]$, where $A \in \Omega$, g is a distortion function, and P is a probability measure on Ω.

Given that F_X is the distribution function of X, a non-negative random variable, and $S_X = 1 - F_X$ is the decumulative distribution function, then,

$$E_\mu[X] = \int_0^\infty g[P(X > x)]dx = \int_0^\infty g[S_X(x)]dx, \tag{3.2}$$

which defines the premium principle developed by Wang (1995a):

$$\rho_g(X) \equiv \int_0^\infty g[S_X(x)]dx = \int_0^\infty S_g^*(x)dx. \tag{3.3}$$

Since X is a non-negative random variable, $\rho_g(X) \equiv E_g[X]$. The distortion effectively changes the measure to allow for risk, sometimes called a risk adjusted measure.

3.1 Properties of Distorted Risk Measures

Theorem 3.1.1 *If g is a concave distortion function, and $S_X(x) = 1 - F_X(x)$, then the distorted risk measure, $\rho_g(X)$ is a coherent risk measure (Wang, 1996b).*

Proof: Given concave distortion function g and an arbitrary risk X, to prove $\rho_g(X)$ is a coherent risk measure, we show that $\rho_g(X)$ satisfies properties A1-A5.

For A1: Bounded above by maximal loss.

Since g is an increasing function and $g(0) = 0$ and $g(1) = 1$, then

$$g(S(x)) \leq 1 \quad for \quad x \leq \max(X), \tag{3.4}$$

$$\text{and} \quad g(S(x)) = 0 \quad for \quad x > \max(X). \tag{3.5}$$

$$\text{Thus,} \quad \rho_g(X) = \int_0^\infty g(S_X(x))dx \leq \int_0^{\max(X)} 1 \quad dx = \max(X) \quad \square \tag{3.6}$$

For A2: Bounded below by expected loss.

Since g is an increasing concave function of S(x),

$$\begin{aligned} g(S_X(x)) \quad &\geq \quad S_X(x) \quad \text{for all} \quad x \geq 0 && (3.7)\\ E[X] = \int_0^\infty S_X(x)\ dx \quad &\leq \quad \int_0^\infty g(S_X(x))\ dx = \rho_g(X) \quad \square && (3.8) \end{aligned}$$

For A3, A4: Scalar multiplicative and scalar additive.

For $a \geq 0$ and $b \geq 0$,

$$S_{aX+b}(u) = \begin{cases} 1 & \text{for} \quad 0 < u < b \\ S_X(\frac{u-b}{a}) & \text{for} \quad u \geq b \end{cases} \tag{3.9}$$

$$\rho_g(aX+b) = \int_0^b 1\ du + \int_b^\infty g(S_X(\frac{u-b}{a}))du \tag{3.10}$$

$$= b + a\int_0^\infty g(S_X(t))dt = a\rho_g(X) + b \quad \square \tag{3.11}$$

For A5: Subadditive. (Based on proof in Wang, 1995b)

First note that if g is increasing and concave, then for $0 < a < b$ and $x > 0$,

$$g(b+x) - g(a+x) \leq g(b) - g(a). \tag{3.12}$$

For any arbitrary increasing concave distortion function g, we define $\rho_g(X) = \int_0^\infty g(S_X(x))\ dx$. Using mathematical induction for every g and related ρ_g, we prove the result for arbitrary loss random variable V, and U a discrete loss random variable taking values in $\{0, ..., n\}$. By A4 the proof also holds for $U \in \{k, ..., n+k\}$ and by A3 for $U \in \{hk, ..., (n+k)h\}$, $h > 0$. Any random variable can be approximated arbitrarily closely by a discrete variable with small span h.

By mathematical induction:

(i) For $n = 0, U_0 = 0$ almost surely, and $\rho(U_0) = 0$, so for any V

$$\rho(V + U_0) = \rho(V) + 0. \tag{3.13}$$

(ii) For $n, U_n \in \{0, ..., n\}$, we assume that

$$\rho(U_n + V) \leq \rho(U_n) + \rho(V). \tag{3.14}$$

(iii) For $n+1$: Consider (U_{n+1}, V) with $U_{n+1} \in \{0, 1, ..., n+1\}$, and let (U^*, V) be distributed as $(U_{n+1}, V|U_{n+1} > 0)$. By (ii) and A4 the result holds for $U^* \in \{1, ..., n+1\}$. Thus

$$\rho(U^* + V) \leq \rho(U^*) + \rho(V). \tag{3.15}$$

With $\omega_0 = \Pr(U = 0)$ and $S_{V|0}(t) = \Pr(V > t|U = 0)$, we have for $t > 0$ that

$$\begin{aligned} S_U(t) &= (1-\omega_0) S_{U^*}(t), & (3.16) \\ S_V(t) &= \omega_0 S_{V|0}(t) + (1-\omega_0) S_V(t), & (3.17) \\ S_{U+V}(t) &= \omega_0 S_{V|0}(t) + (1-\omega_0) S_{U^*+V}(t). & (3.18) \end{aligned}$$

This yields (according to Equation 3.12) for $t > 0$,

$$\begin{aligned} g[S_{U+V}(t)] &- g[S_U(t)] - g[S_V(t)] \\ &= g[\omega_0 S_{V|0}(t) + (1-\omega_0) S_{U^*+V}(t)] & (3.19) \\ &\quad -g[(1-\omega_0) S_{U^*}(t)] - g[\omega_0 S_{V|0}(t) + (1-\omega_0) S_V(t)] \\ &\leq g[(1-\omega_0) S_{U^*+V}(t)] & (3.20) \\ &\quad -g[(1-\omega_0) S_{U^*}(t)] - g[(1-\omega_0) S_V(t)] & (3.21) \\ &= g(1-\omega_0) \left\{ \frac{g[(1-\omega_0) S_{U^*+V}(t)]}{g(1-\omega_0)} \right. & (3.22) \\ &\quad \left. - \frac{g[(1-\omega_0) S_{U^*}(t)]}{g(1-\omega_0)} - \frac{g[(1-\omega_0) S_V(t)]}{g(1-\omega_0)} \right\} \end{aligned}$$

Now define $h[S(t)] = \frac{g[(1-\omega_0)S(t)]}{g(1-\omega_0)}$, a new increasing concave distortion function, and since $g(1-\omega_0)$ is a positive constant, integration over t on both sides implies that the right hand side is less than zero by Equation 3.15, and this yields

$$\rho(U+V) \leq \rho(U) + \rho(V). \quad \square \tag{3.23}$$

It is interesting to note that, similar to risk averse utility functions, coherent distortion functions are increasing and concave, but distortion functions modify the probability and keep the wealth function unchanged, whereas utility functions modify the wealth and keep the probability unchanged.

Theorem 3.1.2 *All distorted risk measures with increasing, concave distortion functions are layer additive (B4) and have an increasing relative risk load (B6) (Wang, 1996b).*

Proof: B4: $X = I_{(x_0,x_1]}(X) + I_{(x_1,x_2]}(X) + ...$, where $x_0 = 0$ and

$$I_{(x_i,x_{i+1}]}(X) = \begin{cases} 0 & X < x_i \\ X - x_i & x_i \leq X < x_{i+1} \\ x_{i+1} - X & x_{i+1} \leq X \end{cases} \tag{3.24}$$

For layer $(x_i, x_{i+1}]$,

$$\rho_g(I_{(x_i,x_{i+1}]}) = \int_{x_i}^{x_{i+1}} g(S_X(t))dt \tag{3.25}$$

$$\text{so that,} \quad \sum_{i=0}^{\infty} \rho_g(I_{(x_i,x_{i+1}]}) = \int_0^{\infty} g(S_X(t))dt = \rho_g(X). \tag{3.26}$$

For B6:

$$\text{Let} \quad \phi(x) = \lim_{h\to 0} \frac{\rho_g(I_{(x,x+h]})}{E[I_{(x,x+h]}]} \tag{3.27}$$

be the relative risk load for the infinitesimal layer $(x, x+dt]$. Then

$$\phi(x) = \frac{\rho_g(I_{(x,x+dt]})}{E[I_{(x,x+dt]}]} = \frac{g(S_X(x))}{S_X(x)} \tag{3.28}$$

and substituting $u = S_X(x)$, which is a decreasing function of x, then

$$\phi(x) = \frac{g(u) - g(0)}{u - 0} \quad \text{as} \quad g(0) = 0 \tag{3.29}$$

and since g is increasing concave, $\phi(x)$ is a decreasing function of $S_X(x)$, and thus an increasing function of t. $\square$

Also note that

$$\frac{g(S_X(x))}{S_X(x)} \begin{cases} \geq 1 & \text{if } g \text{ is concave} \\ > 1 & \text{if } g \text{ is strictly concave} \\ = \text{constant} & \text{if } g \text{ is linear} \end{cases} \tag{3.30}$$

Theorem 3.1.3 *All distorted risk measures with increasing, concave distortion functions preserve FSD(C1).*

Proof: C1. Let $X \prec_{FSD} Y$. Then $S_X(t) \leq S_Y(t)$ for all $t \geq 0$. Since g is an increasing, concave distortion function, then

$$g(S_X(t)) \quad \leq \quad g(S_Y(t)) \tag{3.31}$$

$$\text{and} \quad \int_0^\infty g(S_X(t))dt \quad \leq \quad \int_0^\infty g(S_Y(t))dt \tag{3.32}$$

$$\text{which implies,} \quad \rho_g(X) \quad \leq \quad \rho_g(Y). \tag{3.33}$$

Theorem 3.1.4 *All distorted risk measures with increasing, strictly concave distortion functions preserve ordering of dangerousness(C2) and SSD(C3).*

Proof: C2, C3. Due to Proposition 1.4.1 (page 15) we only have to prove that the increasing, strictly concave distortion risk measures preserve SSD with the once crossing rule. (based on proof from Wang (1996c))

Let $E[X] \leq E[Y]$, $X \prec_{SSD} Y$ and let t_0 be the once crossing point, so that

$$\begin{aligned} S_X(t) &\geq S_Y(t) \quad \text{for } t < t_0 \\ S_X(t) &\leq S_Y(t) \quad \text{for } t \geq t_0 \end{aligned} \tag{3.34}$$

and since $X \prec_{SSD} Y$ either

$$S_X(t) \quad < \quad S_Y(t) \;\; \text{for some} \;\; t > t_0 \tag{3.35}$$

$$\text{and/or } S_X(t) \quad > \quad S_Y(t) \;\; \text{for some} \;\; t < t_0 \tag{3.36}$$

Next, construct a new ddf,

$$S_Z(t) = \max\{S_X(t),\ S_Y(t)\} = \begin{cases} S_X(t) & t < t_0 \\ S_Y(t) & t \geq t_0 \end{cases} \tag{3.37}$$

Since we have already shown that the distorted risk measures have increasing relative risk loadings at upper layers,

$$\rho_g(Z) - \rho_g(X) \geq \frac{g[S_X(t_0)]}{S_X(t_0)} \int_{t_0}^{\infty} [S_Y(t) - S_X(t)]dt \tag{3.38}$$

and

$$\rho_g(Z) - \rho_g(Y) \leq \frac{g[S_X(t_0)]}{S_X(t_0)} \int_{0}^{t_0} [S_X(t) - S_Y(t)]dt \tag{3.39}$$

with at least one of the above inequalities being a strict inequality.

Subtracting the last two equations, we obtain

$$\rho_g(Y) - \rho_g(X) > \frac{g[S_X(t_0)]}{S_X(t_0)} \int_0^\infty [S_Y(t) - S_X(t)]dt \geq 0 \tag{3.40}$$

$$\text{Thus,} \quad \rho_g(Y) > \rho_g(X). \quad \square \tag{3.41}$$

Corollary 3.1.1 *All piece-wise linear distortion functions are coherent but do not preserve SSD(C3).*

Proof: The proof for the corollary parallels the proof of Theorem 3.1.4, however over any linear portion of the distortion function $\frac{g[S_X(t)]}{S_X(t)} = M$, a constant, the slope of the linear portion.

For any risk X we can construct a risk Y such that $E[X] = E[Y]$ and $X \prec_{SSD} Y$, and where t_0, the once crossing point, is such that $S_X(t_0) = S_Y(t_0) = b$ and $g(b)$ lies on one linear portion of the distortion function. Also suppose that the linear portion containing $g(b)$ covers the range from $g(a)$ to $g(c)$ where

$a < b < c$. Then

$$S_Y(t) \begin{cases} = S_X(t) & \text{for } t \geq t_a \\ = a & \text{for } t = t_a \\ \leq S_X(t) & \text{for } t_a \geq t \geq t_b \\ = b & \text{for } t = t_b \\ \geq S_X(t) & \text{for } t_b \geq t \geq t_c \\ = S_X(t) & \text{for } t \leq t_c \end{cases} \tag{3.42}$$

which implies that

$$\int_0^{t_c} [S_Y(t) - S_X(t)]dt = 0 \tag{3.43}$$

and

$$\int_{t_a}^{\infty} [S_Y(t) - S_X(t)]dt = 0. \tag{3.44}$$

Constructing the same equations as in (3.38), (3.39), (3.40), we obtain

$$\rho_g(Z) - \rho_g(X) = \frac{g[S_X(t_0)]}{S_X(t_0)} \int_{t_b}^{t_a} [S_Y(t) - S_X(t)]dt \tag{3.45}$$

and

$$\rho_g(Z) - \rho_g(Y) = \frac{g[S_X(t_0)]}{S_X(t_0)} \int_{t_c}^{t_b} [S_X(t) - S_Y(t)]dt \tag{3.46}$$

which gives,

$$\rho_g(Y) = \rho_g(X). \tag{3.47}$$

That is, we construct a risk such that $X \prec_{SSD} Y$ but $\rho_g(Y) = \rho_g(X)$. □

Thus increasing concave distortion functions are coherent and increasing strictly concave distortion functions are coherent and preserve second order stochastic dominance.

3.2 Special Distortions

Using distorted probabilities, it is possible to derive a distortion function that reproduces the CTE risk measure using Wang's premium principle (see Equation 3.3 on page 50). Define the distortion function as:

$$g_c(t) = \begin{cases} 1 & if \quad 1-\alpha < t \leq 1, \\ \frac{t}{1-\alpha} & if \quad 0 < t < 1-\alpha. \end{cases} \tag{3.48}$$

Wang's distorted risk measure will replicate the CTE_α risk measure where q_α is the tail boundary value. Since this distortion function is increasing and concave (see Figure 3.1), and the portion of the loss distribution that we are concerned with is the positive losses, the premium principle using this distortion function is a member of the family of premium principles with concave distortion functions and thus satisfies the properties of a coherent risk measure (see Definition 1.4.1). However, since the CTE distortion function is piecewise linear and not strictly

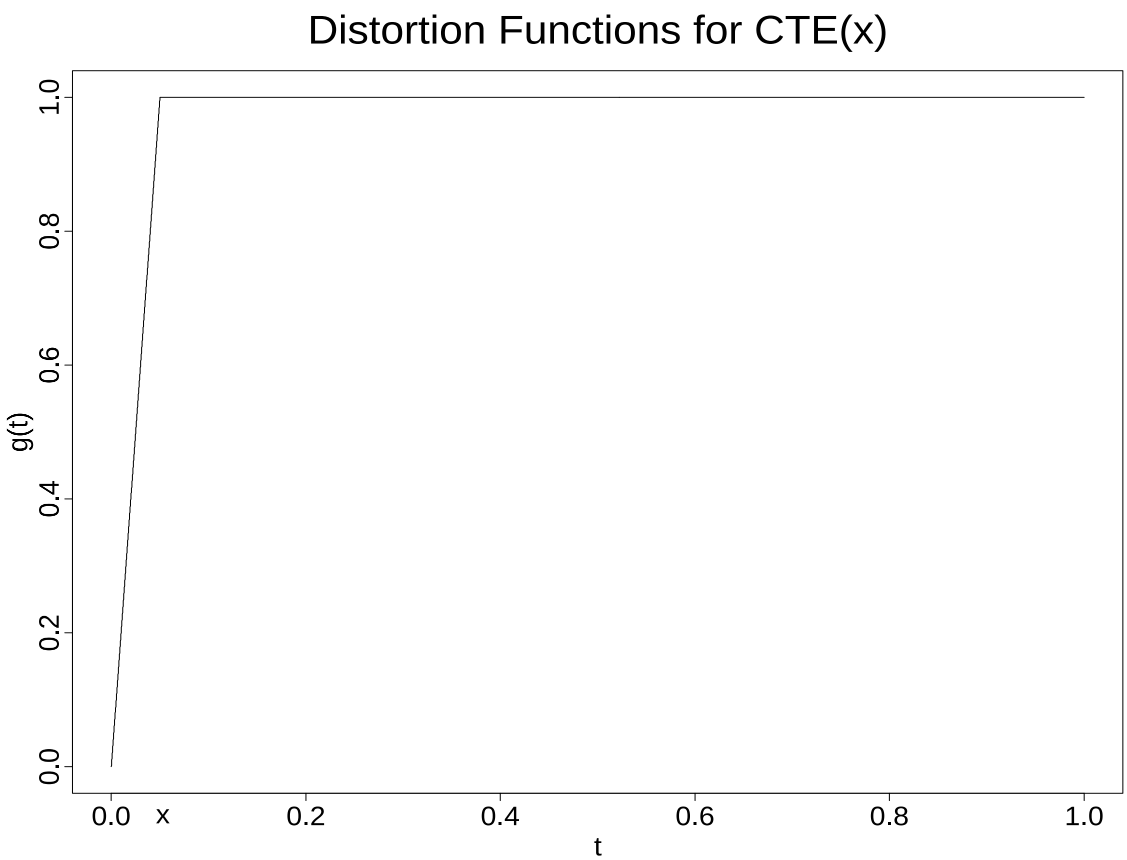

Figure 3.1: Distortion Function for Conditional Tail Expectation Function

concave it need not preserve second order stochastic dominance. To illustrate this, consider the expected value function which is a special case of the CTE (CTE_0).

Using distorted probabilities, it is also possible to define a distortion that will produce the percentile-VaR, V_α, as the risk measure. Define the distortion function as

$$g_V(t) = \begin{cases} 1 \quad if \quad 1-\alpha < t \le 1, \\ 0 \quad if \quad 0 < t < 1-\alpha. \end{cases} \tag{3.49}$$

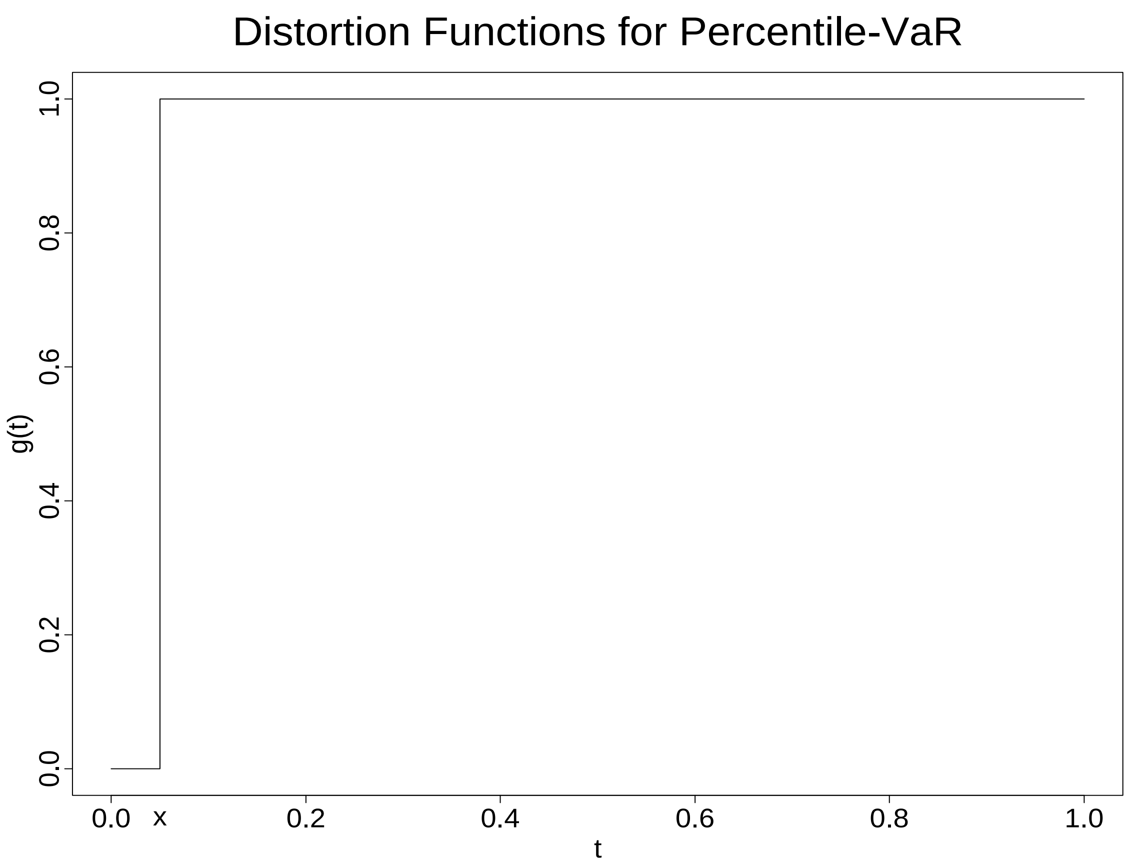

Figure 3.2: Distortion Function for Percentile-VaR

In this case, the distortion function is a step function(see Figure 3.2), with a discontinuity at $1-\alpha$. Thus, this distortion function is not concave, and the premium principle is not coherent. As we have already shown, VaR can be super-additive, illustrating that VaR is not a coherent risk measure.

The set of all concave distortion functions defines a large class of premium principles. There are some special cases of distortion functions that have intuitive explanations. These are described below:

Piecewise Linear Distortions: The CTE is an example of a piecewise linear distortion. The CTE can be interpreted as the mean loss given that the loss is greater than a specified value (discussed above). Piecewise linear distortion divides the risk based on the linear segments, and multiplies the probability of each subset by the slope of the linear segment. This assigns an outcome more "probability" where the distortion is steep, and less when it is flat.

The dual-power distortion: (Wang, 1996b) The function

$$g_d(t) = 1 - (1-t)^{\kappa}, \quad \kappa \geq 1 \tag{3.50}$$

gives the dual-power risk measure,

$$\rho_d(X) \;=\; \int_0^{\infty} 1 - [1 - S_X(x)]^{\kappa} dx, \tag{3.51}$$

which can be interpreted as the expected value of the maximum of κ observations. Using extreme value theory from Section 2.4, if $Y_1, Y_2, ..., Y_\kappa$ is a set of κ *i.i.d.* random variables with corresponding order statistics $Y_{(1)}, Y_{(2)}, ..., Y_{(\kappa)}$, then, $\rho_d(X)$ is equivalent to

$$E[Y_{(\kappa)}] \approx F^{-1}\left(\frac{\kappa}{\kappa+1}\right). \tag{3.52}$$

To obtain an approximation for the 95th or 99th percentile using the dual-power risk measure, one could use $\kappa = 19$ and $\kappa = 99$ respectively.

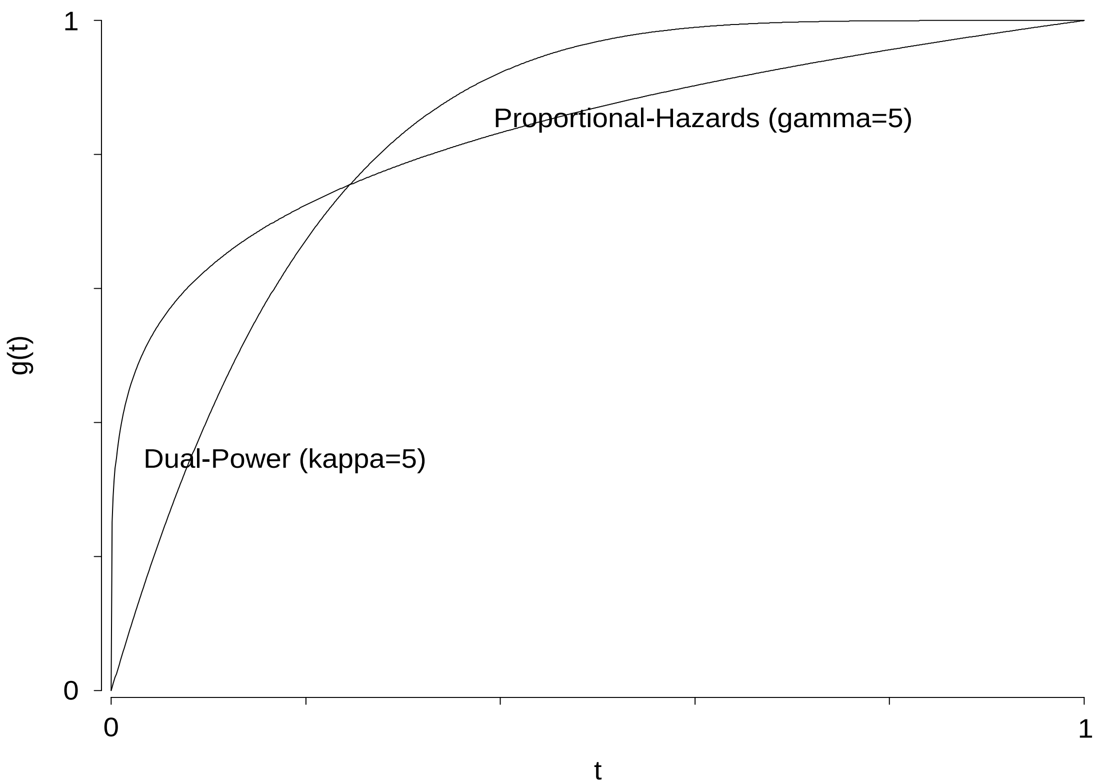

Figure 3.3: Proportional-Hazards and Dual-Power Distortion Functions

The proportional hazards (PH) distortion: (Wang, 1996a) The function

$$g(t) = t^{\frac{1}{\gamma}}, \quad \gamma \geq 1. \tag{3.53}$$

gives the PH-distortion risk measure,

$$\rho_g(X) \;=\; \int_0^\infty S_X(x)^{1/\gamma} dx, \quad \gamma > 1, \tag{3.54}$$

which can be interpreted as a risk-adjusted risk measure, where γ is the risk-averse index. Wang (1996a) shows how this resembles the risk neutral valuation method in option pricing theory.

Of these risk measures only the PH-distortion principle has $g\prime(0) = \infty$ where $g\prime(0)$ is the upper bound for the relative loading at upper layers. Wang (1996a) shows that in comparing a two-point risk and a Pareto risk with the same mean, only the PH-distortion calculates the risk measure for the Pareto risk to be more than the two point risk. Thus when $g\prime(0)$ is finite, the relative risk loading does not increase fast enough at the upper layers, as the relative risk loading is limited by a function of $g\prime(0)$.

Wang, Young and Panjer (1997) consider premium functionals. They use a similar set of axioms to A1-A5, and add a sixth axiom called the Reduction of Compound Bernoulli Risks which states:

> Let $X = IY$ be a compound Bernoulli risk, where the Bernoulli frequency random variable I is independent of the loss severity random variable $Y = X|X > 0$, and let ρ be the distortion function. Then the market prices for risks $X = IY$ and $I\rho(Y)$ must be equal.

An equivalent condition for this property is that $g(wq) = g(w)g(q)$ (Wang, Young and Panjer, 1997), or that g is multiplicative (property B2). These six axioms result in defining the market premium functional ρ uniquely as the proportional hazard distortion risk measure.

The proportional hazard distortion functions are a special subclass of coherent distortion functions that follow a larger set of properties and have proven useful for insurance premium calculations. The PH-distortion applied to the survivor function $S(x)$ is also known as the Lehmann family of survivor functions and is related to proportional hazards models in statistics (Barlow and Proschan, 1965). The application of the PH-distortions to premium principles has been considered extensively in Wang (1996a) and Wang (1996c).

The subclass of premium principles that use PH-distortion functions preserves all the same properties as the concave increasing distortion functions, however it also satisfies the multiplicativity property(B2), which is equivalent to the compound Bernoulli property from above and has a derivative of $+\infty$ at zero.

3.3 Illustrations

This section illustrates how the PH-distortion, dual-power distortion and the CTE improve upon the percentile-VaR risk measure.

Often insurance data, or loss data, is recorded without the related data on gains, or the probability of a gain. Example 3.3.1 identifies why the censored distribution is used, and the problems that would result if the full (uncensored) distribution or the truncated distribution were used. The next three Examples (3.3.2, 3.3.3 and 3.3.4) were used in section 2.2, to identify situations where percentile-VaR did not adequately compare risks. The CTE, PH-distortion and dual-power distortion risk measures are applied to these examples, and the results are discussed.

Example 3.3.1 Let X, Y be discrete loss random variables defined by,

$$f(X=x)=\begin{cases} 0.45, & x=-10 \\ 0.32, & x=-5 \\ 0.18, & x=0 \\ 0.04, & x=5 \\ 0.01, & x=10 \end{cases} \tag{3.55}$$

and,

$$f(Y=y)=\begin{cases} 0.71, & y=-10 \\ 0.04, & y=-5 \\ 0.0, & y=0 \\ 0.2, & y=5 \\ 0.05, & y=10. \end{cases} \tag{3.56}$$

A histogram for this example is shown in Figure 3.4. Using the full distribution, the expected values for these distributions are the same :

$$E[X]=E[Y]=-5.8. \tag{3.57}$$

Considering the full distributions, the probability of having a loss in Y is 5 times greater than having a loss in X, however the gains in Y are also greater than the gains in X. Thus the variance in Y is greater than the variance of X, but the expected returns are the same. For second order stochastic dominance, comparing

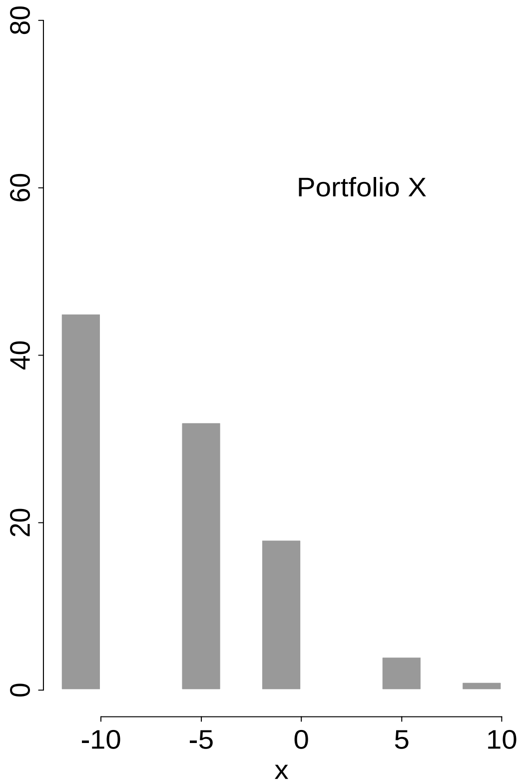

Figure 3.4: Histograms of Losses for Example 3.3.1

the mean and variance of two random variables is not enough. To show that X is

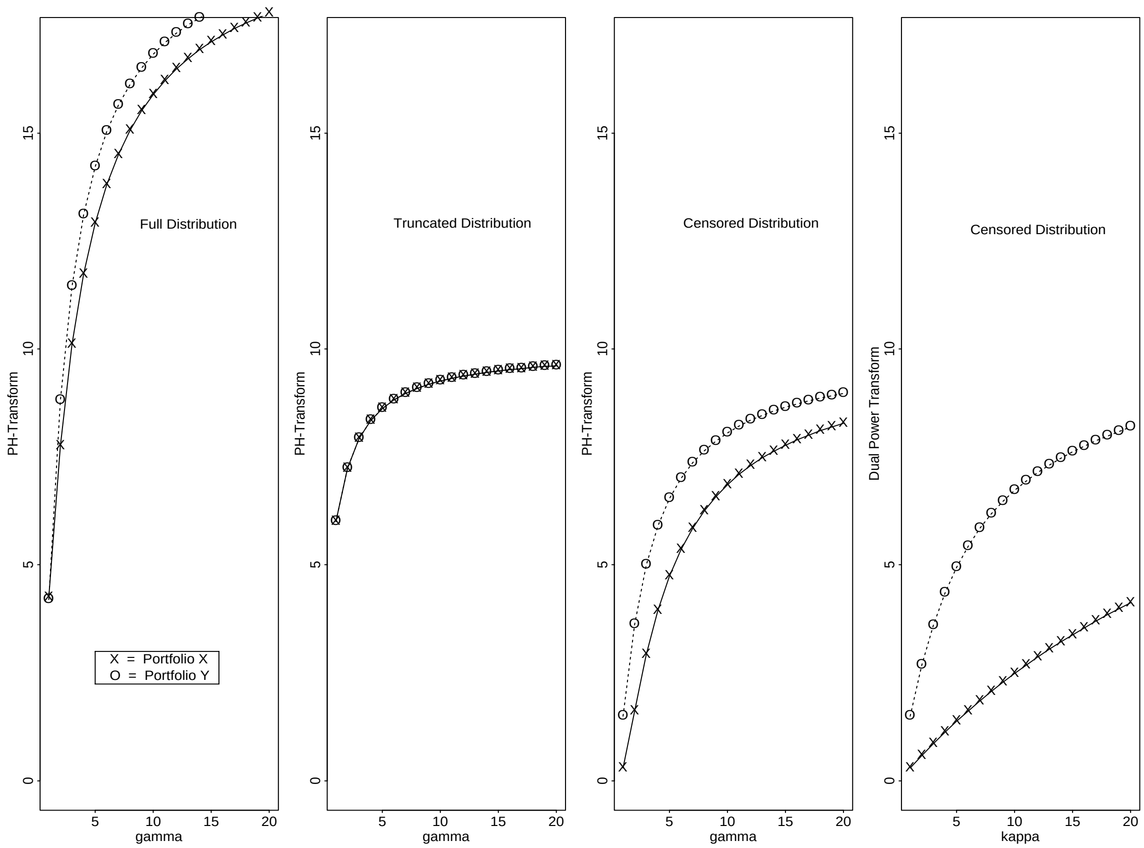

Figure 3.5: Capital Requirements for Example 3.3.1 using Proportional-Hazards and Dual-Power Distortions

less risky than Y, we compare ddfs.

$$\int_x^{\infty} S_X(t) - S_Y(t)dt = \begin{cases} 0 & \text{for} \;\; t \leq -10 \\ -1.3 + 0.26(-5 - x) & \text{for} \;\; -10 < x \leq -5 \\ -1.2 - 0.02(0 - x) & \text{for} \;\; -5 < x \leq 0 \\ -0.2 - 0.20(5 - x) & \text{for} \;\; 0 < x \leq 5 \\ 0 - 0.04(10 - x) & \text{for} \;\; 5 < x \leq 10 \\ 0 & \text{for} \;\; 10 < x \end{cases} \tag{3.58}$$

For any value of x, this integral is always negative, implying that risk Y dominates risk X using second order stochastic dominance (SSD). Based on the full distribution, any risk averse individual would prefer to invest in X. In terms of VaR, $V_{.9}(X) = V_{.9}(Y) = 0$, however $V_{.95}(X) = 0$, which is less than $V_{.95}(Y) = 10$.

Comparing these two random variables, the expected value of the loss, given that the loss is greater than zero, is the same for each risk. That is,

$$E[X|X > 0] = E[Y|Y > 0] = 6.0. \tag{3.59}$$

As well, if only the positive losses were recorded (in other words, if the loss distribution is truncated at 0) then the two truncated loss distributions would be the same:

$$f_T(X = t) = f_T(Y = t) = \begin{cases} .8, & t = 5 \\ .2, & t = 10. \end{cases} \tag{3.60}$$

Using the truncated distribution, there is no difference between these risks. However, since we know the true full distribution of these risks, it is clear that any risk measure must not be based solely on a truncated loss distribution.

If instead a censored distribution is used, where all gains are recorded as zero losses, the censored distributions would be:

$$f(X = x) = \begin{cases} .95, & x = 0 \\ .04, & x = 5 \\ .01, & x = 10 \end{cases} \tag{3.61}$$

$$f(Y=y)=\begin{cases} .75, & y=0 \\ .2, & y=5 \\ .05, & y=10. \end{cases} \tag{3.62}$$

and the expected, censored, losses would be:

$$E[\max(0,X)]=0.3 \qquad \text{and} \qquad E[\max(0,Y)]=1.5. \tag{3.63}$$

Figure 3.5 applies the PH-distortion risk measure to the full distribution of X and Y, the censored distribution, and the truncated distribution. The expected value definition of the distortion risk measure (see Equation 3.3) is used in the full distribution case, so that the integral of the distorted decumulative distribution function would be over the entire real number line. This does not make sense, as gains would add to the risk measure and the risk measure exceeds the maximum loss. Thus, the full distribution should not be used in this way. For premium calculations Wang (1999) uses a different distortion for gains, the dual-power distortion. However for the purpose of capital adequacy, the size of a possible gains is not relevant when a loss occurs. The second graph in Figure 3.5 applies the PH-distortion to the truncated distributions, and there is no difference between the risk measures for portfolio X and portfolio Y. Since the losses in each portfolio are considerably different, this implies that the probability that a gain occurs is of importance and that the truncated distribution should not be used. The third and fourth graphs apply the PH-distortion and the dual-power distortion risk measures to the censored data. These graphs order the risks in accordance with second order

Risk Measure	Parameter	Portfolio X	Portfolio Y
Dual-Power	19	3.98	8.09
	99	8.12	9.97
Proportional Hazards	4	3.95	5.90
	19	8.19	8.92

Table 3.1: Risk Measures for Example 3.3.1 using Censored Data

stochastic dominance, are bounded below by the expected loss, are bounded above by the maximal loss and follow the properties of coherence. Thus for the rest of this thesis, we consider only the censored distribution. Using the censored distribution, Table 3.1 compares portfolio X and portfolio Y using some specific risk measures.

Using the censored distribution, it seems that the dual-power distortion is more sensitive than the PH-distortion to changes in the probability of a loss occurring. However, for the dual-power the ordering of portfolios in terms of riskiness depends on the value of the parameter, as is shown in the next example.

Example 3.3.2 This example is a continuation of Example 2.2.1. The histograms and decumulative distribution functions for portfolio X and Y are shown in Figures 3.6 and 3.7 respectively. Based on $V_{.95}$, portfolio X is more risky. However, if we consider a person who is looking for financial security, a 5% probability of losing everything may seem more risky. Based on a 95% quantile the CTE for portfolio X is 1.246, and for portfolio Y is 2.689, which ranks the two portfolios in the same order as the PH-distortion risk measure when the parameter is greater than 2, or the dual-power risk measure when the parameter is over 8.

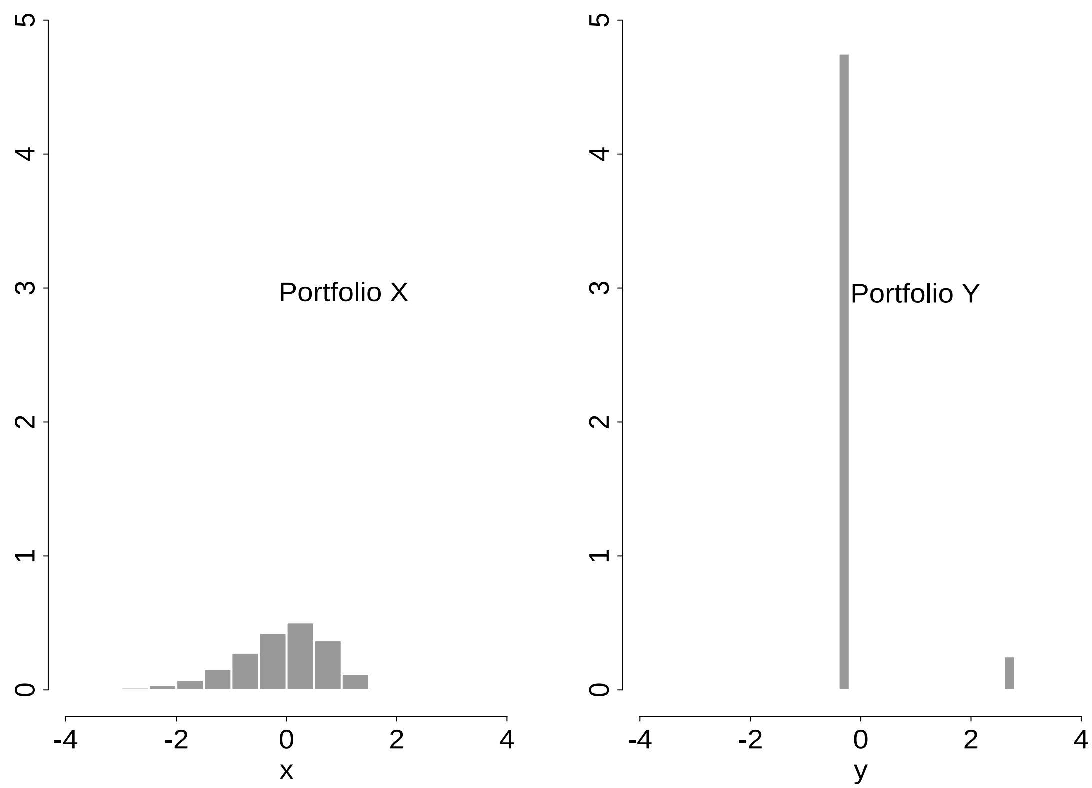

Figure 3.6: Histograms of Losses for Example 3.3.2

Example 3.3.2 shows that the ordering of risk using the PH-distortion or dual-power distortion may depend on the parameter chosen. The parameter relates to a measure of risk aversion, so it is important to use the same parameter when comparing portfolios; however, there is no strict rule for selecting a parameter value.

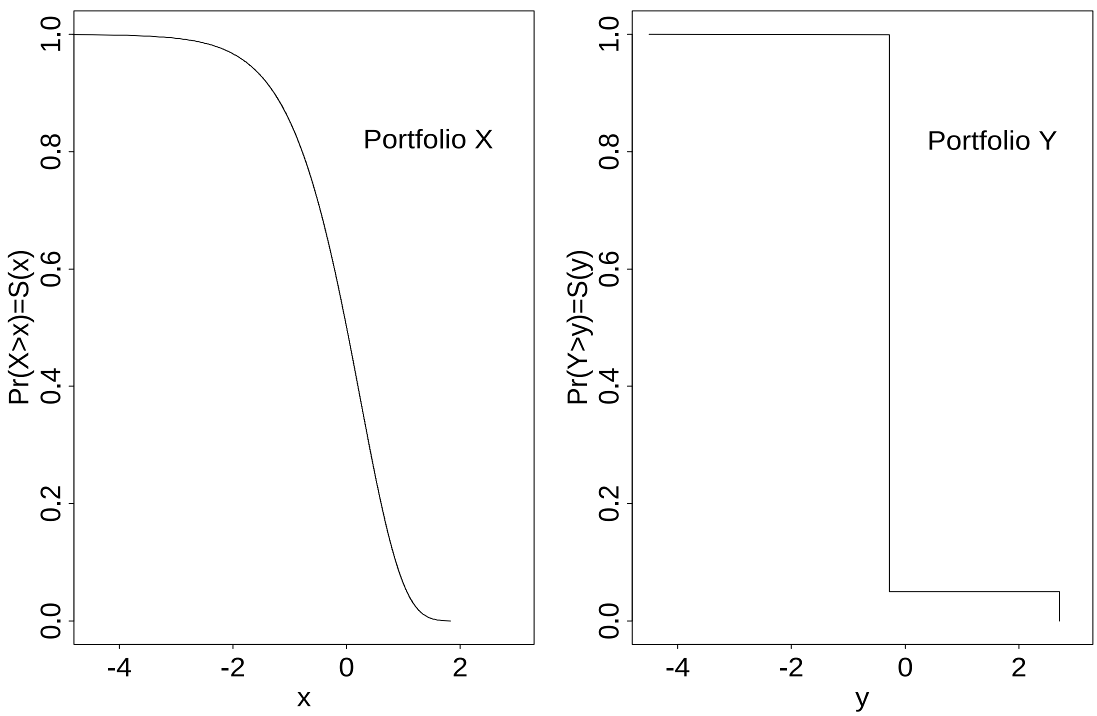

Figure 3.7: $S_X(x)$ and $S_Y(y)$ for Example 3.3.2

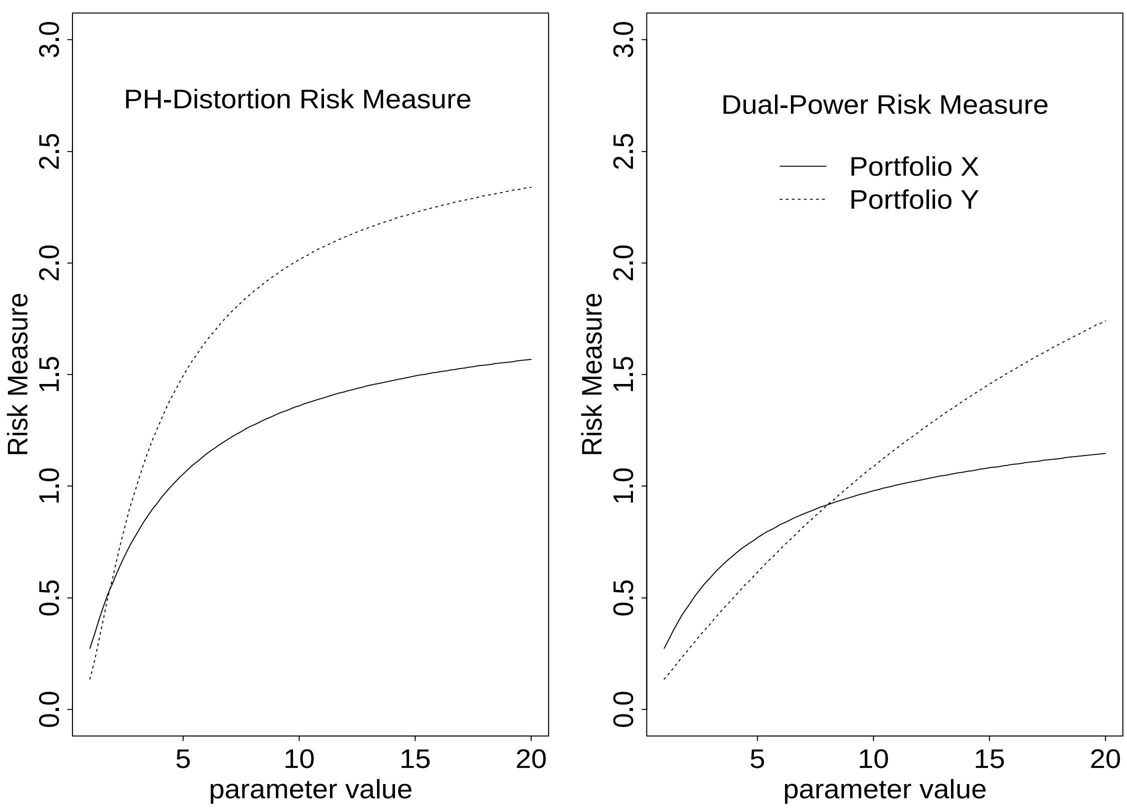

Figure 3.8: Capital Requirements for Example 3.3.2 using Proportional-Hazards and Dual-Power Distortions

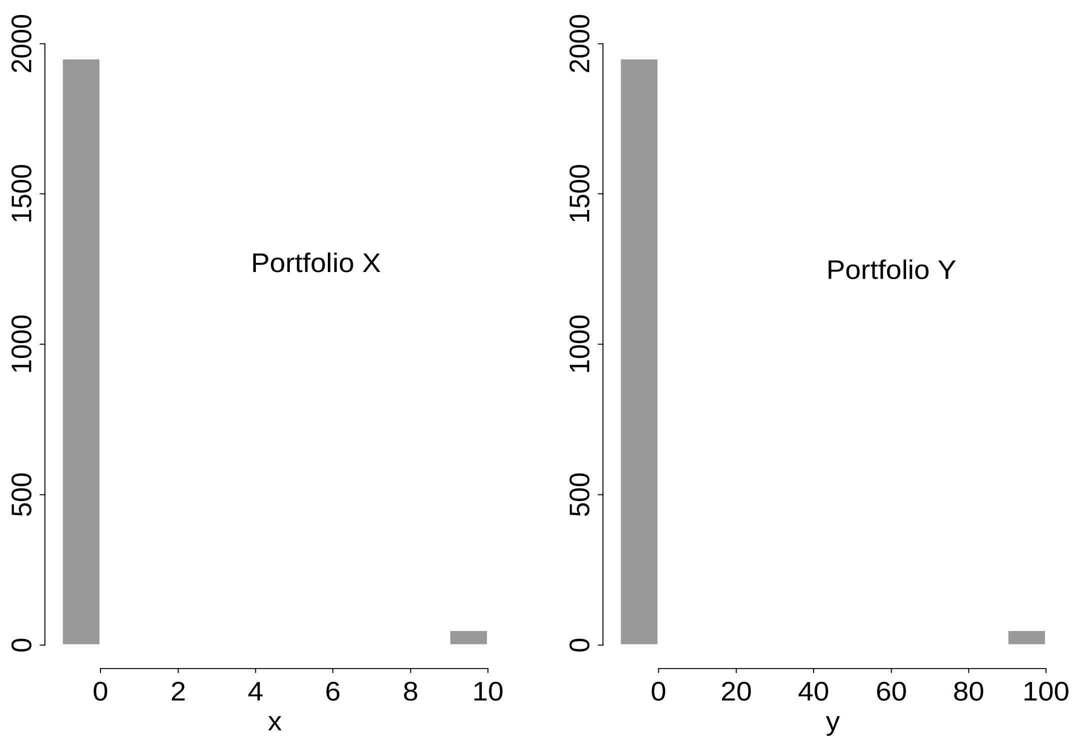

Figure 3.9: Histograms of Losses for Example 3.3.3

Example 3.3.3 This example is described in Example 2.2.2. The histograms for portfolio X and Y are shown in Figure 3.9. Both portfolios have a $V_{.95}$ of zero, which suggests that the two risks are equally risky, even though the potential loss for portfolio Y is ten times that of portfolio X.

Based on a 95% quantile, the CTE for portfolio X is 4.950, and the CTE for portfolio Y is 49.504. The PH-distortion and the dual-power distortion risk measures also evaluate portfolio X as the less risky portfolio, independent of the parameters see Figure 3.10.

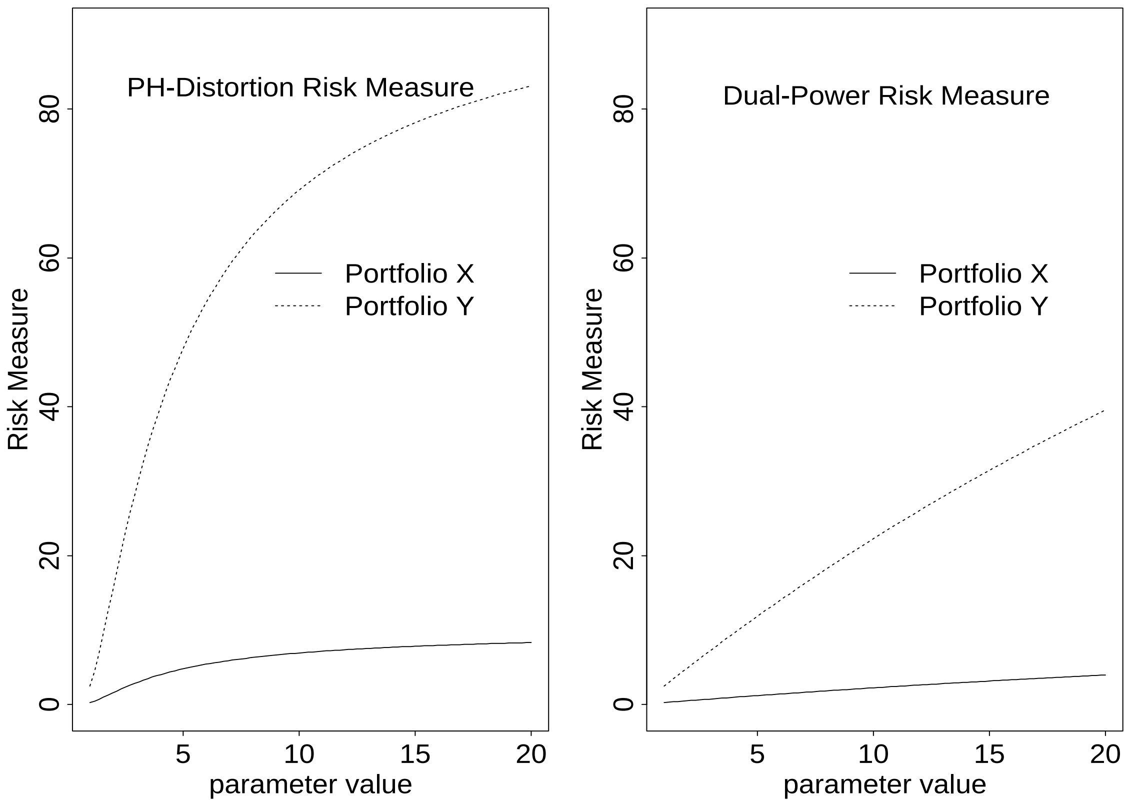

Figure 3.10: Capital Requirements for Example 3.3.3 using Proportional-Hazards and Dual-Power Distortions

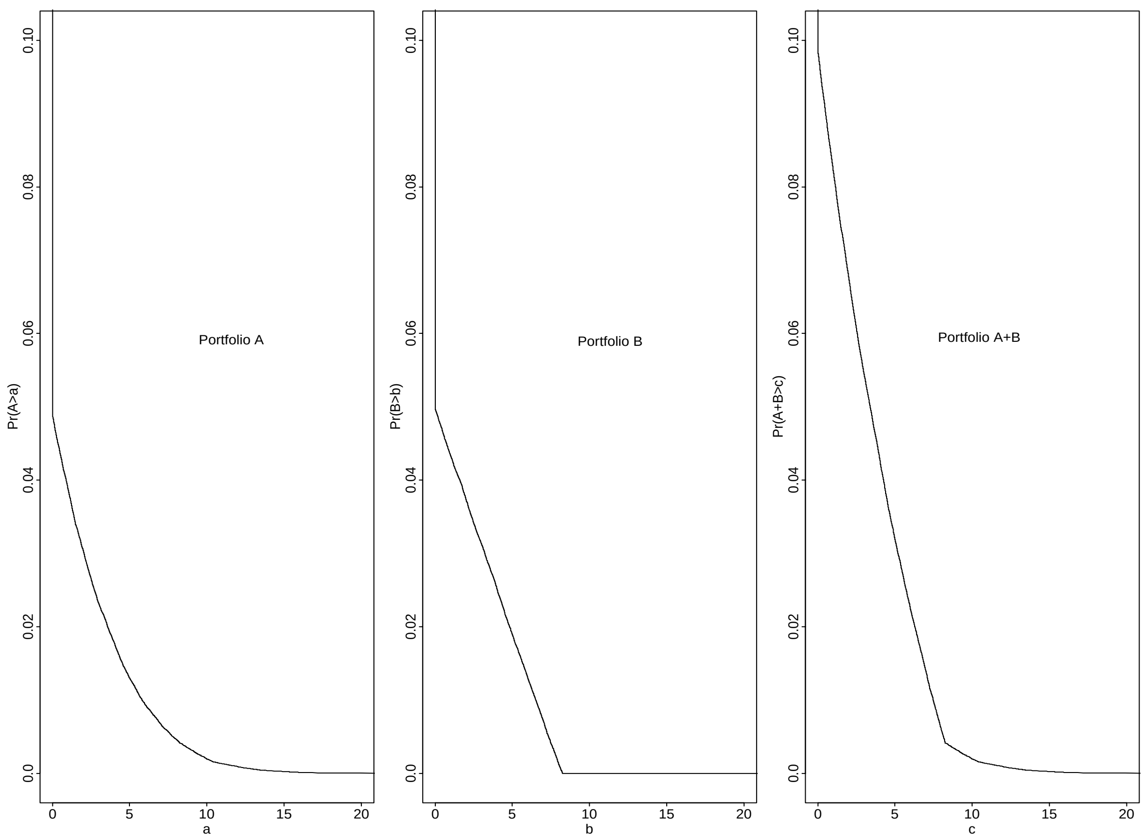

Figure 3.11: $S_A(a)$, $S_B(b)$ and $S_{A+B}(c)$ for Example 3.3.4

Example 3.3.4 This example is described in Example 2.2.3. The decumulative distribution functions for portfolio A and B and $A + B$ are shown in Figure 3.11.

The probability that the reverse butterfly spread will produce a loss is 5%, and the probability that the short put will produce a loss is 5%. Note that the support of the loss distributions is non-overlapping, and based on $V_{.95}$, the combined portfolio has a higher percentile-VaR than the two portfolios separately. This shows that percentile-VaR can be superadditive, and thus the percentile definition of value-at-risk fails to adhere to the subadditivity property(A4) of a consistent measure.

Table 3.2 illustrates the capital required using the risk measure discussed so far. The PH-distortion and the dual-power distortion risk measures maintain the same ordering between portfolios A and B; and the comined portfolio has a sub-additive capital requirement.

Portfolio:	A	B	A+B
Mean	0.1796	0.2058	0.3854
Maximum	89.18	8.249	89.18
$V_{.95}$	0.0	0.0	3.42
$\text{CTE}_{.90}$	1.796	2.058	3.854
PH 19	23.29	6.693	23.57
DP 19	2.722	2.952	4.573

Table 3.2: Comparison of Risk Measures for Example 2.2.3

3.4 Comparison of Distortion Methods

In this section, we use the distortion functions discussed previously to compare the same two parameter distributions that were used in section 2.6. For both the PH-distortion and the dual-power distortion, the value of the parameter determines the risk aversion inherent in the risk measure. Figure 3.12 shows how the PH-distortion compares the four distributions. For all values of the risk aversion parameter, the PH-distortion consistently ranks the distributions, identifying the Pareto distribution as the most risky, as expected.

Figure 3.13 compares the dual-power distortion for the same four distributions. In this figure, it is apparent that the ranking of the risks depends upon the

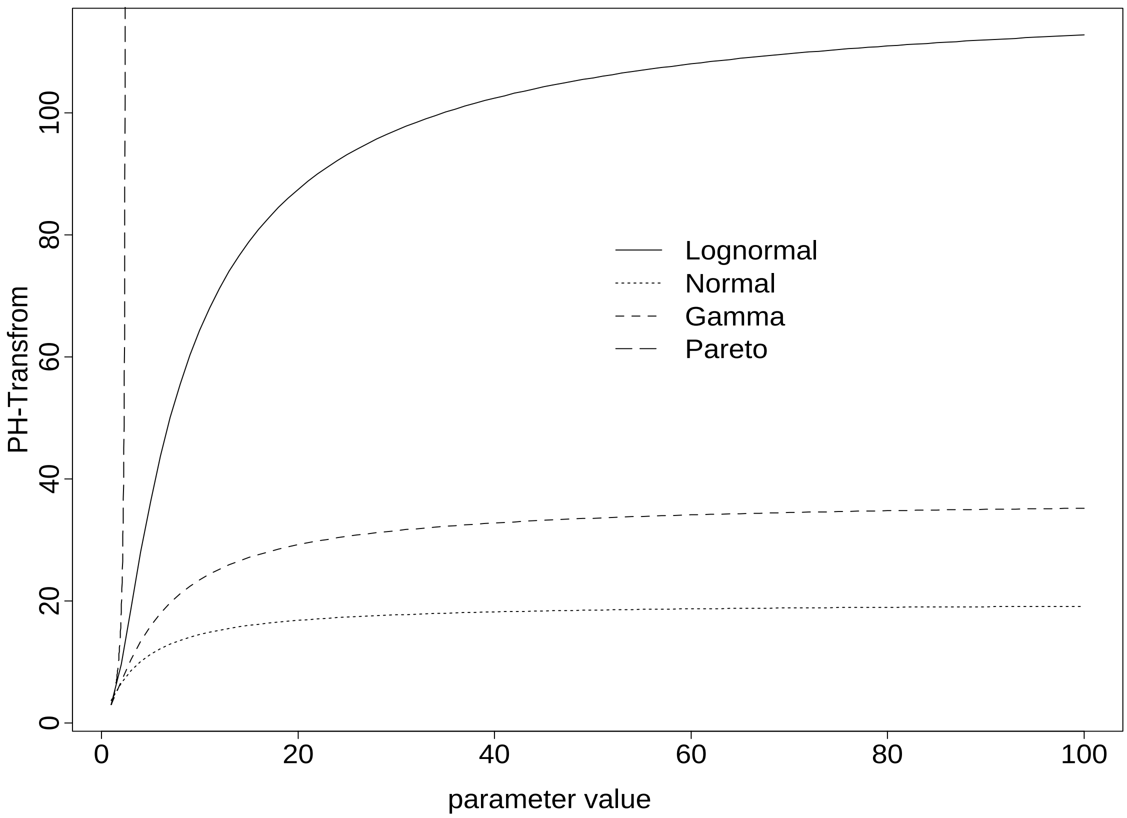

Figure 3.12: Capital Requirements for Two-Parameter Distributions (Mean= 3 and 95%ile= 10.415) using Proportional-Hazards Distortion

value of the risk aversion parameter. When the parameter is large enough, the order of risks is the same as the order seen when using the PH-distortion. However, when using a small parameter, the Pareto distribution is shown to be the least risky, which is the opposite of what would be expected for any risk averse person. The CTE also indicates a change in risk ordering depending upon the percentile chosen for truncation (see Figures 2.7 and 2.8).

The two-parameter distributions with the same mean and variance produce similar results (see Figures 3.14, 3.15, 2.9 and 2.10); however, the dual-power risk

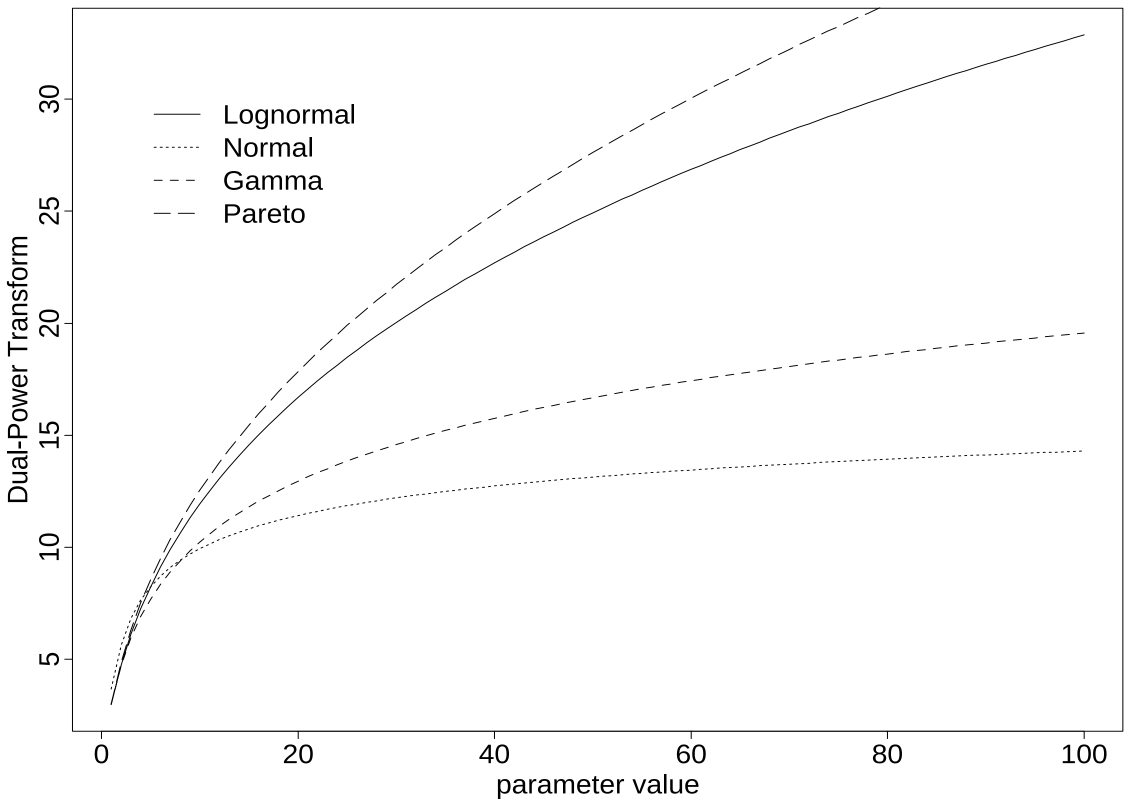

Figure 3.13: Capital Requirements for Two-Parameter Distributions (Mean= 3 and 95%ile= 10.415) using Dual-Power Distortion

measure does not rank the Pareto distribution as the most risky until very far out in the tail. The dual-power distortion risk measure for the gamma distribution crosses the lognormal distribution and will cross the Pareto distributions for some parameter value greater than 100.

This change in the ordering of risks has been postulated by Wang (1996b) to be due to the finite bound on the relative risk aversion at upper limits when using the dual-power distortion or the CTE distortion function. Unfortunately, Wang (1996b) has also shown that although an infinite bound of the relative risk aversion

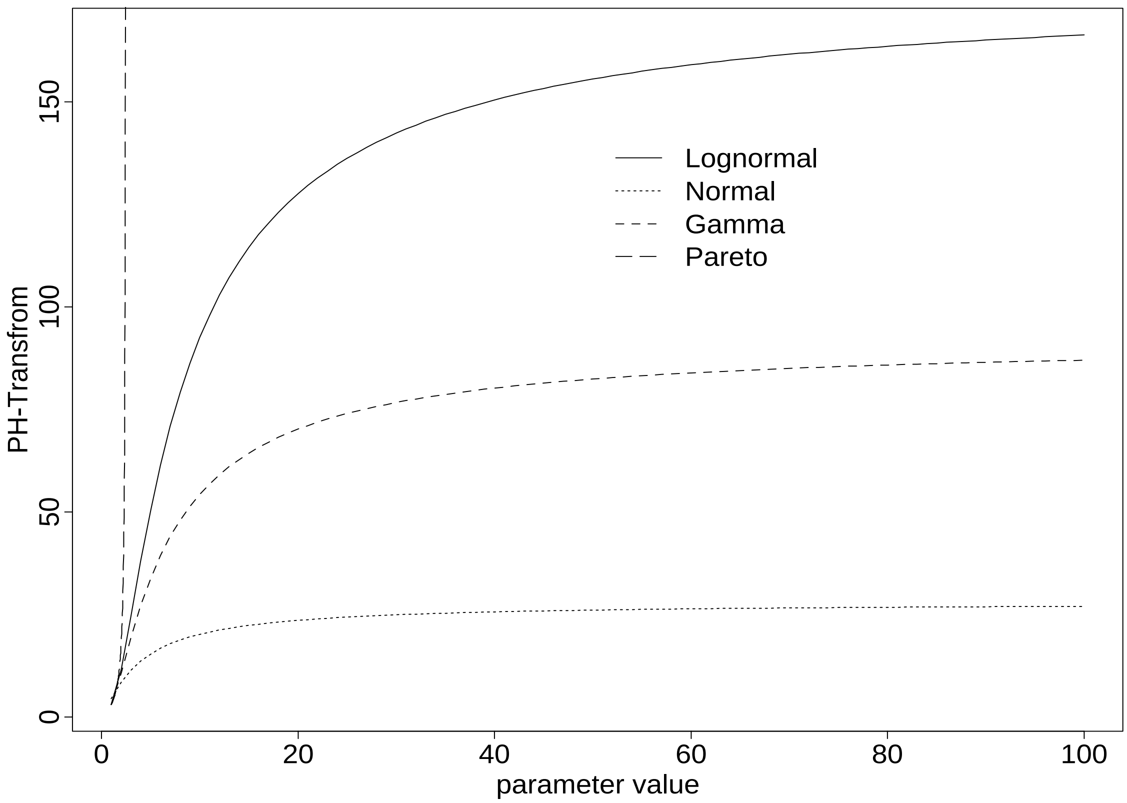

Figure 3.14: Capital Requirements for Two-Parameter Distributions (Mean= 3 and Variance= 45) using Proportional-Hazards Distortion

is necessary, it is not sufficient to ensure a consistent ranking of risks.

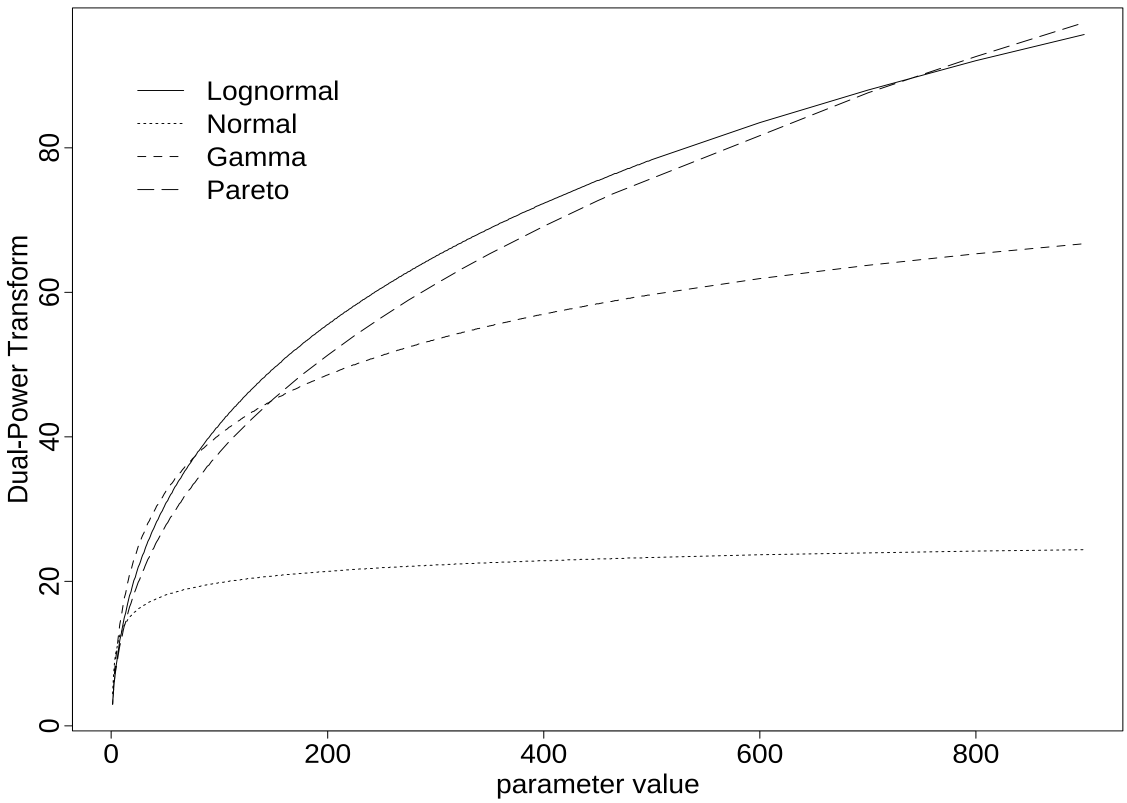

Figure 3.15: Capital Requirements for Two-Parameter Distributions (Mean= 3 and Variance= 45) using Dual-Power Distortion

Chapter 4

Features of the Beta Distortion

In this chapter, we recognize that both the PH-distortion and the dual-power distortion come from the beta family of distributions. Because of this generalization, we consider the full class of beta distributions and find a set of beta distortions that is coherent. As well, we consider characteristics of these risk measures in order to compare them with the current standards and identify appropriate parameter values.

4.1 The Beta Distortion

The PH-distortion and the dual-power distortion, discussed in Chapter 3, are special cases of the beta distortion function. Each of these distortion functions is an incomplete beta function (Hogg and Klugman, 1984):

$$
\begin{aligned}
g_\beta(S(x)) = \beta(a,b;S(x)) &= \int_0^{S(x)} \frac{1}{\beta(a,b)} t^{a-1}\,(1-t)^{b-1}\,dt, && (4.1)\\
&= F_\beta(S(x)) && (4.2)
\end{aligned}
$$

where $F_\beta(.)$ is the cdf of the Beta(a,b) distribution, and $S(x)$ is the decumulative distribution function, so that $0 \leq S(x) \leq 1$, and $\beta(a,b)$ is the beta function with parameters $a \geq 0$ and $b \geq 1$

$$
\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}\,(1-t)^{b-1}\,dt. \tag{4.3}
$$

To obtain the PH-distortion function, let $a = \frac{1}{\gamma}$,and $b = 1$, so that

$$
\begin{aligned}
g_P(S(x)) &= \int_0^{S(x)} \frac{1}{\beta(\frac{1}{\gamma},1)} t^{\frac{1}{\gamma}-1}\,(1-t)^{1-1}\,dt && (4.4)\\
&= S(x)^{\frac{1}{\gamma}}. && (4.5)
\end{aligned}
$$

To obtain the dual-power distortion function, let $a = 1$, and $b = \kappa$, so that

$$
\begin{aligned}
g_D(S(x)) &= \int_0^{S(x)} \frac{1}{\beta(1,\kappa)}\, t^{1-1}\,(1-t)^{\kappa-1}\,dt && (4.6)\\
&= 1-(1-S(x))^{\kappa}. && (4.7)
\end{aligned}
$$

Definition 4.1.1 *The general beta distortion risk measure is defined as*

$$\rho_\beta(X) \;=\; \int_0^\infty \int_0^{S_X(x)} \frac{1}{\beta(a,b)} t^{a-1}\,(1-t)^{b-1}\,dt\;dx. \tag{4.8}$$

for $a > 0$ *and* $b > 0$.

This risk measure is clearly differentiable for any $a > 0$ and $b > 0$. In order to limit the beta class of distortion functions to the subset that have the properties we need for coherence, we apply the following theorem:

Theorem 4.1.1 *The beta risk measure with parameters* a *and* b *is coherent if and only if* $0 < a \leq 1$ *and* $b \geq 1$.

Proof: The beta distortion risk measures (see Definition 1.4.1) satisfy $g_\beta(0) = 0$ and $g_\beta(1) = 1$, since the beta distortions are beta distribution functions over the interval $[0, 1]$. In addition, since $a > 0$ and $b > 0$, all beta distortion functions are twice differentiable all coherent beta risk measures can be found by determining the parameters a and b such that

$$g'_\beta(p) \;\geq\; 0, \quad g''_\beta(p) \;\leq\; 0, \qquad \text{for all} \;\; p, \;\; \text{where} \;\; 0 \leq p \leq 1. \tag{4.9}$$

Since, $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \;\geq\; 0$, for all $a > 0$ and $b > 0$.

$$g'_\beta(p) = \frac{1}{\beta(a,b)} p^{a-1}\,(1-p)^{b-1} \geq 0 \qquad \text{for all} \;\; 0 \leq p \leq 1, \tag{4.10}$$

and all beta distortion functions are increasing.

Since,

$$g''_\beta(p) = \frac{1}{\beta(a,b)} p^{(a-2)}\ (1-p)^{(b-2)}((a-1)+(2-a-b)p) \leq 0 \tag{4.11}$$

for all $0 \leq p \leq 1$,

$$\text{and} \quad \frac{1}{\beta(a,b)} p^{(a-2)}\ (1-p)^{(b-2)} \geq 0, \tag{4.12}$$

then the beta distortion function is concave if an only if

$$(a-1)+(2-a-b)p \leq 0 \quad \text{for all} \quad 0 \leq p \leq 1. \tag{4.13}$$

Splitting this into 3 cases we have

$$(a-1) \leq 0 \qquad \text{for } p = 0, \tag{4.14}$$

$$1-b \leq 0 \qquad \text{for } p = 1, \tag{4.15}$$

$$(a-1)(1-p)+(1-b)p \leq 0 \qquad \text{for } 0 < p < 1. \tag{4.16}$$

It is easy to see that for all $a \leq 1$ and $b \geq 1$, all three cases hold. So that the beta distortion function is concave if $a \leq 1$ and $b \geq 1$.

To show that the beta distortion function is concave only if $a \leq 1$ and $b \geq 1$, assume $g''_\beta(p) \leq 0$ for any p. Then $a \leq 1$ from setting $p = 0$.

Also, $g''_\beta(p) \leq 0$ if and only if

$$(a-1) + (-a-b+2)p \leq 0 \quad \text{for any} \quad p$$

and for $0 < a \leq 1$, this is true if and only if

$$(-b+1) \leq 0 \quad \text{by setting} \quad p = 1.$$

Hence, $g''_\beta(p) \leq 0$ for any p implies $a \leq 1$ and $b \geq 1$.

Thus any beta distortion risk measure, $\rho_\beta(X)$, is coherent if and only if $0 < a \leq 1, \ 1 \leq b \leq \infty$. $\square$

The coherent beta distortion risk measures are defined as:

$$\rho_\beta(X) = \int_0^\infty \int_0^{S_X(x)} \frac{1}{\beta(a,b)} t^{a-1} (1-t)^{b-1} \, dt \, dx. \tag{4.17}$$

where $0 < a \leq 1, \ 1 \leq b \leq \infty$.

In order to compare the beta distortion with the PH-distortion and the dual-power distortion, set $a = \frac{1}{\gamma}$ and $b = \kappa$, so that the beta distortion is

$$g_\beta(t) = \int_0^t \frac{1}{\beta(\frac{1}{\gamma}, \kappa)} t^{\frac{1}{\gamma}-1} (1-t)^{\kappa-1} \, dt, \tag{4.18}$$

and the dual-power distorted risk measure is

$$\rho_\beta(X) \;=\; \int_0^\infty \int_0^{S_X(x)} \frac{1}{\beta(\frac{1}{\gamma},\kappa)} t^{\frac{1}{\gamma}-1}\,(1-t)^{\kappa-1}\;dt\;dx. \tag{4.19}$$

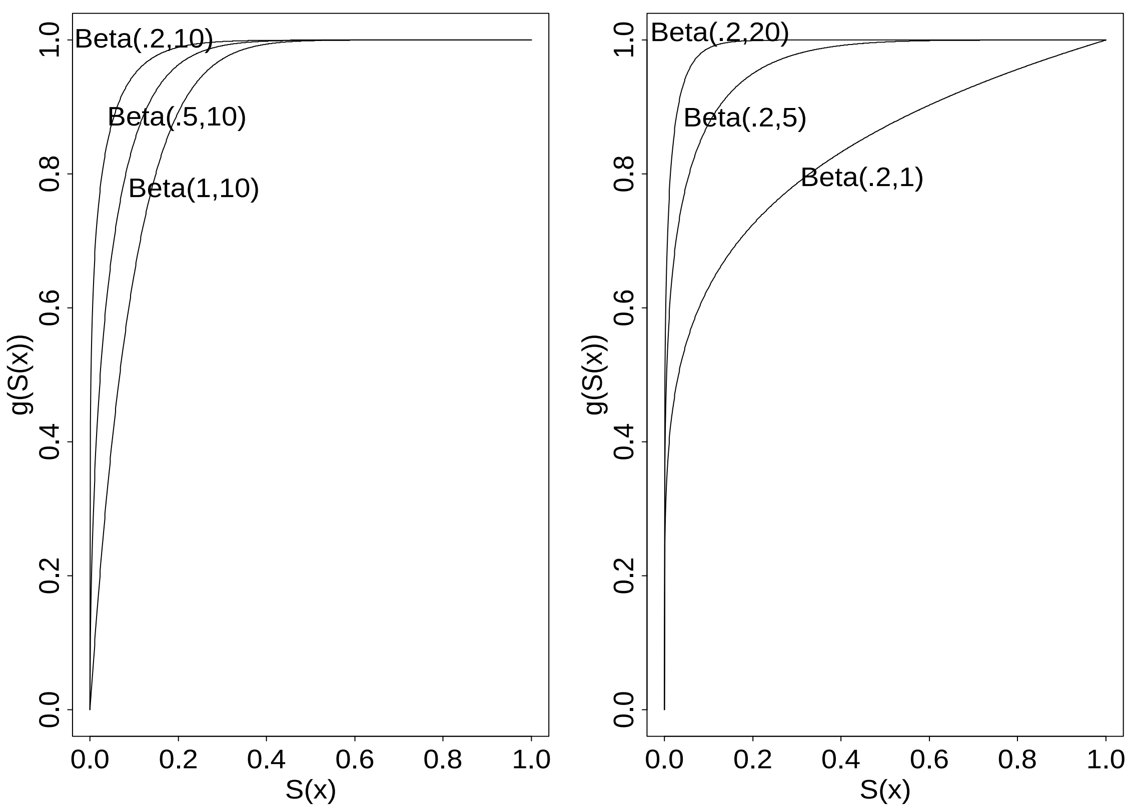

Figure 4.1: Beta($\frac{1}{\gamma},\kappa$) Distortion functions

Some examples of the beta distortion of the survival function are demonstrated in Figure 4.1. These graphs illustrate the effect of varying the parameters. The first graph demonstrates that higher values of the γ parameter increase the initial gradient of the distortion, which corresponds to higher aversion to the far right tail of the loss distribution. The second graph demonstrates that higher values of the κ parameter does little to affect the initial gradient, but increases the gradient for the moderately extreme events, which corresponds to higher risk aversion overall.

Since all coherent distorted risk measures satisfy first and second order stochastic dominance and are comonotonic (B4 on page 12), then we may consider further ordering properties for comparison of the beta distortion.

Definition 4.1.2 *For risk measure ρ with parameter space Ω, and arbitrary risks X and Y, given some $\omega_0 \in \Omega$ assume $\rho(X(\omega_0)) < \rho(Y(\omega_0))$.*

$$\textit{If } \rho(X(\omega)) \leq \rho(Y(\omega)) \quad \textit{for every} \quad \omega \in \Omega,$$

then ρ is defined as a **consistent** *risk measure, and Y is said to dominate X in ρ-order.*

For coherent beta distortions where $\gamma + \kappa > 2$, these measures order risks consistently with second order stochastic dominance. When risks are not comparable using second order stochastic dominance, the PH-distortion risk measure still seems to consistently compare risks independently of the parameter.

Definition 4.1.3 *For any risks X and Y, if $\rho_P(X) < \rho_P(Y)$ for any $\gamma \geq 1$, then we say that Y dominates X in PH-order.*

If Y dominates X, we say that X is less risky than Y. If we assume that the order of risks defined by the PH-distortion is what we want, and we call this ordering PH-order, then the problem we would like to solve is this:

If X is less risky than Y in PH-order, then we would like to show that

$$\frac{1}{\beta(\frac{1}{\gamma},\kappa)}\int_0^{\infty}\int_0^{S_X(x)} t^{\frac{1}{\gamma}-1}\,(1-t)^{\kappa-1}\,dt\,dx$$

$$\leq \frac{1}{\beta(\frac{1}{\gamma},\kappa)}\int_0^{\infty}\int_0^{S_Y(y)} t^{\frac{1}{\gamma}-1}\,(1-t)^{\kappa-1}\,dt\,dy \tag{4.20}$$

for all $0<\frac{1}{\gamma}\leq 1$ and $1\leq\kappa$.

We already know that this is true for any risks X and Y where Y dominates X with respect to second order stochastic dominance (which includes FSD, net stop loss order, order of dangerousness); however, we have not been able to prove that beta-order implies PH-order in general. Analyzing the beta distortion $g_\beta(X)$ for $\gamma\neq 1$, we have $g'_\beta(0)=\infty$, and if $1<\kappa$ then $g'(1)=0$, whereas when $\kappa=1$ (PH-distortion) $g'(1)=\frac{1}{\gamma}$. The gradient of $g(x)$ at $x=1$ is less than one, and determines the risk-adjusted probability allocated to minimal losses.

In chapter 3, we noted that the PH-distortion satisfies the multiplicativity property (B2), which is equivalent to the compound Bernoulli property, and as a result has a derivative of $+\infty$ at zero. The beta distortion does not satisfy the multiplicativity property (B2) but is submultiplicative and also has an infinite derivative at zero as long as γ is strictly greater than 1.

The compound Bernoulli property is desired for premium evaluations. If there is a transfer of risk between two parties, economic no arbitrage theory implies that there should be no free lunch for either party. However, in calculating capital adequacy margins, there is a difference between the capital you should hold to cover the frequency and severity risk and the capital you should hold to cover only the frequency risk with a predetermined constant severity. Since the frequency and severity risks are partially diversifiable, pooling the risk should lead to a lower capital requirement. Thus a comparable axiom could be that the capital requirement for risk $\rho(IY)$ should be less than or equal to the capital requirement for the unpooled risk, $\rho(I\rho(Y))$. Using the same method as Wang (1997), distortion function g must satisfy,

$$g(xy) \;\leq\; g(x)g(y), \qquad \text{where} \quad 0 \leq x, y \leq 1. \tag{4.21}$$

The functions g that satisfy this inequality are not as simple as in the case of equality; however we consider the case of our beta distortion, g_β with $\gamma \geq 1$ and $\kappa \geq 1$ and prove that these functions satisfy the submultiplicative property.

First we need to show that the beta distortion dominates the PH-distortion:

Proposition 4.1.1 *Given $\kappa \geq 1$ and $\gamma \geq 1$, the beta distortion dominates the PH-distortion for all $x \in (0,1)$; that is $g_\beta(x) \geq g_P(x)$ for all $x \in (0,1)$.*

Proof: In other words, we need to show that

$$g_\beta(x) \geq g_P(x) \qquad \text{for all} \quad x \in (0,1). \tag{4.22}$$

When $\kappa = 1$, we have that $g_\beta(x) = g_P(x)$ by definition, so assume $\kappa > 1$. We also have that $g_P(x) = x^{\frac{1}{\gamma}}$ by definition. So consider the function

$$f(y) = g_\beta(y) - y^{\frac{1}{\gamma}}. \tag{4.23}$$

Then, $f(y) = 0$ when $y = 0$ or $y = 1$, and

$$\frac{d}{dy} f(y) = y^{(\frac{1}{\gamma}-1)} \left(\frac{(1-y)^{\kappa-1}}{\beta(\frac{1}{\gamma},\kappa)} - \frac{1}{\gamma} \right) \tag{4.24}$$

which implies that $\frac{d}{dy} f(y) = 0$ at $y = 0$ or at $y = 1 - \left(\frac{\beta(\frac{1}{\gamma},\kappa)}{\gamma} \right)^{\frac{1}{\kappa-1}}$, which lies in the interval $(0,1)$ since $\beta(\frac{1}{\gamma},\kappa) \leq \gamma$, and $\frac{d}{dy} f(y)|_{y=1} = -\frac{1}{\gamma}$. So, $f(y)$ does not have a root between 0 and 1, and $f(y) > 0$ for all $y \in (0,1)$. Thus $g_\beta(y) > y^{\frac{1}{\gamma}}$ for all $y \in (0,1)$. □

Proposition 4.1.2 *Let $X = IY$ be a compound Bernoulli risk, where the Bernoulli frequency random variable I is independent of the loss severity random variable $Y = X|X > 0$. Then the capital requirement for risk $X = IY$ is less than or equal to the capital requirement for risk $I\rho(Y)$ for the beta distortion risk measure.*

Proof: Based on the proof of Theorem 3 in Wang (1997), it is sufficient to prove that g_β, the beta distortion function with parameters $0 < \frac{1}{\gamma} \leq 1$ and $\kappa \geq 1$

satisfies submultiplicativity. That is, we want to prove that

$$g_\beta(x)\ g_\beta(y) \geq g_\beta(xy), \quad \text{for all} \ \ x, y \ \in \ [0,1]. \tag{4.25}$$

For $\kappa = 1$, g_β is the PH-distortion and the proof from Wang (1997) shows that $g_P(x)\ g_P(y) = g_P(xy)$, so we consider the case for $\kappa > 1$.

Since $g_\beta(x)$ is a distortion function, we know that $g_\beta(0) = 0$ and $g_\beta(1) = 1$. As well, $g_\beta(x)$ is a smooth continuous function, and the derivative exists for all $x \ \in \ (0,1)$. Since Equation 4.1 is symmetric in x and y, we choose any arbitrary y in $(0,1)$, and consider the function

$$f(x) = g_\beta(x)\ g_\beta(y) - g_\beta(xy) \tag{4.26}$$

$f(x) = 0$ when $x = 0$ or $x = 1$.

To show that there are no other roots of $f(x)$, we prove that $\frac{d}{dx}f(x)$ has a unique root in $(0,1)$.

$$\frac{d}{dx}f(x) = g_\beta(y)g'_\beta(x) - yg'_\beta(xy) \tag{4.27}$$

$$= \frac{x^{\frac{1}{\gamma}-1}}{\beta(\frac{1}{\gamma},\kappa)}\left[g_\beta(y)(1-x)^{\kappa-1} - y^{\frac{1}{\gamma}}(1-xy)^{\kappa-1}\right] \tag{4.28}$$

which has a unique solution of

$$x = \frac{\frac{g_\beta(y)}{y^{\frac{1}{\gamma}}}^{\frac{1}{\kappa-1}} - 1}{\frac{g_\beta(y)}{y^{\frac{1}{\gamma}}}^{\frac{1}{\kappa-1}} - y} \tag{4.29}$$

Since $g_\beta(y) > y^{\frac{1}{\gamma}}$ and $1 > y$, then $x \in (0,1)$.

Next we show that $\lim_{x\to 0^+} \frac{d}{dx} f(x)$ is positive, and since $f(0) = 0$, this proves that $f(x)$ is positive for all $x \in (0,1)$.

We want to show that

$$\frac{d}{dx} f(x) = g_\beta(y) g'_\beta(x) - y g'_\beta(xy) > 0, \quad \text{as} \quad x \to 0^+. \tag{4.30}$$

Which is the same as showing that

$$\frac{g_\beta(y)}{y} - \frac{g'_\beta(xy)}{g'_\beta(x)} > 0, \quad \text{as} \quad x \to 0^+. \tag{4.31}$$

We know that

$$\lim_{x\to 0^+} \frac{g'_\beta(xy)}{g'_\beta(x)} = \lim_{x\to 0^+} \frac{(xy)^{\frac{1}{\gamma}-1}(1-xy)^{\kappa-1}}{(x)^{\frac{1}{\gamma}-1}(1-x)^{\kappa-1}} = y^{\frac{1}{\gamma}-1}. \tag{4.32}$$

And so, we need to show that

$$\frac{g_\beta(y)}{y} - y^{\frac{1}{\gamma}-1} > 0 \quad \text{for some} \quad y \in (0,1) \tag{4.33}$$

$$\text{or} \quad g_\beta(y) > y^{\frac{1}{\gamma}} \quad \text{for some} \quad y \in (0,1) \tag{4.34}$$

However, we know (see Proposition 4.1.1) that

$$g_\beta(y) \geq y^{\frac{1}{\gamma}} \quad \text{for any} \quad y \in (0,1) \tag{4.35}$$

for $\kappa < 1$ and $\gamma > 1$. So we have that

$$g_\beta(x)\ g_\beta(y) \geq g_\beta(xy), \quad \text{for all} \quad x, y \ \in \ [0,1]. \ \square \tag{4.36}$$

We have also illustrated that g_β satisfies (4.1) by drawing contour graphs for $g(x)g(y) - g(xy)$ over a range of parameters for all combinations of x and y. See Figure 4.2.

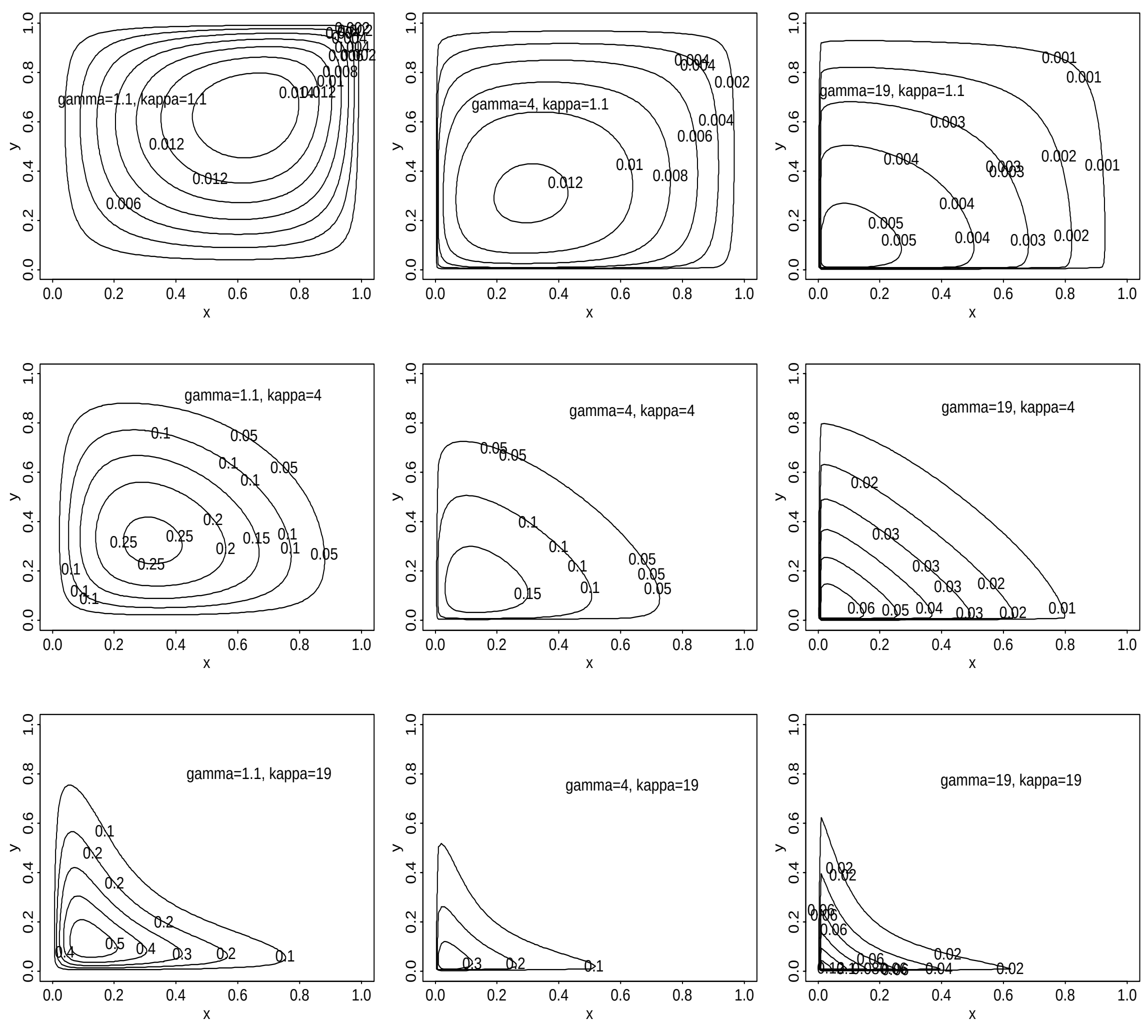

Figure 4.2: Contours of $g(x)g(y) - g(xy)$ for Various Beta Distortion Functions

4.2 Parameterizing the Beta Distortion

Although there are many parameter values that make the beta distortion risk measure coherent risk, there is no strict preference for choice of parameters. In this section we use a uniform distribution for the risk and the von Mises and Kullback-Leibler Information measures to compare the beta risk measures with percentile-VaR and the CTE. Using these comparisons, it is possible to identify parameters that are consistent with current regulatory standards.

4.2.1 Non-Informative Risk

In order to set parameters for the risk measures that are appropriate for many applications in a variety of industries, we must not put too much emphasis on extreme risk exposures. For applications that are concerned with these, an auditor would most likely be involved with analyzing the risk and should identify it as an extreme exposure.

To choose a portfolio to compare the risk measures is subjective. In order to limit the subjectivity of choosing a risk distribution, we rely on the idea of maximum entropy. The entropy of a distribution $f(x)$ is defined as

$$\epsilon(f) = -\int_0^\infty f(x)\log(f(x))dx. \tag{4.37}$$

Entropy is a measure of the missing information needed on average to describe a random variable (Cover and Thomas, 1991). Entropy only depends on the probabilities of the random variable and not on the value of the random variable. The distribution $f(x)$ that maximizes the entropy, is the one that is least subjective with respect to the unknown information (Buchen and Kelly, 1996).

The distribution that maximizes entropy on a finite interval with no prior information, is the uniform distribution. If the interval is infinite, and the mean is known, then the exponential distribution maximizes entropy. Since the maximal loss is usually considered to be finite (the maximal loss before bankruptcy), the uniform distribution maximizes entropy. Similarly, in Bayesian statistics, the uniform distribution is considered to be the most uninformative prior distribution, as all outcomes are equally probable.

To compare parameters, each risk measure is applied to a risk with a uniform distribution. Since the risk distribution is completely arbitrary, the uniform distribution gives the most general results; equating this solution with the regulatory standard gives a method to compare the risk measures and determine parameters for the beta risk measures.

Equating these risk measures when applied to a uniform loss distribution is equivalent to equating the areas under the distortion graphs. If we set the percentile-VaR parameter to $\alpha = 0.95$, then the equivalent CTE parameter is $\alpha = 0.90$. For the PH-distortion risk measure, $\gamma = 19$, the dual power risk measure, $\kappa = 19$, and the general beta risk measure could use any (γ, κ), as long as $\gamma\kappa = 19$. For our purposes, we have chosen to use $\gamma = \kappa = \sqrt{19}$. Figure 4.3 shows the distortion functions using these parameters. Using extreme value theory from Section 2.4, the dual-power distortion with $\kappa = 19$ gives a value approximately equal to $E[Y_{(19)}]$ from a sample of size 20, which is also an estimate of the 95th percentile.

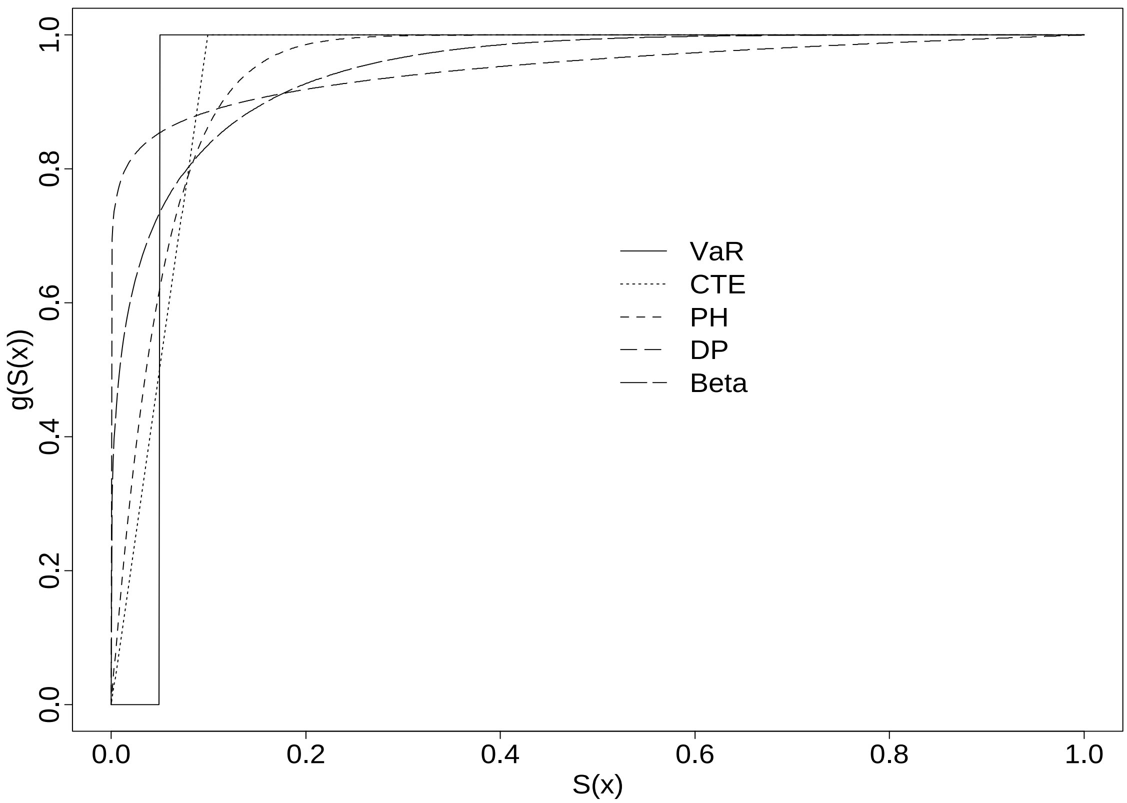

Figure 4.3: Distortion Functions Applied to a Uniform Risk

Both percentile-VaR and the CTE ignore the shape of a large portion of the risk distribution. They may identify some extreme exposures; however, portfolios with the same extreme exposures may vary greatly in the shape of the rest of their distributions, and should not be considered equal. The beta risk measures consider the full distribution. The PH-distortion risk measure parameter γ penalizes extreme tail risks more heavily, whereas the dual-power risk measure parameter κ penalizes moderate risks more heavily. This will be more evident with the illustrations in the next chapter.

4.2.2 Relative Entropy

Using a topic closely linked to entropy, relative entropy (Cover and Thomas, 1991) or the Kullback-Leibler(KL) information, we are able to define a distance between the distorted distribution and the original risk distribution. Relative entropy or the KL information between two density functions is defined as follows:

Definition 4.2.1 *Let $f(x)$ and $h(x)$ be two density functions. Then we define the KL information as:*

$$I(f(x), h(x)) = \int_0^{\infty} f(x) \log\left(\frac{f(x)}{h(x)}\right) dx. \tag{4.38}$$

The KL information can be interpreted as the expected information when the distribution with density function $h(.)$ is transformed into the distribution with density function $f(.)$.

Definition 4.2.2 *The function $M(.,.)$ is a metric on the set of real valued functions Ω, if for any arbitrary functions f, g, $h \in \Omega$, $M(.,.)$ satisfies all of the following:*

1. *Non-negative: $M(f,h) \geq 0$;*

2. *Zero: $M(f,h) = 0$ if and only if $f \equiv h$ almost surely;*

3. *Symmetry: $M(f,h) = M(h,f)$;*

4. *Triangle inequality: $M(f,h) \leq M(f,g) + M(g,h)$.*

The KL information measure is always non-negative. It is equal to zero when $f(x) = h(x)$, and additive for independent risks. However, it is not symmetric. That is, $I(f(x), h(x)) \neq I(h(x), f(x))$, and thus I is not a metric, as the triangle inequality does not hold. It is possible to create a metric form of this measure by defining the metric information as

$$MI(f(x), h(x)) \; = I(f(x), h(x)) \; + \; I(h(x), f(x)). \tag{4.39}$$

In most applications, relative entropy is minimized in order to find the density that is the closest to what is observed. The parameters $\gamma = 1$ and $\kappa = 1$ (in Equation 4.18) minimize relative entropy. However, if minimized, the distortion function would be a uniform distortion and our risk measure would be the expected value of the risk. For capital adequacy, the objective is not entropy minimization. Instead we use this distance to equate the parameters of the various distortions, for an arbitrary loss distribution.

Let $S(x)$ be the pdf of an arbitrary loss random variable X with pdf $h(x)$, and let $f(x)$ be the distorted pdf. Since $h(x) = -S'(x)$ and $f(x) = -\frac{d}{dx}g(S(x)) = g'(S(x))h(x)$, then

$$I(f(x), h(x)) \;\; = \;\; \int_0^\infty \; g'(S(x))h(x) \log g'(S(x) \; dx, \tag{4.40}$$

$$\text{and} \quad I(f(x), h(x)) \;\; = \;\; E_X[g'(S(x)) \log(g'(S(x))]. \tag{4.41}$$

where the expectation is taken with respect to the $h(.)$ measure.

Applying the KL information to the PH-distortion we simplify and obtain:

$$I_P(f(x), h(x)) = \gamma - \log(\gamma) - 1, \tag{4.42}$$

for the dual-power distortion:

$$\begin{aligned} I_D(f(x), h(x)) &= \frac{1 - \kappa + \kappa \log(\kappa)}{k} = \log(k) - 1 + \frac{1}{k} && (4.43) \\ &= I_P(f(x), h(x)) && (4.44) \end{aligned}$$

with $\gamma = \kappa$.

For the beta distortion:

$$\begin{aligned} I_\beta(f(x), h(x)) &= -\log(\beta(\kappa, \frac{1}{\gamma})) && (4.45) \\ &+ \left[\Psi(\frac{1}{\gamma})(\frac{1}{\gamma} - 1) + (\kappa - 1)\Psi(\kappa) - \Psi(\kappa + \frac{1}{\gamma})(\frac{1}{\gamma} + \kappa - 2) \right] \end{aligned}$$

where $\Psi(z) = \frac{d}{dz}\Gamma(z)$ and $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$.

In terms of the KL information, Table 4.1 compares the PH-distortion, the dual-power distortion and the beta distortion information. Table 4.2 lists the parameters that give the same distance between the original loss distribution and the distorted loss distribution.

Since this information measure is not a metric, we consider both the metric form of the KL information, as well as the most common metric used in mathematics, the von Mises distance, or the L_2 norm.

Parameters		**Beta**	**PH-Distortion**	**Dual-Power**
γ	κ			
1	1	0.00	0.00	0.00
2	2	1.75	0.31	0.19
4	4	2.99	1.61	0.64
19	19	18.42	15.06	2.00
$\sqrt{19}$	$\sqrt{19}$	3.39	1.89	0.70

Table 4.1: Kullback-Leibler Distances

	Beta		**PH-Distortion**		**Dual-Power**	
KL Distance	γ	κ	γ	κ	γ	κ
0.765	2	2	4.730	1	1	2.792
2.993	4	4	53.21	1	1	5.741
3.386	$\sqrt{19}$	$\sqrt{19}$	70.28	1	1	6.212

Table 4.2: Parameters giving Equivalent Kullback-Leibler Distances

The metric form of the KL information (MKL), is defined as,

$$MI(f(x), h(x)) = I(f(x), h(x)) + I(h(x), f(x)) \tag{4.46}$$

which can be simplified to,

$$\begin{aligned} MI(f(x), h(x)) &= \int_0^\infty g'(S(x))h(x)\log(g'(S(x))\ dx \\ &\quad - \int_0^\infty h(x)\log(g'(S(x))\ dx \end{aligned} \tag{4.47}$$

where $S(x)$ is the ddf, and $h(x)$ is the pdf of the loss random variable X. Hence,

$$MI(f(x), h(x)) = E_X\left[(g'(S(x)) - 1)\log(g'(S(x)))\right]. \tag{4.48}$$

Applying the MKL information to the PH-distortion, simplifies to

$$MI_P(f(x), h(x)) = \gamma + \frac{1}{\gamma} - 2, \tag{4.49}$$

and for the dual-power distortion,

$$MI_D(f(x), h(x)) = \kappa + \frac{1}{\kappa} - 2. \tag{4.50}$$

Thus, the metric form of the KL information introduces symmetry between the dual-power and the PH-distortion. This symmetry is somewhat evident when the metric is applied to the beta distortion,

$$\begin{aligned} MI_\beta(f(x), h(x)) &= (\frac{1}{\gamma} - 1)(\psi(\frac{1}{\gamma}) - \psi(\kappa + \frac{1}{\gamma})) \\ &+(\kappa - 1)(\psi(\kappa) - \psi(\kappa + \frac{1}{\gamma})) + \frac{1}{\gamma} + \kappa - 2. \end{aligned} \tag{4.51}$$

However, this symmetry is only on the boundaries of the parameter space, that is for the cases when $\kappa = 1$, or when $\gamma = 1$.

Definition 4.2.3 *For a function $f(x)$, the L^p norm is defined as,*

$$L^p(u) = \left(\int_0^\infty \mid f(x) \mid^p dx \right)^{\frac{1}{p}}.$$

The L^2 norm, where $f(x)$ is a difference between two functions is often called the von Mises distance. For two distributions, $f(x)$ and $h(x)$, with respective ddfs of $S^*(x)$ and $S(x)$, the von Mises distance is defined as

$$W^2 = (L^2(S^*(x) - S(x)))^p = \int_0^1 |S^*(x) - S(x)|^2(-dS(x)). \tag{4.52}$$

From Equation 4.52, it can be seen that this distance is a metric, and measures the distance between the distortion and the identity distortion. This measure is strictly positive, symmetric and zero when $S(x) = S^*(x)$. Since $S(x)$ is a $Uniform(0, 1)$ random variable, the von Mises distance is independent of the loss distribution and for the PH-distortion, the distance simplifies to:

$$W_P^2 = \frac{1}{3} - \frac{2}{\gamma + 2} + \frac{1}{2\gamma + 1}. \tag{4.53}$$

For the dual-power distortion, the distance equates to:

$$W_D^2 = \frac{1}{3} - \frac{2}{\kappa + 2} + \frac{1}{2\kappa + 1}. \tag{4.54}$$

The dual-power distortion and the PH-distortion with $\kappa = \gamma$, are power functions of the cumulative or decumulative distribution functions. These distortions are symmetric in $x + y = 1$.

Proposition 4.2.1 *Given any $p \geq 1$, the L^p norm for the difference between $g(S(x))$ and $S(x)$ gives the same distance measure for the dual-power distortion as for the PH-distortion when $\gamma = \kappa$.*

Here, both the dual-power and PH-distortion functions distort the survivor distribution, but they focus the distortion on different parts of the distribution.

Proof: For the PH-distortion, the L_p norm is:

$$\begin{aligned}
\int_0^1 |g(S(x)) - S(x)|^p dx &= \int_0^1 |S(x)^{\frac{1}{\gamma}} - S(x)|^p dx && (4.55)\\
&= \int_0^1 |u^{\frac{1}{\gamma}} - u|^p du && (4.56)\\
&= \int_0^1 u^{\frac{p}{\gamma}}(1 - u^{1-\frac{1}{\gamma}})^p du && (4.57)\\
&= \frac{\gamma}{\gamma - 1} \frac{\Gamma(\frac{p+1}{\gamma-1} + 1)\Gamma(p+1)}{\Gamma(\frac{\gamma(p+1)}{\gamma-1} + p + 2)} && (4.58)\\
&= \frac{1}{\gamma - 1} \frac{\Gamma(\frac{p+1}{\gamma-1})\Gamma(p+1)}{\Gamma(\frac{\gamma(p+1)}{\gamma-1})} && (4.59)
\end{aligned}$$

For the dual-power distortion, the L_p norm is:

$$\begin{aligned}
\int_0^1 |g(S(x)) - S(x)|^p dx &= \int_0^1 |F(x) - F(x)^\gamma|^p dx && (4.60)\\
&= \int_0^1 u^p(1 - u^{\gamma-1})^p dx && (4.61)\\
&= \frac{1}{\gamma - 1} \frac{\Gamma(\frac{p+1}{\gamma-1})\Gamma(p+1)}{\Gamma(\frac{\gamma(p+1)}{\gamma-1})}. \quad \square && (4.62)
\end{aligned}$$

Table 4.3 shows the von Mises and MKL distances for parameters κ and γ such that $\kappa\gamma = 20$. An interesting aspect of this table is that, when we transpose the parameters so that $\kappa > \gamma$, the von Mises and the MKL distances are both greater than when $\gamma > \kappa$. This difference is very small; however this is counter-intuitive for the measures we are using.

Parameters		**Distance**	
γ	κ	**Von Mises**	**MKL**
1	20	0.2668	18.05
20	1	0.2668	18.05
5	4	0.2596	7.31
4	5	0.2598	7.38
10	2	0.2617	9.86
2	10	0.2624	10.17

Table 4.3: Von Mises and MKL Distances for Parameters $\kappa\gamma = 20$

	Beta		**PH Distortion**		**Dual-Power**	
MKL Distance	γ	κ	γ	κ	γ	κ
1.55	2	2	3.24	1	1	3.24
6.21	4	4	8.08	1	1	8.08
38.24	19	19	40.21	1	1	40.21
Von Mises Distance	γ	κ	γ	κ	γ	κ
0.1065	2	2	3.85	1	1	3.85
0.2451	4	4	14.43	1	1	14.43
0.3281	19	19	286.36	1	1	286.36

Table 4.4: Parameters giving Equivalent von Mises and MKL Distances

According to our objectives, we consider losses far out in the tail to be more risky than medium sized losses; although the von Mises metric does suggest that both parameters are equally important (see Table 4.4), the Kullback-Leibler distance gives us results that are much more intuitive for capital adequacy purposes.

The von Mises metric for the general beta distortion can not be as easily simplified,

$$W^2 = \int_0^1 \left\{ u - \int_0^u \frac{1}{\beta(\frac{1}{\gamma},\kappa)} t^{\frac{1}{\gamma}-1}(1-t)^{\kappa-1}dt \right\}^2 du. \tag{4.63}$$

In selecting parameters for the beta distortion, we rely on the KL information. Using $\kappa = 19$ for the dual-power distortion would give an approximation for the 95th percentile. Using current regulatory standards, the current approach uses a 95 percent VaR, and then multiplies by three. The KL information for the dual-power $\kappa = 19$ measure is 2.00. By equating the distorted risk measures applied to a risk with a uniform distribution, the equivalent parameter for the PH-distortion is $\gamma = 19$, which is quite extreme when applied to practical examples. The KL information for the PH-distortion with $\gamma = 19$ is 15.06. The equivalent beta parameters, when applied to a uniform risk, are $\gamma\kappa = 19$. If we choose $\gamma = \kappa = \sqrt{19}$, the KL information is 3.39, which is not that much higher than choosing $\gamma = \kappa = 4$ with an information of 2.99. In calculating premiums, Wang suggests a PH-distortion parameter of 2, which would have a KL information of 0.31, which is extremely low; however the risks used in his applications are often diversifiable. At this point, we postulate that a beta distortion with parameters $\gamma = \kappa = 4$ seems appropriate for capital adequacy purposes. These parameters will be tested in Chapter 5.

4.3 Tail Behaviour Under Distortions

There are many ways of describing the tail behaviour of a distribution. Often in actuarial science, we discuss distributions with increasing mean residual lifetimes or decreasing failure rates. In statistics and finance, these are often referred to as heavy-tailed distributions, or distributions with kurtosis higher than the normal distribution. In this chapter, we would like to classify these characteristics and determine how the distortions methods affect these characteristics. To start, we define the tail measures that are currently used and discuss any associations between them. Then we apply these measures to an unknown survival distribution and compare this to the distorted survival distribution under the PH-distortion, dual-power distortion and beta distortions, as well as the CTE and VaR distortions. Then, we use the Weibull distribution, which can be heavy-tailed or light-tailed depending on the parameters chosen, and see how these distortions affect the tail. Next, we see that there are conclusions that translate to other families of distributions. Lastly, we consider the four two-parameter distributions and apply some of these measures to illustrate our observations.

4.3.1 Tail Measures

There are two functions that help to classify characteristics of the tail of a distribution. The first is the Mean Excess Loss (MEL), also known as the Mean Residual Lifetime (MRL) (Klugman, Panjer and Willmot, 1998):

Definition 4.3.1 *Given $S(x)$, the ddf of a non-negative loss random variable X, the mean residual lifetime $e(x)$ is defined as:*

$$e(x) = \int_0^\infty \frac{S(x+y)}{S(x)}\,dy, \qquad x \geq 0. \tag{4.64}$$

Equivalently,

$$e(x) = \int_0^\infty \frac{S(t)}{S(x)}\,dt, \qquad \text{as long as} \quad 0 < e(0) < \infty \ \text{ exists.} \tag{4.65}$$

If the MRL is large for large values of x, then the expected loss for $X - x$ is large, and the distribution is heavy-tailed. The MRL is a linear transformation of the CTE,

$$e(x) = E[X - x | X > x] = CTE(x) - x. \tag{4.66}$$

The second function to help characterize the tail of a distribution is the failure rate, also called the hazard rate or the force of mortality:

Definition 4.3.2 *Given $S(x)$ the ddf of loss random variable X, the failure rate $\lambda(x)$ is defined as:*

$$\lambda(x) = -\frac{d}{dx}\log S(x). \tag{4.67}$$

Equivalently,

$$\frac{1}{\lambda(x)} = \frac{f(x)}{1 - F(x)}, \qquad \text{where } f(x) \text{ is the pdf of } x. \tag{4.68}$$

In terms of a loss distribution, for small dx, $\lambda(x)dx$ is the probability that the loss is close to x, given that the loss is greater than or equal to x. If $\lambda(x)$ is small for large x, the loss is likely to be larger than x and the distribution is heavy-tailed.

If a distribution has a monotone decreasing failure rate (DFR) or a monotone increasing mean residual lifetime (IMRL) then the distribution is said to have a heavy tail. If a distribution has a monotone increasing failure rate (IFR) or a monotone decreasing mean residual lifetime (DMRL) then the distribution is said to have a light tail. If a distribution has monotone increasing failure rate (non-decreasing), then it has a DMRL. As well, if distribution has monotone decreasing failure rate (non-increasing), then it has an IMRL. Thus, monotone DFR implies IMRL and monotone IFR implies DMRL; however the reverse is not implied (Klugman, Panjer and Willmot, 1998).

Another connection between the failure rate and the MRL lies in their limits:

$$\lim_{x\to\infty} e(x) = \lim_{x\to\infty} \frac{1}{\lambda(x)} \qquad \text{if the limits exist.} \tag{4.69}$$

It is possible for a distribution to have both an increasing and a decreasing failure rate. The lognormal failure rate starts out at zero, increases to a maximum, and then decreases to a limit of zero. The maximum failure rate is attained at a solution to

$$N(z) \;=\; 1 \;-\; \frac{1}{\sqrt{2\pi}(\sigma + z)} e^{\frac{1}{2}z^2}, \quad \text{where} \quad z = \frac{\log(x) - \mu}{\sigma}, \quad x \geq 0 \qquad (4.70)$$

and $N(.)$ is the Normal cdf.

It is also possible for a distribution to have either an increasing or a decreasing failure rate, depending on the value of its parameters. The Weibull and Gamma distributions are examples of this. As a special case, the exponential distribution has constant failure rate, λ, and is considered to have both an increasing and a decreasing failure rate. Table 4.5 illustrates some characteristics of the failure rate for the Pareto, Lognormal, Weibull, Gamma and Normal distributions.

Another measure of the tail is the coefficient of **kurtosis** (Hogg and Craig, 1995), which is defined by:

$$K \;=\; \frac{E[(X - \mu)^4]}{\sigma^4} \qquad (4.71)$$

where μ is the mean and σ is the standard deviation of a random variable X. The kurtosis of the normal distribution is 3. A fatter tailed distribution would have a larger kurtosis.

To consider how distortion functions affect tail behaviour, the failure rate of a loss is compared to the distorted failure rates. To start, we define the survival function under each of the five distortions:

Distribution	$\lambda(\mathbf{0})$	$\lim_{\mathbf{x}\to\infty} \lambda(\mathbf{x})$	**Failure Rate**
Pareto (α,θ)	$\frac{\alpha}{\theta}$	0	DFR
Lognormal (μ,σ)	0	0	neither
Weibull (τ,θ)	$0 \Leftrightarrow \tau > 1$ $\frac{\tau}{\theta} \Leftrightarrow \tau = 1$ $\infty \Leftrightarrow \tau < 1$	$\infty \Leftrightarrow \tau > 1$ $\frac{\tau}{\theta} \Leftrightarrow \tau = 1$ $0 \Leftrightarrow \tau < 1$	IFR $\Leftrightarrow \tau > 1$ constant $\Leftrightarrow \tau = 1$ DFR $\Leftrightarrow \tau < 1$
Gamma (α,θ)	$0 \Leftrightarrow \alpha > 1$ $\frac{1}{\theta} \Leftrightarrow \alpha = 1$ $\infty \Leftrightarrow \alpha < 1$	$\frac{1}{\theta}$	IFR $\Leftrightarrow \alpha > 1$ constant $\Leftrightarrow \alpha = 1$ DFR $\Leftrightarrow \alpha < 1$
Normal (μ,σ)	$\frac{\phi(\frac{-\mu}{\sigma})}{N(\frac{-\mu}{\sigma})}$	∞	IFR

Table 4.5: Failure Rate Analysis for Two-Parameter Distributions

Value-at-Risk:

$$S_V(x) = g_V(S(x)) = \begin{cases} 0 & \text{if} \quad x > x_\alpha \\ 1 & \text{otherwise} \end{cases} \tag{4.72}$$

Conditional Tail Expectation:

$$S_C(x) = g_C(S(x)) = \begin{cases} \frac{S(x)}{1-\alpha} & \text{if} \quad x > x_\alpha \\ 1 & \text{otherwise} \end{cases} \tag{4.73}$$

Proportional Hazard:

$$S_P(x) = g_P(S(x)) = S(x)^{\frac{1}{\gamma}} \tag{4.74}$$

Dual Power:

$$S_D(x) = g_D(S(x)) = 1 - (1 - S(x))^{\kappa} \tag{4.75}$$

Beta:

$$S_B(x) = g_B(S(x)) = \int_0^{S(x)} \frac{1}{\beta(\frac{1}{\gamma}, \kappa)} t^{\frac{1}{\gamma}-1}(1-t)^{\kappa} dt \tag{4.76}$$

In terms of the failure rate, $\lambda(x)$, the distorted failure rates are:

Value-at-Risk:

$$\lambda_V(x) = -\frac{d}{dx}\log S_V(x) = \begin{cases} 0 & if \quad x \neq x_\alpha \\ undefined & otherwise \end{cases} \tag{4.77}$$

Conditional Tail Expectation:

$$\lambda_C(x) = -\frac{d}{dx}\log S_C(x) = \begin{cases} 0 & if \quad x < x_\alpha \\ \lambda(x) & otherwise \end{cases} \tag{4.78}$$

Proportional Hazard:

$$\lambda_P(x) = -\frac{d}{dx}\log S_P(x) = \frac{\lambda(x)}{\gamma} \tag{4.79}$$

Dual Power:

$$\lambda_D(x) = -\frac{d}{dx}\log S_D(x) = \frac{\kappa F(x)^{\kappa-1} f(x)}{1 - F(x)^{\kappa}} \tag{4.80}$$

Beta:

$$\lambda_B(x) = -\frac{d}{dx}\log S_B(x) = -\frac{d}{dx}\log\left(\int_0^{S(x)} \frac{1}{\beta(\frac{1}{\gamma},\kappa)} t^{\frac{1}{\gamma}-1}(1-t)^{\kappa-1}dt\right) \quad (4.81)$$

$$= \frac{\frac{1}{\beta(\frac{1}{\gamma},\kappa)} S(x)^{\frac{1}{\gamma}}(1-S(x))^{\kappa-1}}{S_\beta(x)}\lambda(x) \quad (4.82)$$

Under the percentile-VaR distortion, there is no rate of change of the failure rate. Under the CTE distortion, the failure rate in the tail is the same as the original failure rate. Under the dual-power distortion, the failure rate gets smaller, and the rate of change has the same sign in the tail; however, $\lambda_D(0) = 0$, and for DFR distributions, the distorted failure rate increases to a maximum, and then decreases to the same limit as the original failure rate. For the PH-distortion, the failure rate gets smaller, and the rate of change has the same sign. Under the beta distortion, the failure rate is smaller than the failure rate under the PH-distortion, and the rate of change has the same sign in the tail; however, $\lambda_\beta(0) = 0$, and for DFR distributions, the distorted failure rate increases to a maximum, bounded by the PH-distorted rate, and then decreases to the same limit as the PH-distorted failure rate. The maximum failure rate for the beta distortion failure rates depends on the distribution of the underlying risk.

Using the PH-distortion, the failure rate is multiplied by a positive constant $\frac{1}{\gamma}$. The derivative of the distorted failure rate maintains its same sign. Thus, a distribution that had an IFR still has an IFR, only the failure rate is smaller and increasing at a slower rate, creating a risk loading for an IFR distribution. For a DFR distribution, even though the derivative of the distorted failure rate decreases, it remains negative, and the failure rate is proportionately smaller, creating a risk loading for a DFR distribution.

Proposition 4.3.1 *For any risk X, with ddf $S(x)$ and failure rate $\lambda(x)$, if X has an increasing failure rate (IFR) in the tail, or $\frac{d}{dx}\lambda(x) > 0$ for some x, then there exists some ξ such that for any $x > \xi$, $\frac{d}{dx}\lambda_\beta(x) > 0$.*

Corollary 4.3.1 *If X has a decreasing failure rate in the tail, or $\frac{d}{dx}\lambda(x) < 0$ for some x, then there exists some ξ such that for any $x > \xi$, $\frac{d}{dx}\lambda_\beta(x) < 0$.*

This indicates that the sign of the rate of change of the failure rate is preserved in the tail, under coherent beta distortions.

Proof: For the PH-distortion, the risk adjusted failure rate is $\lambda_P(x) = \frac{\lambda(x)}{\gamma}$. Thus, the PH-distortion maintains the sign of the rate of change of the failure rate of the original risk. Based on this, we show that the beta distorted failure rate approaches the PH-distortion failure rate with the same slope as $x \to \infty$.

By definition

$$\lambda_\beta(x) = -\frac{d}{dx}\log\int_0^{S(x)} \frac{1}{\beta(\frac{1}{\gamma},\kappa)} t^{\frac{1}{\gamma}-1}(1-t)^{\kappa-1}dt \tag{4.83}$$

$$= -\frac{d}{dx}\log S_\beta(x) \tag{4.84}$$

For smooth, continuous $S(x)$ with smooth continuous derivatives

$$\lim_{x\to\infty}\frac{d}{dx}\lambda_\beta(x) = \lim_{x\to\infty}\frac{d}{dx} - \frac{1}{S_\beta(x)}\frac{d}{dx}S_\beta(x) \tag{4.85}$$

$$= \lim_{x\to\infty}\frac{d}{dx} - \frac{1}{S_\beta(x)}\frac{1}{\beta(\frac{1}{\gamma},\kappa)}S(x)^{\frac{1}{\gamma}-1}(1-S(x))^{\kappa-1}(-f(x)) \tag{4.86}$$

$$= \frac{d}{dx}\lim_{x\to\infty}\frac{1}{\beta(\frac{1}{\gamma},\kappa)}\frac{S(x)^{\frac{1}{\gamma}-1}(1-S(x))^{\kappa-1}f(x)}{S_\beta(x)} \to \frac{0}{0}, \tag{4.87}$$

by l'Hôpital,

$$\begin{aligned}
&= \frac{d}{dx}\lim_{x\to\infty}(S(x)^{\frac{1}{\gamma}-1}(1-S(x))^{\kappa-1}f(x))^{-1} \\
&\quad \Bigg\{(k-1)S(x)^{\frac{1}{\gamma}-1}(1-S(x))^{\kappa-2}f(x)^2 \qquad (4.88) \\
&\quad +(1-S(x))^{\kappa-1}\left(f'(x)S(x)^{\frac{1}{\gamma}-1} - (\frac{1}{\gamma}-1)S(x)^{\frac{1}{\gamma}-2}f(x)^2\right)\Bigg\} \\
&= -\frac{d}{dx}\lim_{x\to\infty}\frac{(\kappa-1)f(x)}{1-S(x)} - \frac{(\frac{1}{\gamma}-1)f(x)}{S(x)} + \frac{f'(x)}{f(x)} \qquad (4.89) \\
&= \frac{d}{dx}\lim_{x\to\infty}(\frac{1}{\gamma}-1)\lambda(x) + \frac{f'(x)}{f(x)} \qquad (4.90) \\
&= \frac{d}{dx}\lim_{x\to\infty}\frac{\lambda(x)}{\gamma} = \lim_{x\to\infty}\frac{d}{dx}\frac{\lambda(x)}{\gamma} \qquad (4.91) \\
&= \lim_{x\to\infty}\frac{d}{dx}\lambda_P(x) \qquad (4.92)
\end{aligned}$$

This also implies that

$$\lim_{x\to\infty} [\lambda_\beta(x) - \lambda_P(x)] = 0. \tag{4.93}$$

Thus, the beta distorted failure rate approaches the same failure rate as the PH-distortion with the same slope in the tail, and thus maintains the sign of the rate of change of the failure rate in the tail. □

Since $\lim_{x\to\infty} \lambda(x) = \lim_{x\to\infty} \frac{1}{e(x)}$, we also have that

$$\lim_{x\to\infty} e_\beta(x) = \lim_{x\to\infty} e_P(x). \tag{4.94}$$

In terms of the mean excess loss,

$$e(x) = \int_x^\infty \frac{S(t)}{S(x)} dt, \tag{4.95}$$

the distorted MELs are:

Value-at-Risk:

$$e_V(x) = \int_x^\infty \frac{S_V(t)}{S_V(x)} dt = \begin{cases} x_\alpha - x & \text{if} \quad x < x_\alpha \\ 0 & \text{otherwise} \end{cases} \tag{4.96}$$

Conditional Tail Expectation:

$$e_C(x) = \int_x^\infty \frac{S_C(t)}{S_C(x)} dt = \begin{cases} e(x) & \text{if} \quad x > x_\alpha \\ x_\alpha - x + e(x_\alpha) & \text{otherwise} \end{cases} \tag{4.97}$$

Proportional Hazard:

$$e_P(x) = \int_x^\infty \frac{S_P(t)}{S_P(x)} dt = \int_x^\infty \left(\frac{S(t)}{S(x)}\right)^{\frac{1}{\gamma}} dt \tag{4.98}$$

Dual Power:

$$e_D(x) = \int_x^\infty \frac{S_D(t)}{S_D(x)} dt = \int_x^\infty \frac{1-(1-S(t))^\kappa}{1-(1-S(x))^\kappa} dt \tag{4.99}$$

Beta:

$$e_B(x) = \int_x^\infty \frac{S_B(t)}{S_B(x)} dt = \int_x^\infty \frac{\int_0^{S_B(t)} \frac{1}{\beta(\frac{1}{\gamma},\kappa}) t^{\frac{1}{\gamma}-1}(1-t)^{\kappa-1} dt}{\int_0^{S_B(x)} \frac{1}{\beta(\frac{1}{\gamma},\kappa}) t^{\frac{1}{\gamma}-1}(1-t)^{\kappa-1} dt} dt \tag{4.100}$$

If the failure rate is monotone, then the direction of the rate of change in the MRL is already known. Most of the distributions we consider have a monotone increasing or a monotone decreasing failure rate, dependent on the parameters, and since it is much more difficult to simplify the MRL under each of these transformations, we predominantly use results pertaining to the failure rate. However, if $\lambda(x)$ and $e(x)$ are known, for some loss distribution X, then the rate of change of the MRL is equal to,

$$\frac{d}{dx} e(x) = \lambda(x) e(x) - 1. \tag{4.101}$$

To compare the tails of two distributions, a simple comparison can be done by considering the limit of the ratio of their ddfs.

$$\text{If } \lim_{t\to\infty} \frac{S_X(t)}{S_Y(t)} \to \begin{cases} 0 & \Longrightarrow X \quad \text{has a heavier tail} \\ \text{constant} & \Longrightarrow X, Y \quad \text{have proportionate tails} \\ \infty & \Longrightarrow Y \quad \text{has a heavier tail} \end{cases} \tag{4.102}$$

4.3.2 Weibull Tails

The Weibull distribution has pdf:

$$f(x) = \frac{\tau \left(\frac{x}{\theta}\right)^{\tau} e^{-\left(\frac{x}{\theta}\right)^{\tau}}}{x} \tag{4.103}$$

and decumulative distribution function:

$$S(x) = e^{-\left(\frac{x}{\theta}\right)^{\tau}}. \tag{4.104}$$

The failure rate for the Weibull distribution is:

$$\lambda(x) = \left(\frac{x}{\theta}\right)^{\tau} \tag{4.105}$$

and so,

$$\frac{d}{dx}\lambda(x) = \frac{1}{\theta^{\tau}}\tau(\tau-1)x^{\tau-2}. \tag{4.106}$$

This implies that the Weibull distribution has an IFR for $\tau \geq 1$, and a DFR for $\tau \leq 1$. This also implies that the Weibull distribution has a DMRL for $\tau \geq 1$, and an IMRL for $\tau \leq 1$. The density and failure rate functions of two Weibull distributions are illustrated in Figure 4.4. One distribution has light tails, $\tau = 10$ and kurtosis $= 3.5701$, and the other has heavy tails $\tau = 0.5$ and kurtosis $= 87.72$. Figure 4.5 illustrates how the distortion functions distort the failure rates for the light-tailed and heavy-tailed Weibull distributions. In the light-tailed example, the original distribution has an IFR, and all of the distorted failure rates are also increasing. Comparing this with the heavy-tailed distribution, the original distribution has a DFR. Under the PH-distortion, the distribution is still DFR; however for the dual power and the beta distortions, the failure rate first increases and then decreases. The dual-power distorted failure rate approaches the original failure rate in the tail, and the beta distorted failure rate approaches the PH-distorted failure rate.

The Weibull distribution is a two-parameter distribution. As such, we can compare it to the four two-parameter distributions used in Chapter 2. If we choose $\tau = 0.802412$ and $\theta = 2.653522$, the Weibull distribution has a mean of 3 and a 95th percentile of 10.415.

Table 4.6 compares some of the characteristics of each of these distributions. The Pareto, Weibull and Gamma have monotone decreasing failure rates, and thus also have IMRLs. The lognormal distribution has an increasing then a decreasing failure rate, and the normal distribution has a monotone increasing failure rate.

The distorted failure rates for the Pareto and lognormal distributions are

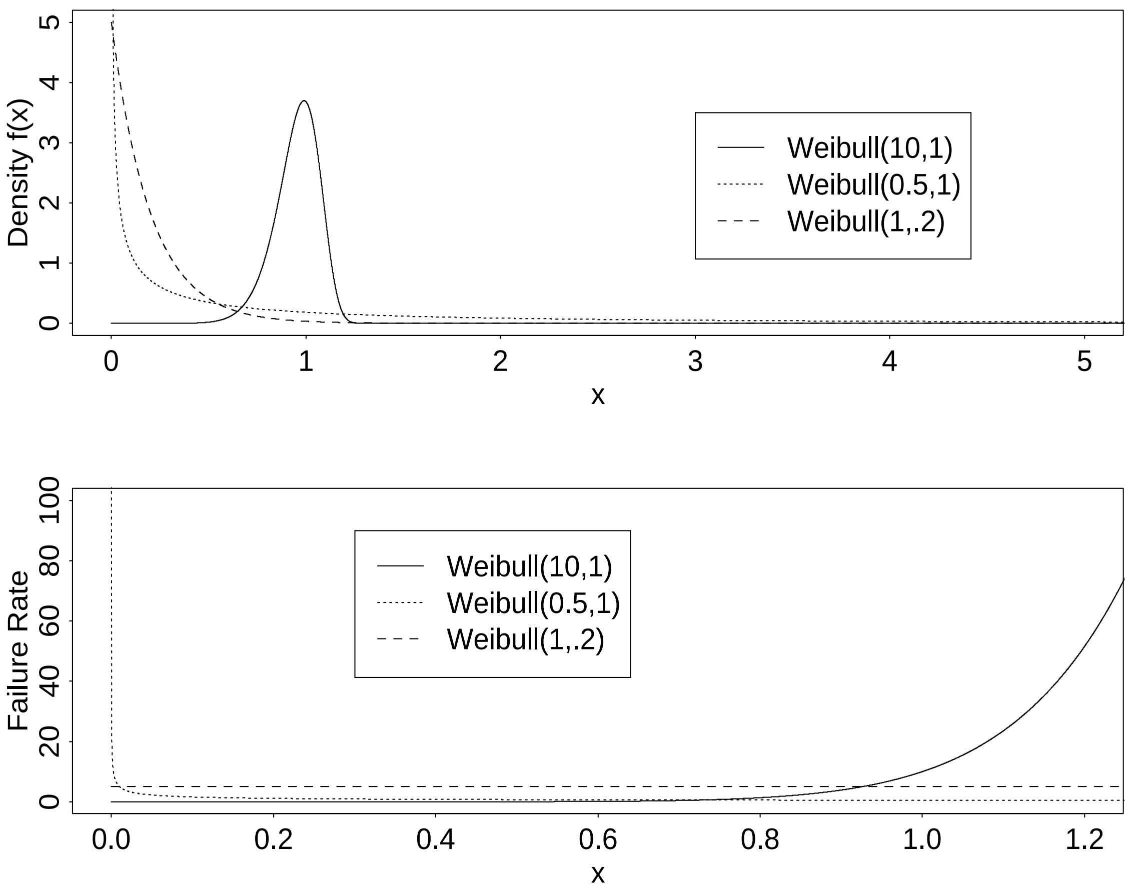

Figure 4.4: Weibull Density and Failure Rate Functions

illustrated in Figure 4.6 and for the normal and gamma distributions in Figure 4.7. If $\tau = 0.5$ and $\theta = 1.5$, the Weibull distribution has a mean of 3 and a variance of 45 and can be compared to the four two-parameter distributions with equated mean and variance. Table 4.7 compares some of the characteristics of each of these distributions.

From Table 4.7, we see that the normal distribution has a kurtosis independent of its parameters. The kurtosis for the Pareto distribution does not exist, as the

Distribution			Kurtosis	$\lambda(\mathbf{0})$	$\lim_{\mathbf{x}\to\infty} \lambda(\mathbf{x})$	FR
Pareto	$\alpha = 2.5$	$\theta = 4.5$	∞	0.556	0	DFR
Lognormal	$\mu = 0.403$	$\sigma = 1.180$	437.6	0	0	neither
Weibull	$\tau = 0.803$	$\theta = 2.654$	15.61	∞	0	DFR
Gamma	$\alpha = 0.663$	$\theta = 4.527$	12.05	∞	0.221	DFR
Normal	$\mu = 3$	$\sigma = 4.508$	3.0	0.095	∞	IFR

Table 4.6: Tail Statistics for Distributions with Mean=3, $x_{95} = 10.415$

Distribution			Kurtosis	$\lambda(\mathbf{0})$	$\lim_{\mathbf{x}\to\infty} \lambda(\mathbf{x})$	FR
Pareto	$\alpha = 2.5$	$\theta = 4.5$	∞	0.556	0	DFR
Lognormal	$\mu = 0.203$	$\sigma = 1.339$	1834.	0	0	neither
Weibull	$\tau = 0.5$	$\theta = 1.5$	87.72	∞	0	DFR
Gamma	$\alpha = 0.2$	$\theta = 15$	33	∞	0.067	DFR
Normal	$\mu = 3$	$\sigma = 6.708$	3.0	0.080	∞	IFR

Table 4.7: Tail Statistics for Distributions with Mean= 3, and Variance= 45

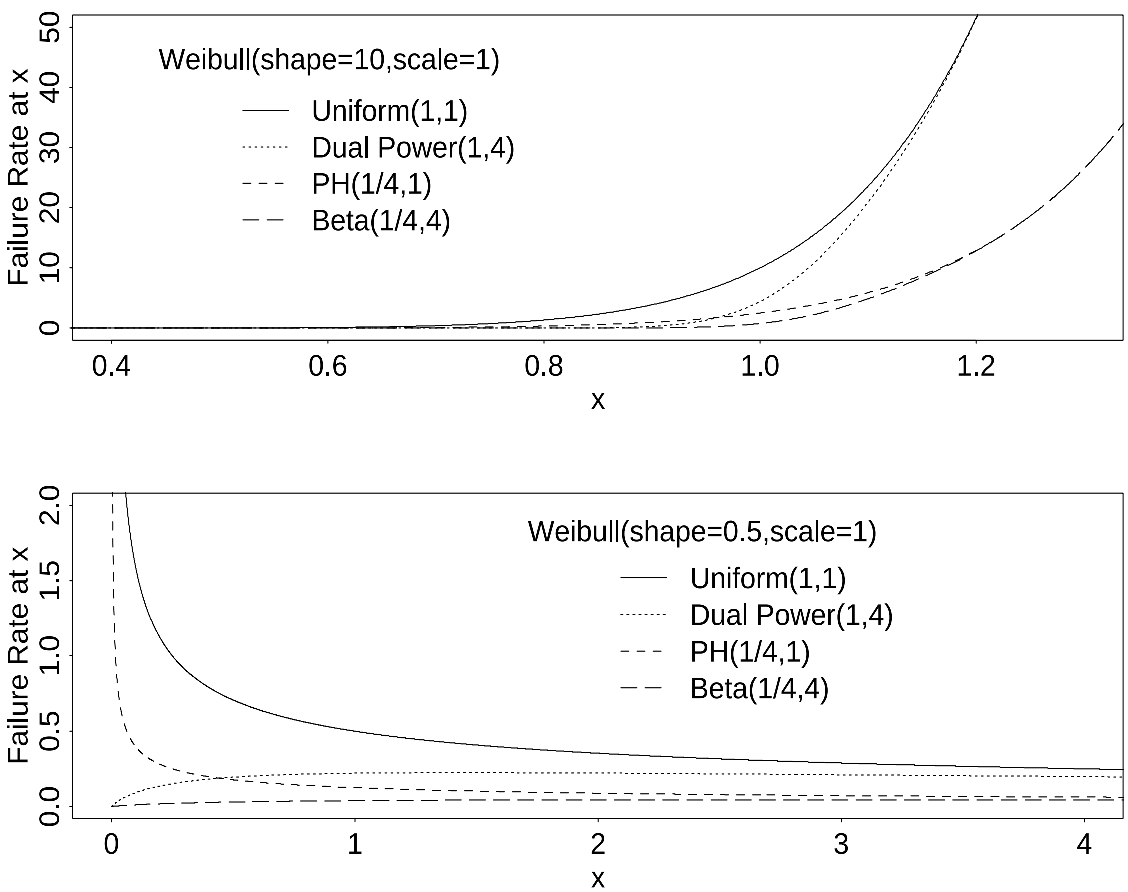

Figure 4.5: Distorted Failure Rates for Heavy and Light-Tailed Weibull Distributions

parameter α is less than 4. The other three distributions have significantly larger kurtosis than the distributions with equated 95th percentile, and the limit of the failure rate for the gamma distribution is significantly smaller. The direction of the failure rates for the five distributions is unchanged. The Pareto distribution is identical to the one used in the previous illustration, and the distorted failure rates for the Pareto can be seen in Figure 4.6. The distorted failure rates for the Weibull and lognormal distributions are illustrated in Figure 4.8 and for the normal and gamma distributions in Figure 4.9.

Although the beta distortion transforms a DFR distribution into a distribution with increasing then decreasing failure rates, the tail of the distorted distribution maintains the decreasing failure rate and is still bounded by the failure rate of the PH-distortion.

All the graphical illustrations in this section have used the parameters $\kappa = \gamma = 4$. Choosing larger parameters would decrease the failure rate further. If $\kappa < \gamma$ the failure rates are closer to the PH-distortion failure rate curve, and the maximum failure rate is increased. When $\kappa > \gamma$ the failure rates are even smaller, the maximum failure rate increases, and the failure rate in the tail still approaches the PH-distorted failure rates.

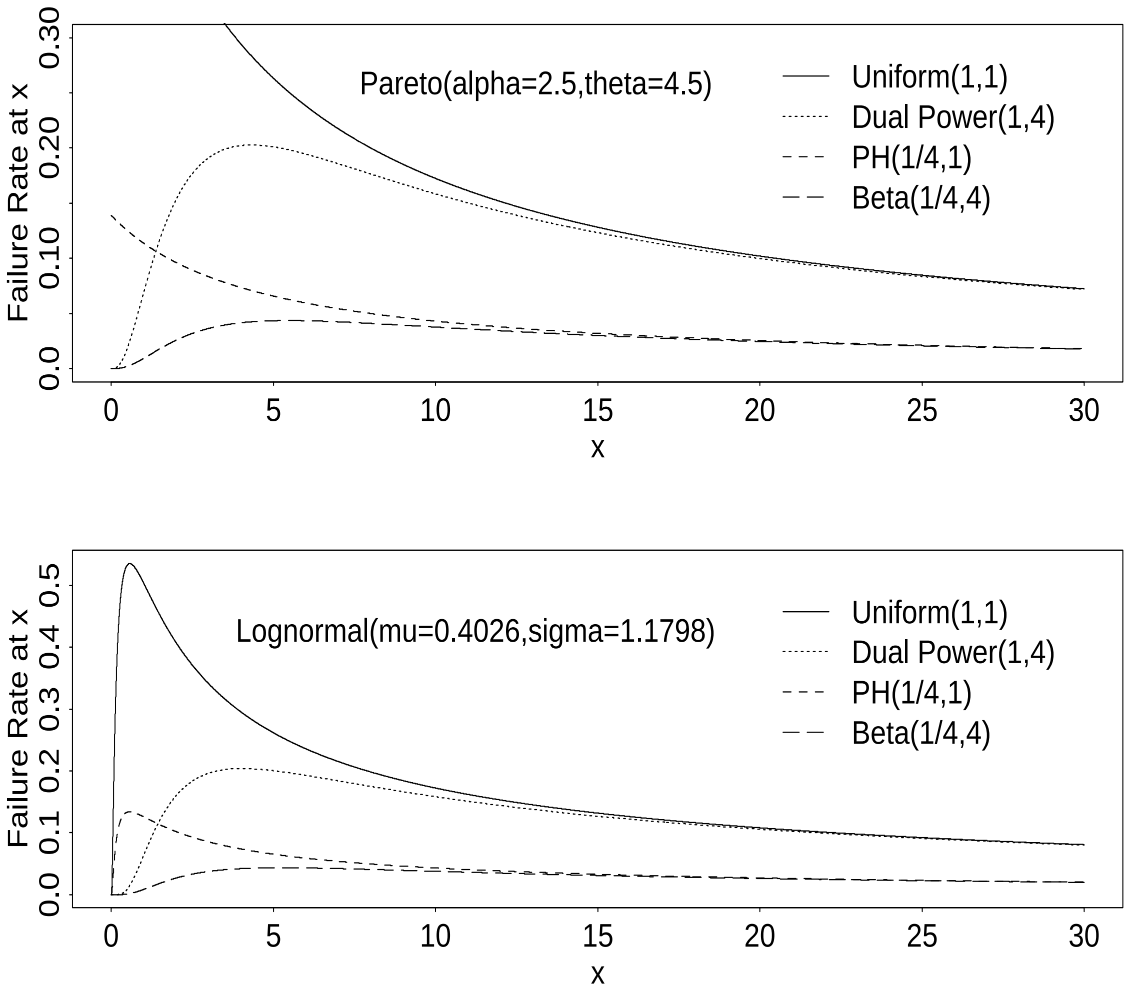

Figure 4.6: Distorted Failure Rates for a Pareto and Lognormal Distributions with $x_{95} = 10.415$

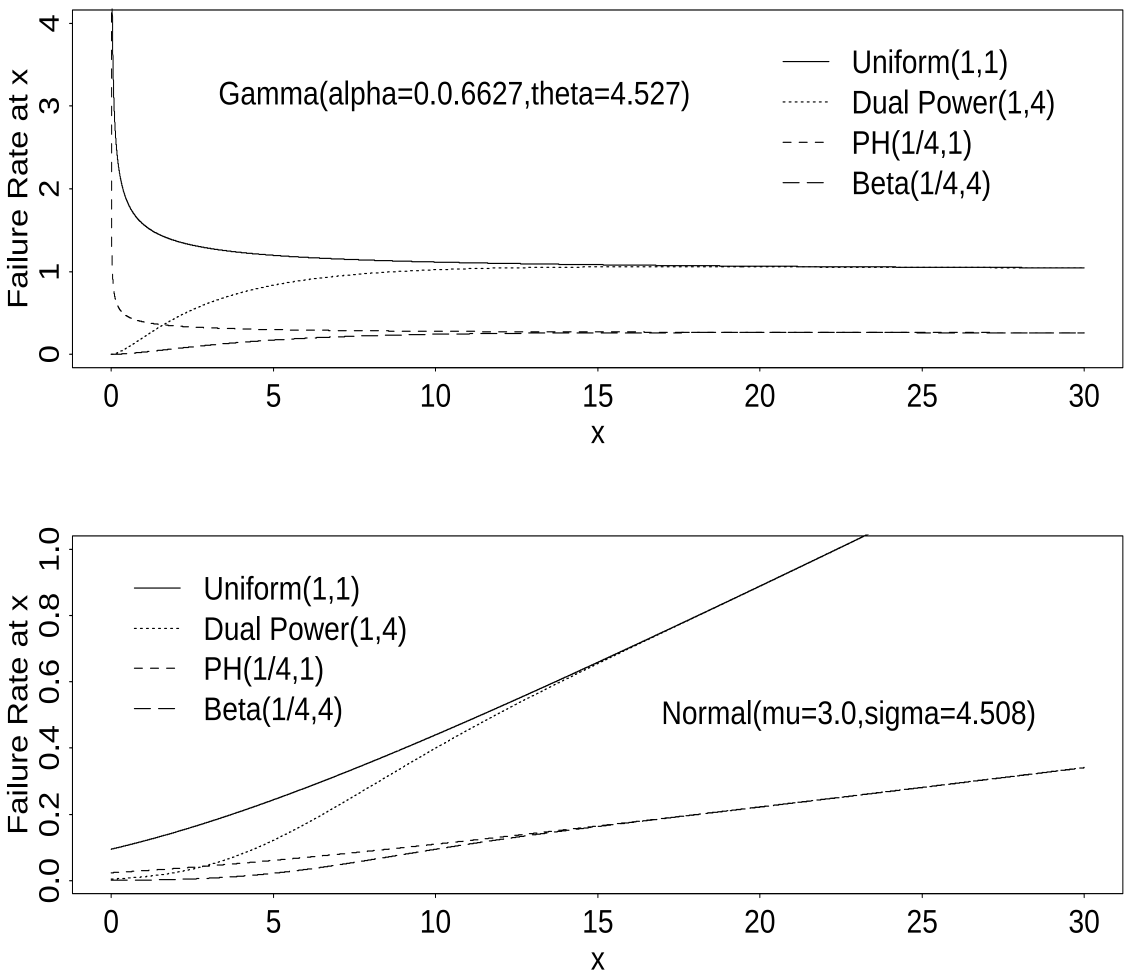

Figure 4.7: Distorted Failure Rates for a Gamma and Normal Distributions with $x_{95} = 10.415$

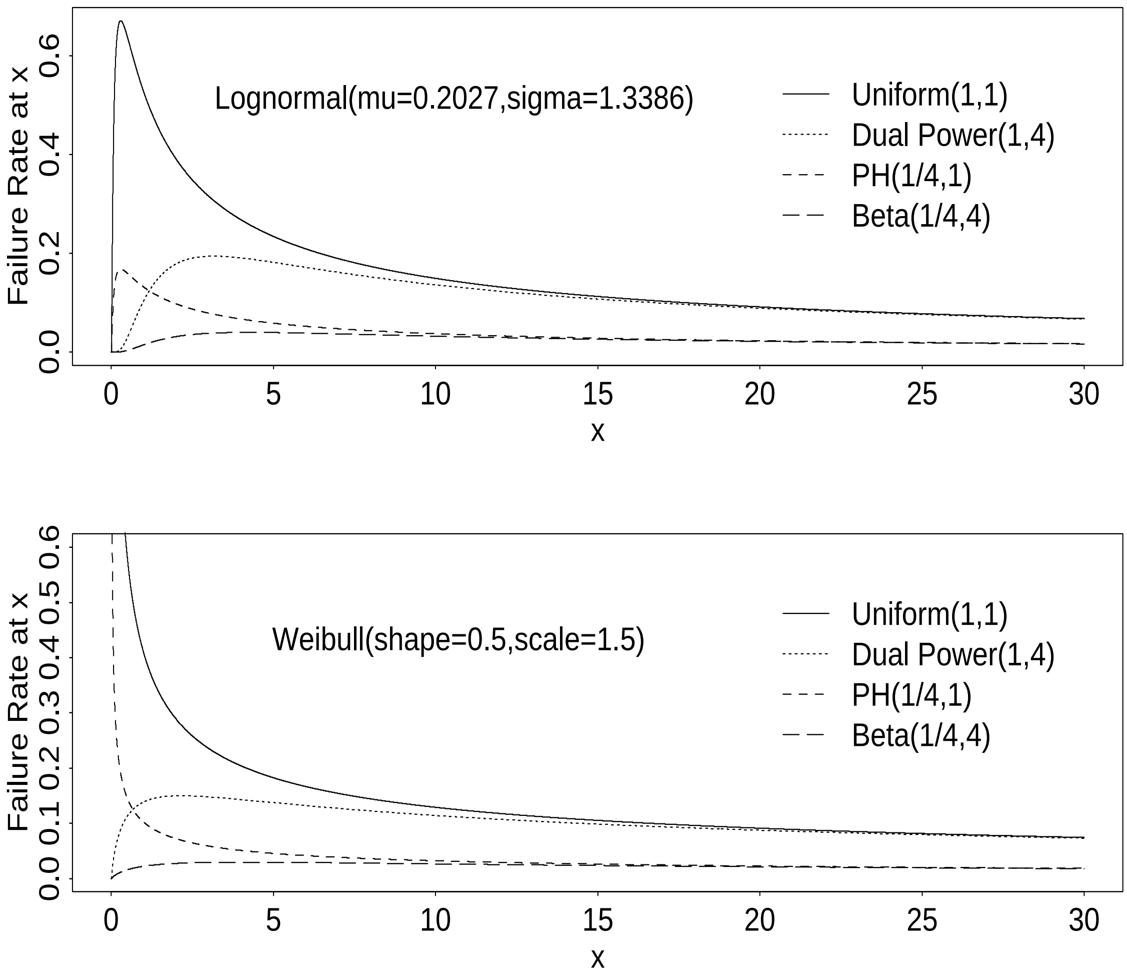

Figure 4.8: Distorted Failure Rates for a Pareto and Lognormal Distributions with Variance = 45

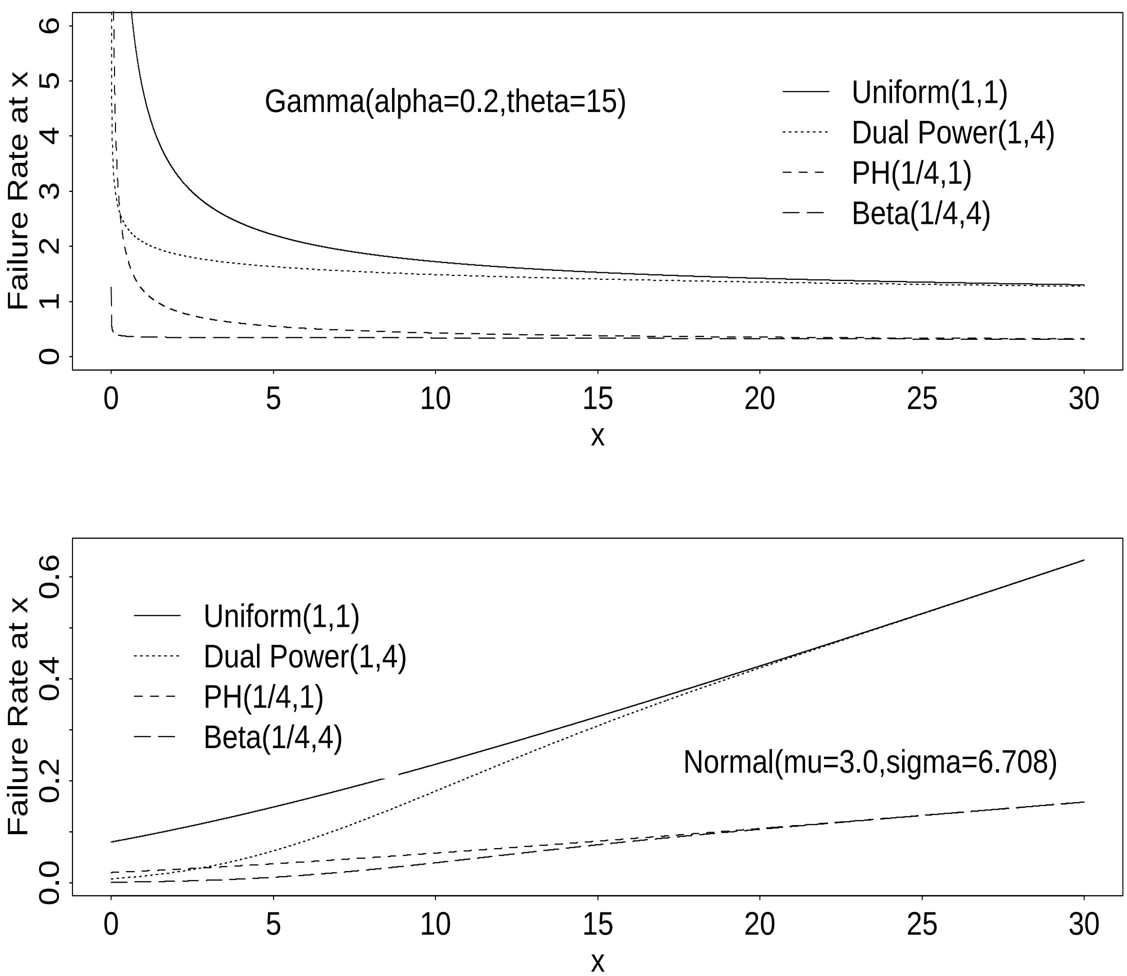

Figure 4.9: Distorted Failure Rates for a Gamma and Normal Distributions with Variance = 45

Chapter 5

Illustrations

In this chapter, the ideas discussed in Chapters 3 and 4 are applied to problems of topical interest that can contribute to our understanding of these risk measures and their applications. In the first illustration a capital requirement is determined for a maturity guarantee. Maturity guarantees often arise in segregated funds as an added option that limits the investor's risk. The second illustration calculates the capital requirement for an annuity rate guarantee, which is often applied in the UK to retirement benefits. An annuity rate guarantee is an option to transfer a lump sum benefit into an annuity benefit at a guaranteed rate. After analyzing these portfolios separately, we combine them using different weighting schemes, and discuss two simple methods for allocating risk capital requirements between the two portfolios. Lastly, we revisit the maturity guarantee and consider the implications hedging has on the capital requirement for the maturity guarantee.

To accurately value any of these portfolios, a model must be chosen for the interest rate process, the mortality process, lapses, inflation, fund accumulation, etc. In order to keep these illustrations simple and transparent, we have chosen to model the fund accumulation process using the Wilkie model and the lognormal distribution. Using the Wilkie model we also consider modeling the long-term interest rate. Some fairly strong assumptions are made about mortality and lapse. In practice, provisions must be made for risk factors not taken into account here. These illustrations are kept fairly simple as they are meant only to demonstrate the capital requirement risk measures considered in the previous chapters.

5.1 The Wilkie Model

In order to model interest rates and stock fund accumulation we have chosen to use the Wilkie stochastic asset model (Wilkie, 1995). This model consists of four discrete annual time series which provide annual values for the retail prices index (a function of the instantaneous rate of inflation), the index of gross equity dividends, the current running gross dividend yield, and the gross yield on consols (equivalent to the long-term interest rates). This model integrates interest rates with accumulation rates for equities. The most recent full Wilkie model now includes simulations for wage inflation, short term interest rates, exchange rates and property accumulation. This model has met with a great deal of support and is often used for applications in actuarial science. However, the Wilkie model has been in the public domain since 1986 and has been subject to scrutiny. Some criticism has been expressed over its appropriateness with respect to short-term forecasting and the subjective decisions Wilkie made in developing the model (Huber, 1997). For our purposes, we feel the Wilkie model is sufficient. We are not exploring the adequacy of this specific model, but using this methodology for our illustrations. As well, our illustrations involve assets with a maturity of 10 years or more, so the short-term forecasting problems are not as relevant. The purpose of these illustrations is to apply our risk measures to topical problems, to determine relationships between portfolios, and to discuss methods for allocating capital.

The time series model was fitted to Canadian data from 1923-1993 (Wilkie, 1995). The model is used to simulate time series using the following equations:

1. Retail Price Index, $Q(t)$

$$\log(\frac{Q(t)}{Q(t-1)}) = 0.034 + 0.64(\log(\frac{Q(t-1)}{Q(t-2)}) - 0.034) + 0.032Z_1(t) \tag{5.1}$$

where $I(t) = \log(\frac{Q(t)}{Q(t-1)})$ is the instantaneous rate of inflation.

> The instantaneous rate of inflation for year t, $I(t)$ is an autoregressive process and is calculated as the mean rate, 0.034, plus a fraction, 0.64, of previous year's deviation from the mean plus a random $Normal(0, 0.032^2)$ term, denoted by $0.032Z_1(t)$.

2. An index of gross equity dividends, $D(t)$

$$\log(\frac{D(t)}{D(t-1)}) = 0.19DM(t) + 0.81\log(\frac{Q(t)}{Q(t-1)}) + 0.001 \tag{5.2}$$

$$- 0.0209Z_2(t-1)) + 0.0406Z_3(t-1) + 0.07Z_3(t))$$

where, $DM(t) = 0.26\log(\frac{Q(t)}{Q(t-1)}) + 0.74DM(t-1)$,

and $Z_3(t), Z_3(t-1)$ and $DM(t-1)$ are independent standard normal random variables.

The change in the logarithm of the dividend index is equal to a function of the current and past instantaneous rates of inflation plus the mean real dividend growth, 0.001, plus an influence, -0.0209, from last years random effect on dividend yield plus and influence, 0.0406, from this years random effect on dividend yield plus a random $Normal(0, 0.07^2)$ term, denoted by $0.07Z_3(t)$.

3. Gross dividend Yield, $Y(t)$

$$\log(Y(t)) = 1.17\log(\frac{Q(t)}{Q(t-1)}) + \log(0.0375) + YN(t) \tag{5.3}$$

where, $YN(t) = 0.7YN(t-1) + 0.19Z_2(t)$,

and $Z_2(t)$ is an independent standard normal random variable.

The logarithm of the dividend yield is equal to its mean value ($\log(0.0375)$) plus an adjustment (1.17) from last year's inflation, plus an influence (0.7) from all the previous random effects on dividend yield plus a random $Normal(0, 0.19^2)$ term, denoted by $0.19Z_2(t)$.

From the dividend index and the dividend yield, we can obtain a price index of ordinary shares as,

$$P(t) = \frac{D(t)}{Y(t)} \tag{5.4}$$

and an accumulation factor for the value of ordinary shares,

$$J(t-1) = \frac{P(t) + D(t)}{P(t-1)}. \tag{5.5}$$

4. Gross yield on Consols (Long-term Interest Rates), $C(t)$

$$\begin{aligned} C(t) &= CM(t) + 0.037exp(CN(t)) && (5.6)\\ CM(t) &= 0.04\log(\frac{Q(t)}{Q(t-1)}) + (0.96)CM(t-1) && (5.7)\\ CN(t) &= 0.95CN(t-1) + 0.019Z_2(t) + 0.185Z_4(t), && (5.8) \end{aligned}$$

where $Z_4(t)$ is an independent standard normal random variable.

The long-term interest rate is calculated in two parts, the first is an allowance for expected future inflation ($CM(t)$), and the second is a real yield, where the logarithm of the real yield is equal to its mean ($\log 0.037$) plus a fraction (0.95) of its past random influences, plus an influence (0.019) from the current random effect from dividend yield plus a random $Normal(0, 0.185^2)$ term, denoted by $0.185Z_4(t)$.

The initial values required for these calculations are:

$$\begin{aligned} CM(0) &= DM(0) = I(0) = 0.034 && (5.9)\\ Y(0) &= 0.0375e^{(0.03978)} && (5.10)\\ Q(0) &= D(0) = J(0) = 1.0 && (5.11)\\ CN(0) &= YN(0) = 0.0 && (5.12) \end{aligned}$$

5.2 Maturity Guarantees for Segregated Funds

In a segregated fund, premiums are invested for the insured by the insurer in the assets of the fund. A management charge is deducted from the fund each year, and benefits are paid out on death or on the maturity date. The death benefit and the maturity benefit vary depending on the success of the fund. It is possible that these benefits will be less than the sum of the premiums invested. Adding a maturity guarantee onto the segregated fund guarantees a minimum benefit for the insured, at an additional cost to the insurer. This cost is often offset by a higher management charge. Our objective is to use risk measures to provide guidance on the amount of capital required to support this segregated fund business. In order to determine the capital needed to cover a maturity guarantee, we consider four different maturity guarantees on single premium 10-year policies based on the same segregated fund. The first guarantees 75% of the premium with a 1% percent management charge, the second guarantees 75% with a 2% management charge. The third guarantees 100% of the premium with a 1% percent management charge, and the fourth guarantees 100% with a 2% management charge. Hardy (1998) and Boyle and Hardy (1998) considered this problem using percentile-VaR to calculate the reserve for the maturity guarantee.

The segregated fund policies are assumed to be single premium policies, purchased at the same time and held to maturity, so that there are no intermediate entrants, lapses or withdrawals. We also assume that there are no reinvestments at maturity, so that the full cost of the guarantee is felt at the end of the tenth year. The capital required is determined at this date and we discount using a risk free instantaneous interest rate which determines the amount of capital invested today in risk free assets so that it will accumulate to the amount of the required capital at the end of 10 years. We also make the assumption that there is no mortality risk. Mortality risk is partially diversifiable. We assume that any undiversifiable mortality risk is funded by a portion of the management charge.

Intuitively, the 75% guarantee with a 1% management charge is the least likely to exercise the maturity guarantee. If the segregated fund earns a zero return, and the 1% management charge is deducted each year, the resulting fund is still more than 90% of the original premium and the guarantee is not needed. Conversely, the 100% guarantee with a 2% management charge is the most likely to exercise the maturity guarantee as the fund must earn more than 2% each year on average to maintain its initial value.

Two models were used to determine the fund accumulation factors. The first model for the accumulation factors uses the Wilkie investment model with parameters fitted from Canadian data from Wilkie (1995), described in the previous section. Using the Wilkie model, the fund value at maturity was simulated 150,000 times. For each simulation if the simulated accumulated fund was less than the original premium, the difference between the original premium and the accumulated fund is discounted using a constant instantaneous rate risk free of interest of 6% to the beginning of the 10 year period, otherwise the cost of the maturity guarantee is zero. Since our risk measures are scalar multiplicative (A3) discounting the cost of the maturity guarantee before applying the risk measure is equivalent to applying the risk measure and then discounting. The censored loss random variable for the maturity guarantee is:

$$L_{MG}^{Wilkie} = \max\left[0, g - (1-m)^n \prod_{t=1}^{n} J(t)\right] e^{-0.06n}, \tag{5.13}$$

where m is the yearly management charge and $n = 10$. Using 150,000 simulations, an empirical distribution function is obtained. The beta risk measure was applied to this empirical distribution function to determine the capital requirements for the risk.

The second method assumes a lognormal distribution for a one-year accumulation function $(1 + i_t)$, with parameters consistent with the Wilkie model $(\mu = 0.081, \sigma = 0.17)$. We also assume that the accumulation factors are independent for each year. The accumulated value of the fund, less any management charges m, has a lognormal distribution, so that $(1 + i)^n(1 - m)^n$ is distributed $lognormal(n(\mu + \log(1 - m)), n\sigma)$. The discounted loss random variable is

$$L_{MG}^{Logn} = [g - (1+i)^n(1-m)^n]^+ e^{-n(\delta)} \tag{5.14}$$

where $[x]^+ = \max(x, 0)$ and δ is the risk free instantaneous interest rate. We calculate the expected cost and use numerical integration to determine the capital requirements for the risk, based on the beta risk measure. Figure 5.1 compares the maturity guarantee loss distributions under the Wilkie and lognormal assumptions. The similarity of these curves suggest that the lognormal closely approximates the Wilkie simulation model for fund accumulation.

To illustrate this problem more carefully, consider a segregated fund for a 10-year single premium endowment policy with a 75% maturity guarantee. At the end of the 10 years, the policy pays the market value of the segregated fund, as long as it is worth at least 75% of the value of the insured's premiums. The loss distribution for this guarantee is bounded, the maximum payout (loss) at maturity in this case is 75% of the invested premiums, the minimum payout is 0. If the fund value at maturity is only 65% of the premiums, the guarantee will be exercised and the cost of the guarantee is 10% (75%-65%) of the invested premium. If the fund value is in excess of 75% of invested premiums, the investor gets the full fund value and the guarantee is not exercised.

In comparison, a 100% maturity guarantee on the same segregated fund has a maximum payout at maturity of 100% of the invested premium, and there is a greater probability of having a payout under this guarantee. It is obvious to see that the 100% maturity guarantee carries a higher risk for the insurer than the 75% guarantee. Conversely, the management charge on the segregated fund is deducted yearly, and reduces the accumulated value of the fund. If the fund has a 2% management charge, the average yearly accumulation rate on the fund must be more than 2.04% in order for the fund to maintain its initial investment. The lower the management charge, the more likely the fund is to maintain its initial investment, and the fund will have a lower probability of needing a payout under the maturity guarantee. Thus the 100% maturity guarantee with a 2% management charge is the most risky of the four combinations, while the 75% maturity guarantee with a 1% management charge is the least risky.

To compare the other two maturity guarantees, we consider the accumulated cost of the management charge and compare this with the guarantees. Assuming no fund accumulations, a 2% management charge for each of the 10 years reduces \$1 of initial investment to $\$1(1-0.02)^{10} = \$0.817$. Since 81.7% is greater than 75%, for a 75% maturity guarantee, the average accumulation rate can be less than zero and no payout is required from the guarantee. A 1% management charge reduces \$1 of initial investment to $\$1(1-0.01)^{10} = \$0.904$. Since 90.4% is less than 100%, for a 100% maturity guarantee, the average accumulation rate must be greater than 1 .01% for the accumulated fund to be greater than the initial investment. Thus a 100% maturity guarantee with a 1% management charge is more risky than the 75% maturity guarantee with a 2% management charge. If the guarantee rates were 85% and 90% instead of 75% and 100% respectively, it would be more difficult to compare these guarantees.

Using the Wilkie investment model to simulate interest rates over the 10 year duration of the segregated fund, and a 6% risk free discount rate, the expected value of the initial cost of the 100% maturity guarantee with a 2% management charge is \$1.44. The discounted value of the maximum loss is \$54.88. For the 75%

	100% Maturity Guarantee		**75% Maturity Guarantee**	
Mgmt. Charge:	2%	1%	2%	1%
Mean	1.437	0.938	0.295	0.175
Std. Dev.	4.707	3.769	1.784	1.348
Maximum	54.88	54.88	41.16	41.16
Kurtosis	19.13	29.18	67.43	110.8
DP 19	15.29	11.70	4.559	2.915
PH 19	34.52	32.98	22.46	21.17
PH 4	14.82	13.02	7.496	6.405
Beta($\frac{1}{2}$, 2)	9.143	7.299	3.452	2.613
Beta ($\frac{1}{4}$, 4)	**21.94**	**19.45**	**11.36**	**9.75**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	23.50	21.02	12.44	10.78
VaR 95%	12.27	7.713	**0.000**	**0.000**
VaR 99%	24.38	21.12	10.66	7.402
CTE 90%	13.72	9.373	2.953	1.752
CTE 95%	19.52	15.75	5.906	3.503

Table 5.1: \$100, 10-Year, Maturity Guarantee Capital Requirements, using the Wilkie Model

maturity guarantee with a 1% management charge the expected value of the initial cost is \$0.18 and the discounted value of the maximum loss is \$41.16. Table 5.1 compares the risk measures discussed in the previous chapters, applied to all four combinations of guarantees and management charges. Table 5.1 illustrates two of the problems with VaR. Based on a 95% VaR, the 75% maturity guarantee does not seem to have any risk even though the mean loss is greater than zero (negative risk loading). As well, VaR 95% does not differentiate between the segregated fund with a 1% management charge and the fund with a 2% management charge (superadditivity). Even though the 75% maturity guarantee is less risky than the

100% guarantee, the 75% guarantee has a much higher kurtosis. The kurtosis of each maturity guarantee is more easily compared when the mean of each portfolio is equated. However, for censored distributions, the kurtosis of the distribution does not directly relate to the thickness of the one-sided tail.

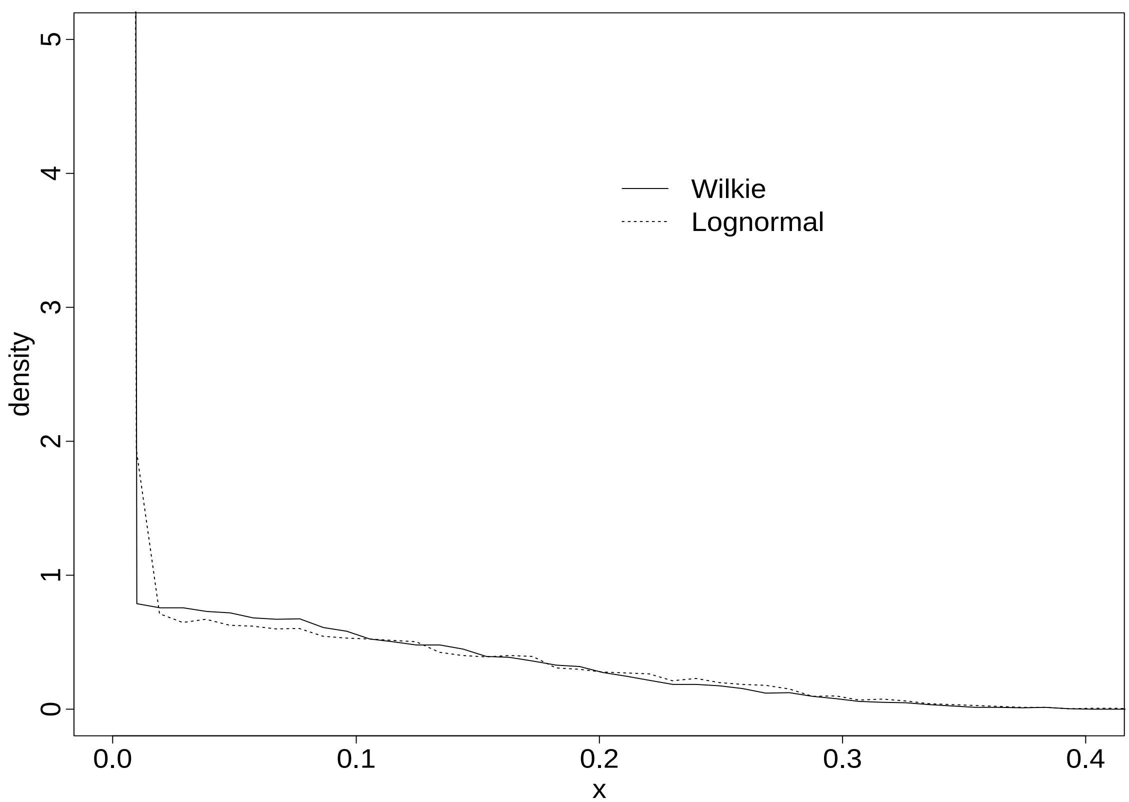

Figure 5.1: Density Functions for the Discounted Loss of a \$1 Maturity Guarantee (2% Management Charge)

As seen in the previous chapter, there are many combinations of beta parameters that can be used. Figure 5.2 shows the effect of varying each parameter for the 100% maturity guarantee with a 2% management charge, whereas Figure 5.3 shows the same effect for the 75% maturity guarantee with a 1% management charge.

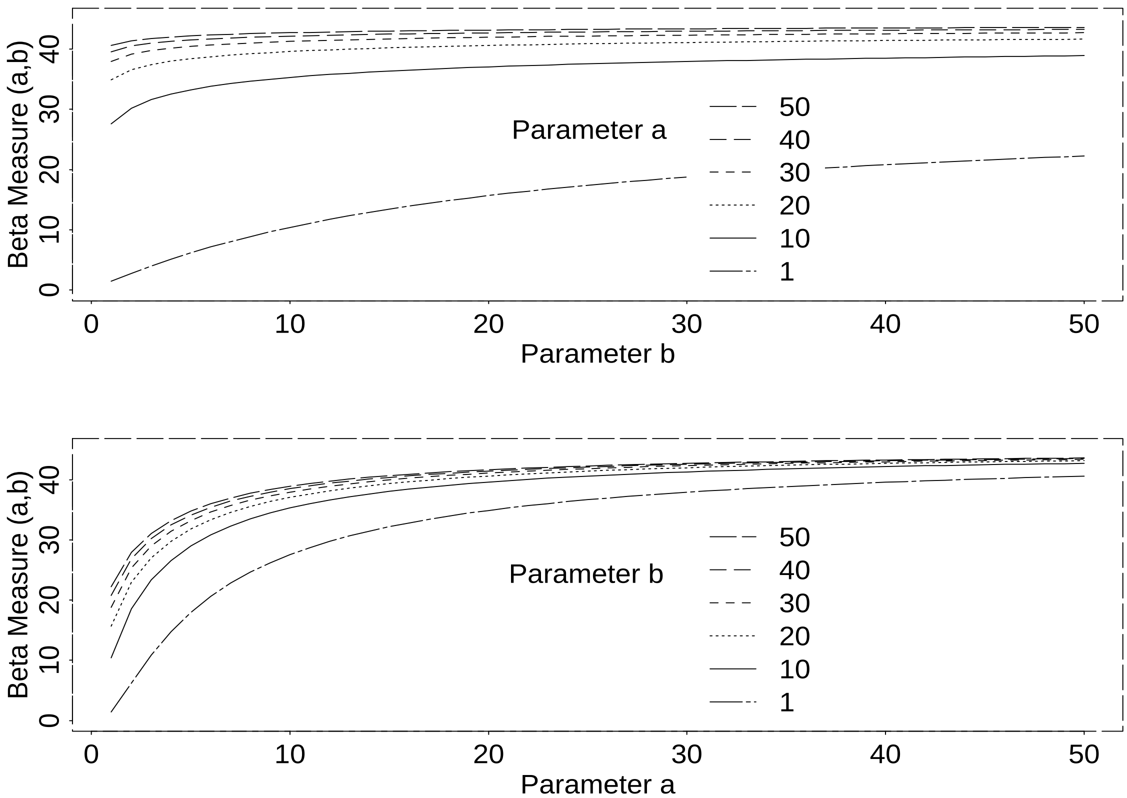

Figure 5.2: Required Capital using the Beta Measures for 100% Guarantees with 2% Management Charge, Assuming Wilkie Investment Model

From these graphs, it is evident that the risk measure is more sensitive to changes in the PH-distortion parameter γ. However, the difference in risk between the 75% maturity guarantee and the 100% maturity guarantee is very slight, and since these liabilities are similar in nature, we consider the 100% maturity guarantee with a 2% management charge as our base maturity guarantee for the rest of this thesis.

The density of the Wilkie simulated maturity guarantee loss distribution, censored at zero, is shown in Figure 5.1. The maturity guarantee loss distributions are far from being normally distributed. Losses for the 100% maturity guarantees

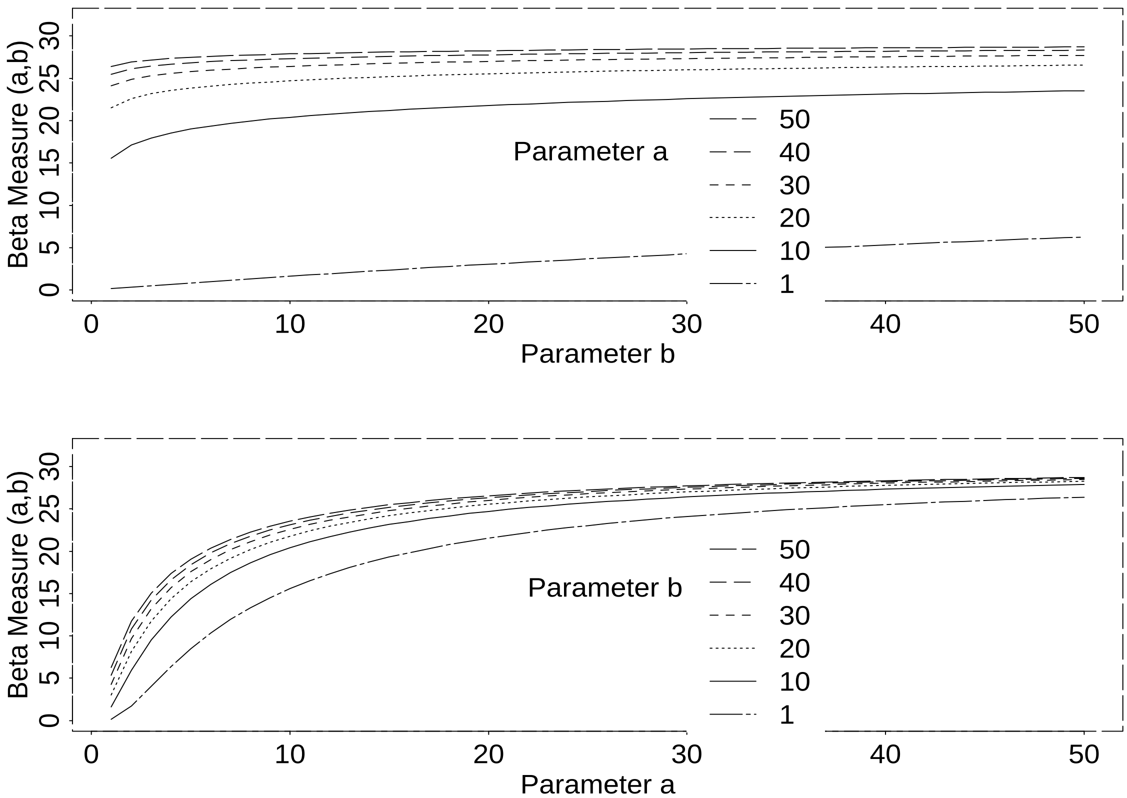

Figure 5.3: Required Capital using the Beta Measures for 75% Guarantees with 1% Management Charge, Assuming Wilkie Investment Model

occur with roughly 10% probability. For the 75% guarantee, losses occur with less than a 5% probability. Losses, in all cases, have an upper bound, but using the Wilkie model or any other reasonable model, it is practically impossible to reach that upper bound. To compare our simulated solutions with a parametric model, we assume that the underlying annual returns are independent and lognormal with parameters that are consistent with the returns from the Wilkie model, $\mu = 0.081$ and $\sigma = 0.17$. Since there is a closed form solution for the beta risk measure, numerical integration instead of simulation is used to calculate the risk measure, and the results are shown in Table 5.2, for 100% and 75% maturity guarantees.

Again we see the same problems with 95% VaR, however all the other risk measures rank the risks consistently. In all cases Beta($\frac{1}{4}$, 4) is larger than 95% VaR and CTE 95%, and is larger than 99% VaR for the 75% maturity guarantee. Since the beta risk measure is coherent, these results lend support to the Beta($\frac{1}{4}$, 4) risk measure.

	100% Maturity Guarantee		**75% Maturity Guarantee**	
Mgmt. Charge:	2%	1%	2%	1%
Mean	1.538(5.037)	1.052	0.365	0.231
Maximum	54.88	54.88	41.16	41.16
DP 19	16.42	12.95	5.502	3.745
PH 19	38.59	37.49	26.56	25.69
PH 4	15.83	14.14	8.465	7.411
Beta($\frac{1}{2}$, 2)	9.782	8.010	4.002	3.137
Beta($\frac{1}{4}$, 4)	**23.43**	**21.10**	**12.77**	**11.22**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	25.10	22.79	14.02	12.44
VaR 95%	13.25	8.800	0.000	0.000
VaR 99%	26.02	22.94	12.30	9.215
CTE 90%	14.76	11.25	3.652	2.30
CTE 95%	21.02	17.40	7.305	4.61

Table 5.2: \$100, 10-Year, Maturity Guarantee Capital Requirements, using the Lognormal Model

Figure 5.1 compares the densities of the maturity guarantee loss distribution using the Wilkie and the lognormal models from 5.1 and 5.2.

The upper bound on the loss for the maturity guarantee, and the 5-10% probability that a loss occurs illustrates how the beta risk measure determine the capital requirement for one type of portfolio. Thus we would like to compare this illustration with another portfolio that has a larger probability of a loss and no bound on the loss.

5.3 Annuity Rate Guarantees

Annuity rate guarantees were popular when interest rates were high, and thus annuity rates were low. An annuity rate guarantee allows the policyholder to convert their insurance policy or endowment benefit into a life annuity at a prespecified rate and a prespecified time. For instance, if a policy holder had a 10-year endowment insurance, with an annuity rate guarantee of 9 at maturity; then for every 9 dollars of endowment insurance that matures at the end of the 10th year, the policyholder has the option to purchase a 1 dollar per year life annuity or $\frac{1}{s_{\overline{12}|\frac{i_{12}}{12}}}$ per month paid at the end of each month for as long as the policy holder lives. Again, if the insured carried whole life insurance with an option to convert at the end of the 15th year at an annuity rate of 8, then the accumulated value of the insurance at the end of the 15th year, could be used to purchase a life annuity at a rate of 8 units life insurance to 1 unit life annuity.

Unfortunately for the insurers, this product was used extensively in the UK as a “free” option when interest rates were high. Since then, interest rates have declined, life expectancy has increased, and the cost of this option is much greater than originally anticipated, and the valuation of these liabilities has become very topical in the UK. The value of an annuity rate guarantee depends on the accumulated value of the insurance fund or the amount of the benefit at the time of conversion, the interest rate at the time of conversion and beyond, as well as the mortality rate of the cohorts purchasing this insurance. This guarantee often was attached to retirement benefits, whereby a lump sum retirement benefit could be transformed into a life annuity, where the age at conversion was approximately 65 and life expectancy at age 65 was roughly 15 years.

In this illustration we assume that the annuity rate guarantee is added to a 10-year endowment insurance with a $1 single premium. The Wilkie investment model used for maturity guarantees is used to simulate the insurance fund accumulation rate, as well as the long-term interest rate at the end of the 10 years. This interest rate is used to determine the present value of the annuity at conversion in 10 years, which is then discounted 10 years at an instantaneous rate of 6%. In practice, each insured could be from a different cohort with different expected lifetimes; however, we have simplified this illustration by assuming that the life insured will live exactly 15 years after conversion instead of having a life expectancy of 15 years. Thus, a 15-year annuity certain is used to determine the present value of the converted annuity.

In calculating the risk of an annuity rate guarantee, the rate of the guarantee is determined at the inception date. At the conversion date the annuity rate guarantee will be exercised if an annuity for the beneficiary costs more than the guarantee rate. Thus, the lower the guarantee rate, the higher the expected cost of the guarantee. As well, the size of a loss is proportional to the size of the endowment fund. An endowment fund with a 1% management charge will accumulate to more than a fund with a 2% management charge. Thus, the risk associated with the annuity rate guarantee decreases as the management charge increases. The loss random variable for the annuity rate guarantee is

$$L_{AG}^{Wilkie} = (1-m)^n \prod_{t=1}^{n} J(t) \max\left[0, a_{\overline{180}|\frac{i_{12}}{12}} - Ag\right] e^{-0.06n}, \tag{5.15}$$

where Ag is the guaranteed annuity rate, and $\frac{i_{12}}{12} = (1+C(n))^{\frac{1}{12}} - 1$.

In order to identify the significance of the risk pertaining to the fund accumulation and the risk pertaining to the interest rate at the end of 10 years, this same annuity rate guarantee is simulated using the Wilkie accumulation factors, but using an assumed year 10 interest rate. This model is referred to as the Partial Wilkie model. The long-term interest rate chosen, $C(n) = 6.01\%$, gives the same expected cost for the guarantee as the long-term interest rate from the complete Wilkie investment simulation. Using this fixed rate, the present value of a 15-year annuity certain is 9.705 and the loss random variable for the Partial Wilkie model is

$$L_{AG}^{Partial} = (1-m)^n \prod_{t=1}^{n} J(t) \max\left[0, 9.705 - Ag\right] e^{-0.06n}. \tag{5.16}$$

Figure 5.4 compares the loss density for the annuity rate guarantee using the Partial Wilkie model (with an assumed annuity present value of 9.705) with the loss density for the annuity rate guarantee using the simulated annuity present values from the Wilkie model.

Comparing the Wilkie investment model assumptions with the Partial Wilkie model, the differences in the densities relate to the tenth year long-term interest

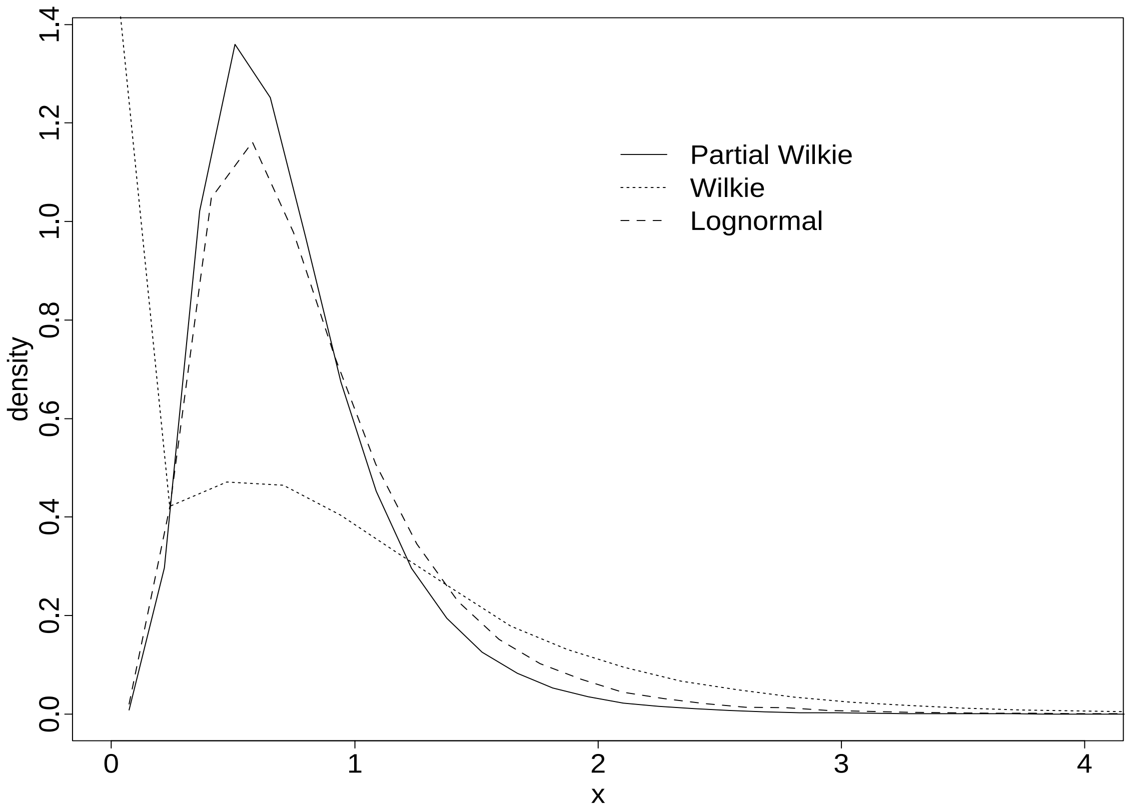

Figure 5.4: Density Functions for the Discounted Loss of an Annuity Rate Guarantee

rate. In practice, an assumed rate may be chosen to maintain the value of a specific risk measure other than the mean, in order to be more conservative.

The loss distribution using the complete Wilkie model is heavier tailed than Partial Wilkie with an assumed tenth year interest rate (see Tables 5.3 and 5.4). There is approximately half of the variability in the model when the tenth year interest rate is assumed. Even though the assumed interest rate guarantees a loss, the variability only relates to the accumulated size of the fund.

Using this same 10th year interest rate, we consider this problem using the

	2% Mgmt Charge	**1% Mgmt Charge**
Mean	0.749	0.829
Std. Dev.	0.806	0.892
Kurtosis	7.292	7.292
DP 19	2.688	2.975
PH 19	7.786	8.618
PH 4	3.005	3.326
Beta($\frac{1}{2}$, 2)	2.053	2.272
Beta($\frac{1}{4}$, 4)	**4.166**	**4.611**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	4.475	4.953
VaR 95%	2.294	2.539
VaR 99%	3.413	3.778
CTE 90%	2.513	2.781
CTE 95%	3.001	3.322

Table 5.3: Annuity Rate Guarantee Capital Requirements, using the complete Wilkie Model

lognormal assumption for the accumulation rate. Figure 5.4 compares the density of the annuity rate guarantee using the lognormal assumption and the same 10th year interest rate of 6.01%, with the Wilkie and Partial Wilkie models. The lognormal assumption gives results between the full Wilkie and the Partial Wilkie model. For the annuity rate guarantee, the lognormal model capital requirements are more sensitive to changes in γ than the Wilkie and Partial Wilkie models. This was not the case when comparing the requirements for the maturity guarantee.

Comparing the annuity rate guarantee to the maturity guarantee based on kurtosis, the loss distribution for the maturity guarantee seems to be much heavier tailed. However, using the beta distortion, the heavier tailed the distribution, the more sensitive the risk measure is to the first beta parameter, γ. Using the

	2% Mgmt Charge	**1% Mgmt Charge**
Mean	0.749	0.829
Std. Dev.	0.387	0.429
Kurtosis	9.598	9.598
DP 19	1.691	1.872
PH 19	4.892	5.415
PH 4	1.966	2.176
Beta($\frac{1}{2}$, 2)	1.548	1.548
Beta($\frac{1}{4}$, 4)	**2.594**	**2.871**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	2.780	3.077
VaR 95%	1.479	1.637
VaR 99%	2.063	2.283
CTE 90%	1.597	1.768
CTE 95%	1.850	2.048

Table 5.4: Annuity Rate Guarantee Capital Requirements, using the Partial Wilkie Model (assumed 10th year interest rate)

lognormal model for the maturity guarantee, the PH-distortion with $\gamma = 19$ gives values 2 to 3 times the values given by the PH-distortion with $\gamma = 4$. Whereas, for the annuity rate guarantee with the lognormal assumption, the PH-distortion with $\gamma = 19$ gives values 7 to 8 times the values given by the PH-distortion with $\gamma = 4$. Under the Wilkie and Partial Wilkie models, the PH-distortion with $\gamma = 19$ are only 2 to 3 times the values given by the PH-distortion with $\gamma = 4$, for both the annuity rate and maturity guarantees. Thus the lognormal assumption for the accumulation rates, creates a heavy-tailed loss distribution for the annuity rate guarantee which is not identified by the kurtosis. Thus the kurtosis may not be a reliable indicator of heavy-tails for non-symmetric censored random variables.

	2% Mgmt Charge	**1% Mgmt Charge**
Mean	0.822	0.909
Std. Dev.	0.480	0.526
Kurtosis	10.57	9.930
DP 19	2.000	2.214
PH 19	18.87	20.88
PH 4	2.515	2.784
Beta($\frac{1}{2}$, 2)	1.633	1.807
Beta($\frac{1}{4}$, 4)	**3.392**	**3.754**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	3.748	4.149
VaR 95%	1.722	1.907
VaR 99%	2.484	2.749
CTE 90%	1.879	2.079
CTE 95%	2.204	2.441

Table 5.5: Annuity Rate Guarantee Capital Requirements, using the Lognormal Model

5.4 Combining Portfolios

When combining portfolios or subsidiaries, and calculating required capital for the combined portfolio, one of the main problems is to determine how much of the required capital should be held by each portfolio. The allocation of capital requirements based on simple subjective methods can lead to a lot of controversy. However, a more accurate division of capital requirements can lead to onerous calculations. Thus we discuss two fairly simple methods of capital allocation; the first is marginal allocation and the second is a beta weighted allocation.

Marginal Capital Allocation: For measuring the risk of a financial conglomerate, where there is a large parent corporation and a number of small subsidiaries, one method to determine the allocation of the capital requirement is to use marginal allocation. First the capital requirement is determined for the parent corporation alone, then one subsidiary is added at a time, possibly in ascending order in terms of size of their portfolios, and the capital requirement is recalculated for the combined portfolio. The capital requirement for the subsidiary is the increase in the capital requirement when the subsidiary is added. This method is repeated for each successive subsidiary.

To illustrate, let X be the risk portfolio for the parent corporation, and let Y and Z be the risk portfolios for two subsidiaries. If $\rho(.)$ is the capital requirement risk measure, then assuming $\rho(Y) > \rho(Z)$, the capital requirement for the parent corporation is $\rho(X)$, the capital requirement for the subsidiary with portfolio Y is $\rho(X+Y) - \rho(X)$, and the capital requirement for the subsidiary with portfolio Z is $\rho(X+Y+Z) - \rho(X+Y)$.

This method is obviously order dependent and puts heavier capital requirements on the risks in the parent company and the subsidiaries that are added first. The sum of all the capital requirements add to the combined requirement $\rho(X+Y+Z)$.

Weighted Capital Allocation: The second method values each portfolio individually assuming some capital requirement measure $\rho(.)$, and also determines the capital requirement for the total combined portfolio. Then the total capital requirement is divided up based on the weighting of their individual measures.

To illustrate using risk portfolios X_1, X_2 and X_3, the combined portfolio has capital requirement $\rho(X_1 + X_2 + X_3)$. The individual portfolio risk measures are $\rho(X_1)$, $\rho(X_2)$ and $\rho(X_3)$ respectively, and the capital requirement for portfolio X_i is

$$\rho(X_1 + X_2 + X_3)\frac{\rho(X_i)}{\rho(X_1) + \rho(X_2) + \rho(X_3)} \tag{5.17}$$

where the sum of all the capital requirements adds to the combined requirement of $\rho(X_1 + X_2 + X_3)$.

To illustrate the two methods of allocating capital requirements among subportfolios, we combine the two portfolios used in the previous section. A maturity guarantee is combined with an annuity rate guarantee, where the accumulation of the underlying asset is based upon the same $lognormal(0.081, 0.17)$ fund accumulation model. Since maturity guarantees have a loss when the fund accumulation is low, and annuity rate guarantees losses are proportional to the fund accumulation factors, these products should provide some insight into combining negatively correlated portfolios. We also assume that the annuity has a certain 15 year term and is valued based on a constant year 10 long-term interest rate of 6.01%.

To start, consider a portfolio with a \$1 initial investment in each fund. Using a Beta$(\frac{1}{4}, 4)$ risk capital requirement from Table 5.6, the marginal capital allocation method requires that the annuity rate guarantee portfolio hold the entire capital requirement, 3.392. Using the proportional allocation method, the maturity guarantee portfolio requires 0.219 and the annuity rate guarantee portfolio would hold 3.173.

	Combined	**Maturity Guarantee**	**Annuity Guarantee**
Mean	0.837	0.015	0.822
Std. Dev.	0.460	0.050	0.480
Kurtosis	11.86	19.12	10.57
DP 19	2.000	0.164	2.000
PH 19	18.87	0.386	18.87
PH 4	2.519	0.158	2.515
Beta($\frac{1}{2}$, 2)	1.634	0.098	1.633
Beta($\frac{1}{4}$, 4)	**3.392**	**0.234**	**3.392**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	3.748	0.251	3.748
VaR 95%	1.723	0.132	1.722
VaR 99%	2.484	0.260	2.484
CTE 90%	1.879	0.148	1.879
CTE 95%	2.204	0.210	2.204

Table 5.6: Combined Portfolio Capital Requirements, using the Lognormal Model

Since both funds accumulate, but only the annuity rate guarantee liability increases with the size of the fund, assuming a \$1 initial investment in each fund leads to a much higher risk in the annuity rate guarantee than in the maturity guarantee. This is illustrated in Figure 5.5 which shows the densities for each portfolio and the density for the combined portfolio. In this example the annuity rate guarantee dominates the portfolio and has a much higher capital requirement; however, the proportional allocation method seems more fair than the marginal allocation method. Figure 5.6 illustrates the failure rates for the two portfolios. Both portfolios are heavy tailed, but the annuity rate guarantee has a much stronger influence on the combined portfolio.

To obtain a more illustrative comparison of the two portfolios, consider a large portfolio of maturity guarantees, so that the mean loss of the maturity guar-

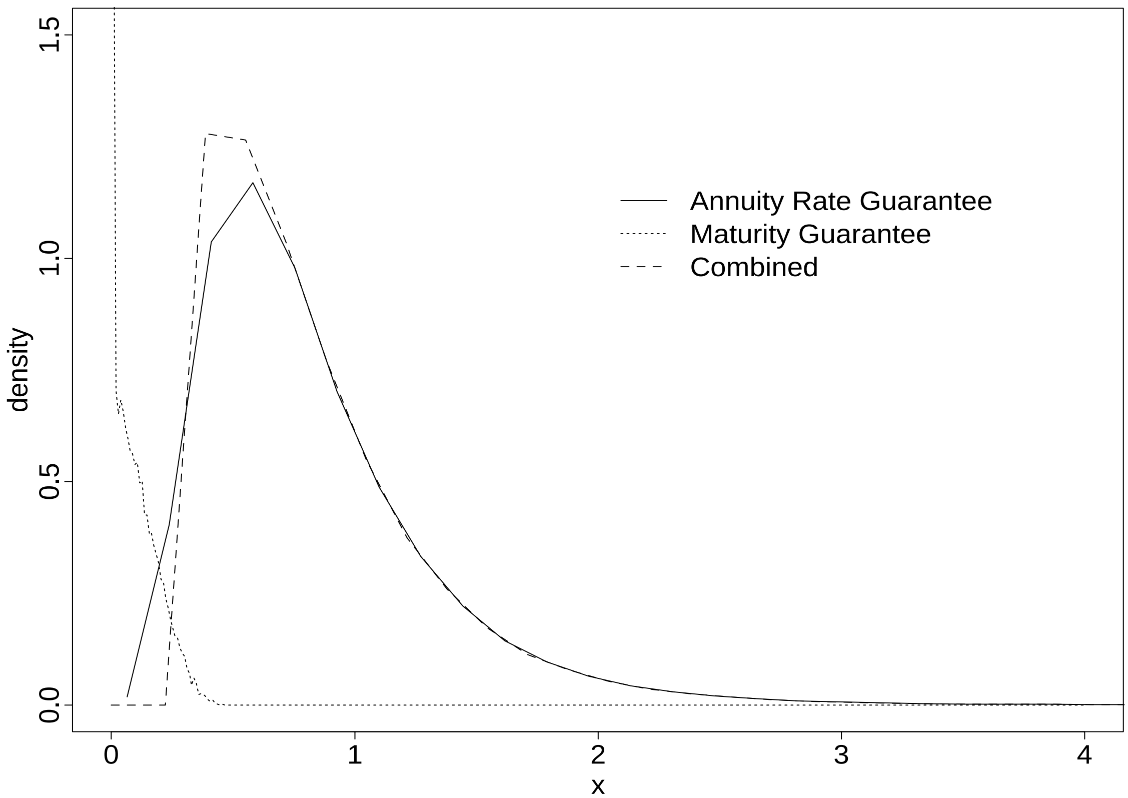

Figure 5.5: Combined Density Function (2% Management Charge), Assuming Lognormal Investment Model

antee portfolio equals that of the annuity guarantee. In effect, for every $1 initially invested in the annuity rate guarantee there is $53.4 invested in the maturity guarantees. Using a 2% management charge, the results are shown in Table 5.7, and a comparison of the individual densities with the combined portfolio density is in Figure 5.7.

In this mean equated portfolio, we have increased the size of the maturity guarantee position 53.4 times, and the maturity guarantee position dominates the portfolio. In terms of capital allocation, using the marginal allocation method with

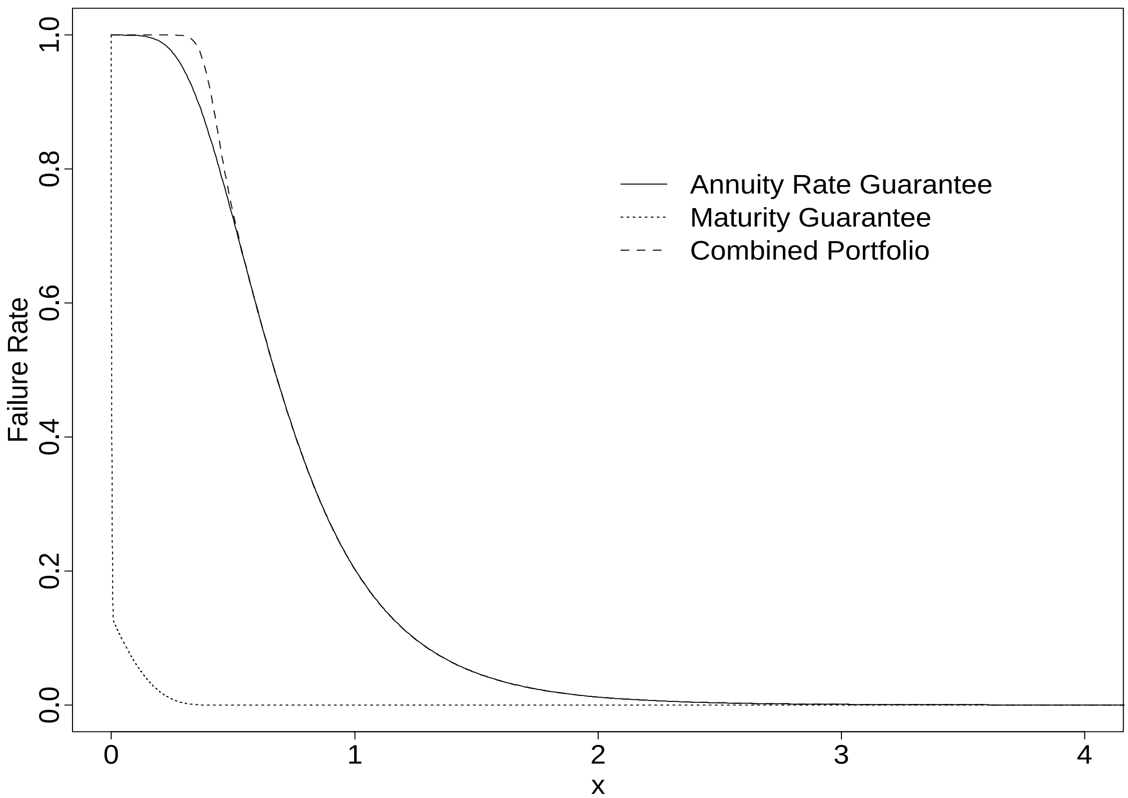

Figure 5.6: Failure Rates for the Combined Portfolio

expected loss to determine the order of inclusion, either portfolio could be added first. If the maturity guarantees portfolio is added first, it would have to hold 12.52 in required capital and the annuity rate guarantee portfolio would require capital of 0.437. If the annuity rate guarantee portfolio is added first, it would have 3.392 in required capital and the maturity guarantees portfolio would hold 9.560. Using the proportional allocation method the maturity guarantee portfolio would have 10.19 and the annuity rate guarantee portfolio would have 2.762.

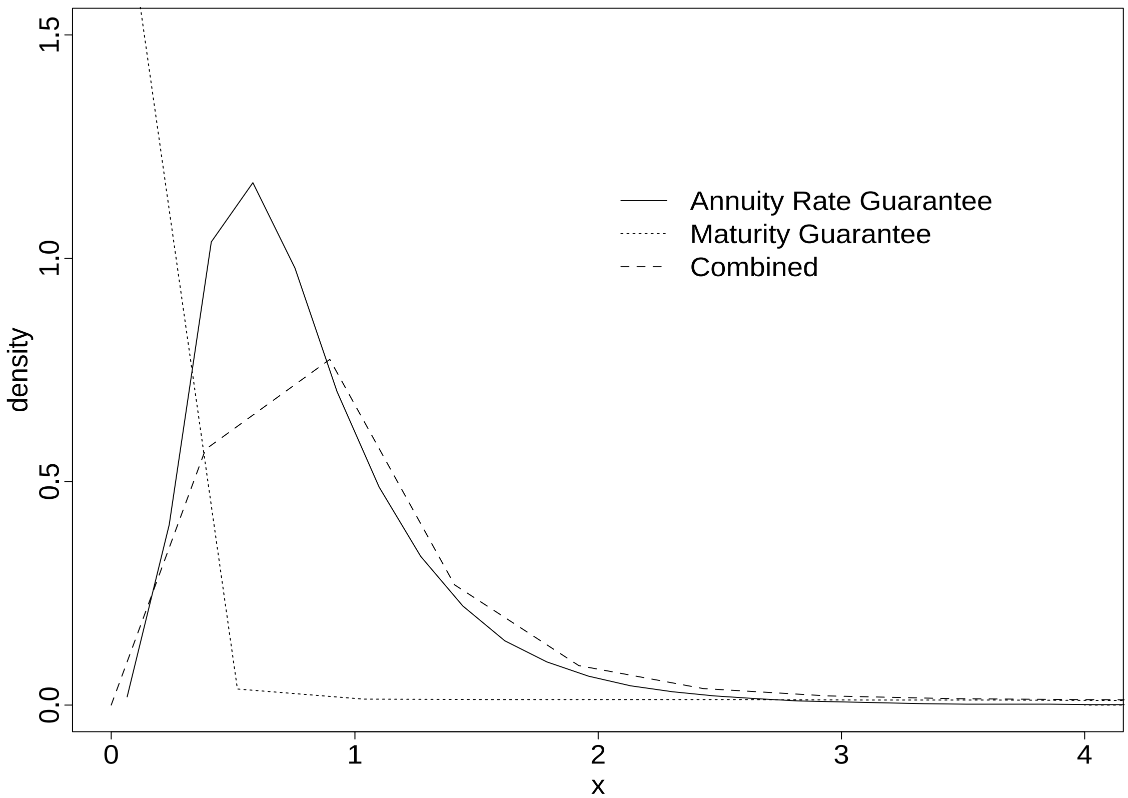

Figure 5.7: Mean Equated Combined Density (2% Management Charge), Assuming Lognormal Investment Model

To consider another weighting scheme, we equate the 95% VaR. In this case, for every \$1 initially invested in the annuity rate guarantee, invest \$13.0 in the maturity guarantees. The results can be seen in Table 5.8, a comparison of the individual densities with the combined portfolio density can be see in Figure 5.8, and a comparison of their failure rates in Figure 5.9. In all these combined portfolio analyses, Beta($\frac{1}{4}$, 4) is larger than 95% VaR and 95% CTE, and is often larger than 99% VaR.

	Combined	**Maturity Guarantee**	**Annuity Guarantee**
Mean	1.643	0.822	0.822
Std. Dev.	2.559	2.691	0.480
Kurtosis	19.29	19.12	10.57
DP 19	9.205	8.772	2.000
PH 19	26.80	20.62	18.87
PH 4	9.042	8.455	2.515
Beta($\frac{1}{2}$, 2)	5.981	5.226	1.633
Beta($\frac{1}{4}$, 4)	**12.95**	**12.52**	**3.392**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	13.83	13.41	3.748
VaR 95%	7.372	7.077	1.722
VaR 99%	14.11	13.90	2.484
CTE 90%	8.200	7.887	1.879
CTE 95%	11.47	11.23	2.204

Table 5.7: Mean Equated Combined Portfolio Capital Requirements, using the Lognormal Model

In terms of capital allocation, using marginal allocation, the annuity rate guarantee portfolio would be added first and have to hold 3.392, and the maturity guarantee portfolio would have to hold 0.667. Using the proportional allocation method, the annuity rate guarantee portfolio would have to hold 2.139 and the maturity rate guarantee portfolio would have 1.920 in required capital.

Considering the two allocation strategies, the marginal allocation method strongly favours smaller risks and does not allow the parent corporation any reduction in terms of capital requirements, for diversification through subsidiaries. The proportional method of allocation of capital requirements allows for hedging or diversification through subsidiaries and is also independent of the order of inclusion, thus it is the preferred approach.

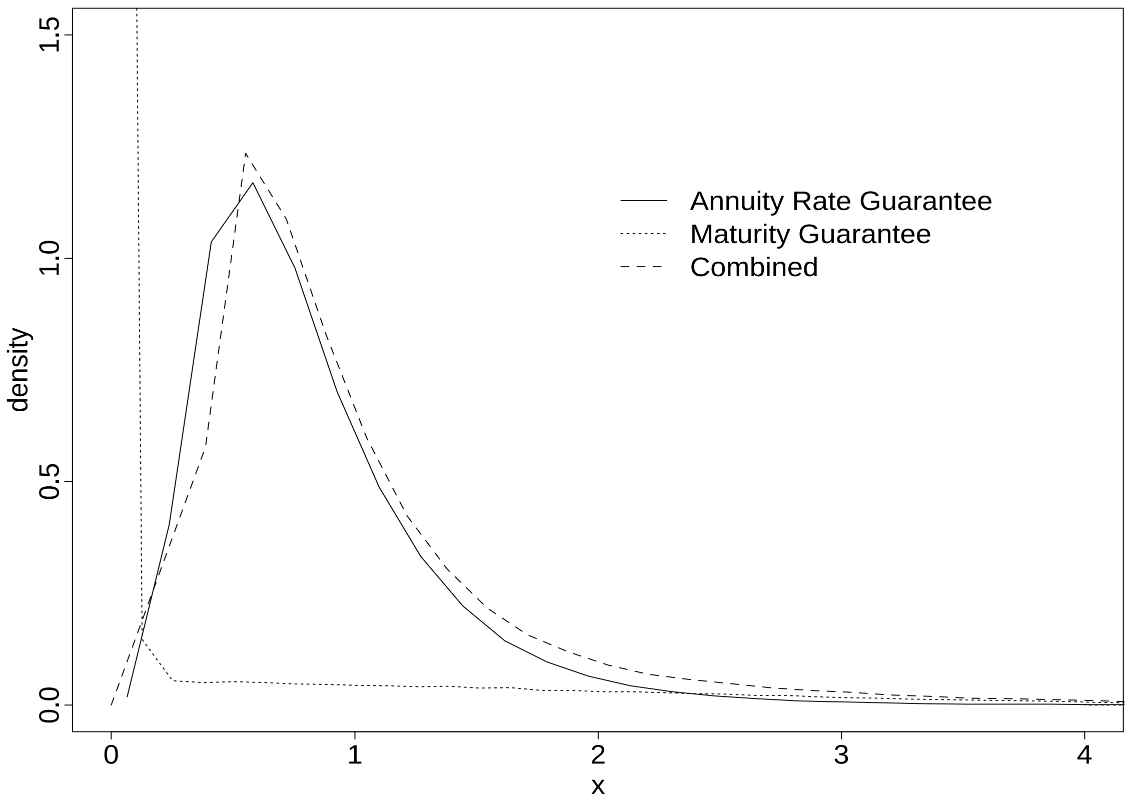

Figure 5.8: VaR Equated, Combined Density (2% Management Charge), Assuming Lognormal Investment Model

5.5 Hedging Error for Maturity Guarantees

Hardy (1998) uses percentile VaR to study hedging and reserving of maturity guarantees for segregated funds. In this section we use some of the same methods to model and hedge maturity guarantees, however we apply the beta distortion risk measures to this problem. Since a hedge portfolio can have gains as well as losses, the combination of the maturity guarantee and the hedge portfolio can not be compared in the same way we compared maturity guarantees and annuity rate

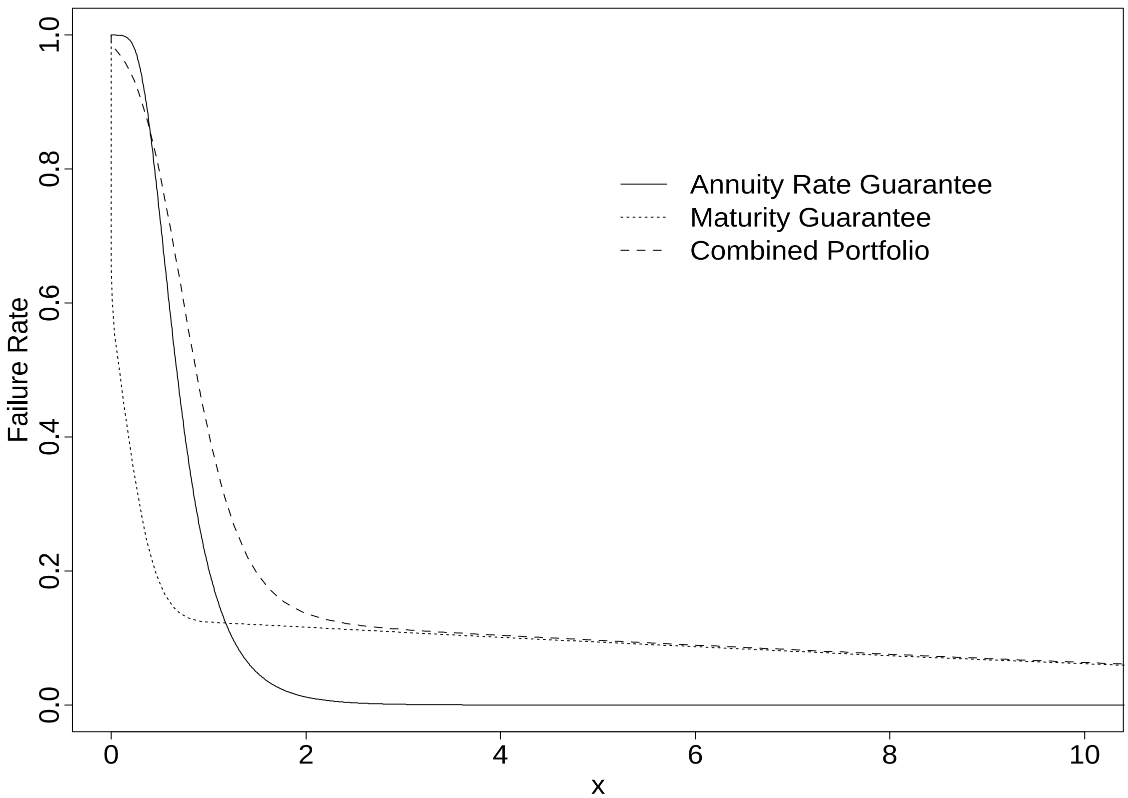

Figure 5.9: Smoothed Failure Rates for VaR Equated Combined Portfolio

guarantees. The losses for the maturity guarantee are directly offset by the gains in the hedge fund. If the maturity guarantee risk is hedged, instead of having some fairly large losses with less than a 10% probability, we will have paid a fixed cost for the hedge and have considerably reduced our risk.

To illustrate, the liability of a maturity guarantee is simply a put option, where the payoff is the maximum of zero and the guarantee amount minus the fund accumu-

	Combined	**Maturity Guarantee**	**Annuity Guarantee**
Mean	1.022	0.200	0.822
Std. Dev.	0.657	0.655	0.480
Kurtosis	10.34	19.12	10.57
DP 19	2.785	2.135	2.000
PH 19	19.15	5.018	18.87
PH 4	3.021	2.058	2.515
Beta($\frac{1}{2}$, 2)	2.119	1.272	1.633
Beta($\frac{1}{4}$, 4)	**4.059**	**3.046**	**3.392**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	4.413	3.263	3.748
VaR 95%	2.375	1.723	1.722
VaR 99%	3.671	3.383	2.484
CTE 90%	2.601	1.920	1.879
CTE 95%	3.151	2.733	2.204

Table 5.8: VaR Equated Combined Portfolio Capital Requirements, using the Log-normal Model

lation at the end of the guarantee.

$$L = \max\left[0, K - S(T)\right] \tag{5.18}$$

Thus during the accumulation period, the insurer can use the Black-Scholes option pricing formula to determine the price of the guarantee as:

$$\begin{aligned} BSP_{Put}(S_t, t) &= Ke^{-r(T-t)}N(-d_2(t)) - S(t)N(-d_1(t)) & (5.19)\\ \text{where } d_1(t) &= \frac{\log(\frac{S(t)}{K}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} & (5.20)\\ \text{and } \ d_2(t) &= d_1(t) - \sigma\sqrt{T-t} & (5.21) \end{aligned}$$

The constant K is the guarantee per initial value of the fund $S(0)$, r is the risk free rate of return, σ is the volatility of the yearly fund accumulation rate and $N(.)$ is the standard normal cumulative probability. The option matures at time T and the Black-Scholes price is determined at time t. Using this formula, one can easily hedge against the liability by investing in $N(-d_2(t))$ units of risk free bonds maturing at time T and taking a short position in $N(-d_1(t))$ units of the asset. In our case, the asset is the segregated fund, $K = 100$, $S(0) = 100(1-0.02)^T$, $T = 10$, $\sigma = 0.17$ and $r = 0.06$.

The Black-Scholes hedging formula assumes continuous adjustments of the hedge ratio. In practice this is neither efficient nor practical, due to transaction costs. In our illustration we assume no transactions costs. If we assume monthly or weekly hedge ratio adjustments, we can determine the initial price of the hedge, and by simulating the fund value determine the adjustments required to rebalance the hedge. In rebalancing the hedge, changes to the investments in bonds and in short assets are needed, the net cost of an adjustment is the difference between the cost of the position that should be held at time t:

$$BSP_{Put}(t, hedge_t) = Ke^{-r(T-t)}N(-d_2(t)) - S(t)N(-d_1(t)) \tag{5.22}$$

and the value at time t of the position taken at time $t-1$, that is $N(-d_2(t-1))$ units of the risk free bond and a short position of $N(-d_1(t-1))$ units of the asset, which has a current time t value of

$$BSP_{Put}(t, hedge_{t-1}) = Ke^{-r(T-t)}N(-d_2(t-1)) - S(t)N(-d_1(t-1)). \tag{5.23}$$

To adjust the hedge portfolio, $N(-d_2(t)) - N(-d_2(t-1))$ units of risk free bonds are purchased and a short position is taken in $N(-d_1(t)) - N(-d_1(t-1))$ units of the asset. The net cost is:

$$\begin{aligned} ADJ_{Put}(t) &= Ke^{-r(T-t)}\left[N(-d_2(t)) - N(-d_2(t-1))\right] \\ &\quad -S(t)\left[N(-d_1(t)) - N(-d_1(t-1))\right]. \end{aligned} \tag{5.24}$$

Any negative values of this adjustment can be viewed as a release of capital, which can be used to fund positive values of the adjustment. The total hedging error is the present value of the series of adjustments, and the risk inherent in the hedged maturity guarantee is based on the distribution of the present value of these adjustments.

Using the Wilkie investment model, readjustments to the hedge only occur yearly, since this model is designed as an annual model. A monthly or weekly model similar to Wilkie's, if available, would have a different structure. In practice, monthly, weekly or daily adjustments to the hedge portfolio are more appropriate than yearly adjustments. Thus we rely on the lognormal distribution for the fund accumulation factors. The parameters of the lognormal distribution, $\mu = 0.081$ and $\sigma = 0.17$ are the same as in the previous sections. For a 100% maturity guarantee with a 2% management charge, the initial cost to hedge the maturity guarantee is \$2.58. Using this model, 100,000 simulations of the ten-year monthly, weekly and daily accumulation processes are used to determine the distribution of the present value of the hedging error. Table 5.9 shows an analysis of the hedging error, when

	Monthly	**Weekly**	**Daily**
MEAN	0.231	0.110	0.041
Std. Dev.	0.432	0.207	0.077
Uncensored Kurtosis	7.518	7.879	8.023
Censored Kurtosis	17.91	18.90	18.99
DP 19	1.437	0.685	0.257
PH 19	4.271	2.299	0.782
PH 4	1.645	0.825	0.301
Beta($\frac{1}{2}$, 2)	1.003	0.484	0.181
Beta($\frac{1}{4}$, 4)	**2.387**	**1.199**	**0.437**
Beta($\frac{1}{\sqrt{19}}$, $\sqrt{19}$)	2.576	1.302	0.472
VaR 95%	1.100	0.525	0.195
VaR 99%	2.038	0.973	0.365
CTE 90%	1.286	0.614	0.230
CTE 95%	1.691	0.805	0.301

Table 5.9: Hedging Error Capital Requirements, using the Lognormal Maturity Guarantee

rebalancing occurs monthly, weekly, and daily. For all three rebalancing periods the beta risk measure, Beta($\frac{1}{4}$, 4), is larger than 99% VaR and 95% CTE. These numbers assume no transaction costs.

Figures 5.10 and 5.11 illustrate the density function for the uncensored and censored hedging errors for a 100% maturity guarantee with a 2% management charge, and Figure 5.12 shows the failure rates. The original capital requirement of \$23.43 is the amount the investor has to hold for the unhedged maturity guarantee. There is more than a 90% probability that none of this capital will actually be needed, since the maturity guarantee will not be exercised. However, if the maturity guarantee is hedged, the cost of the hedge is \$2.58, a fixed cost, and the investor

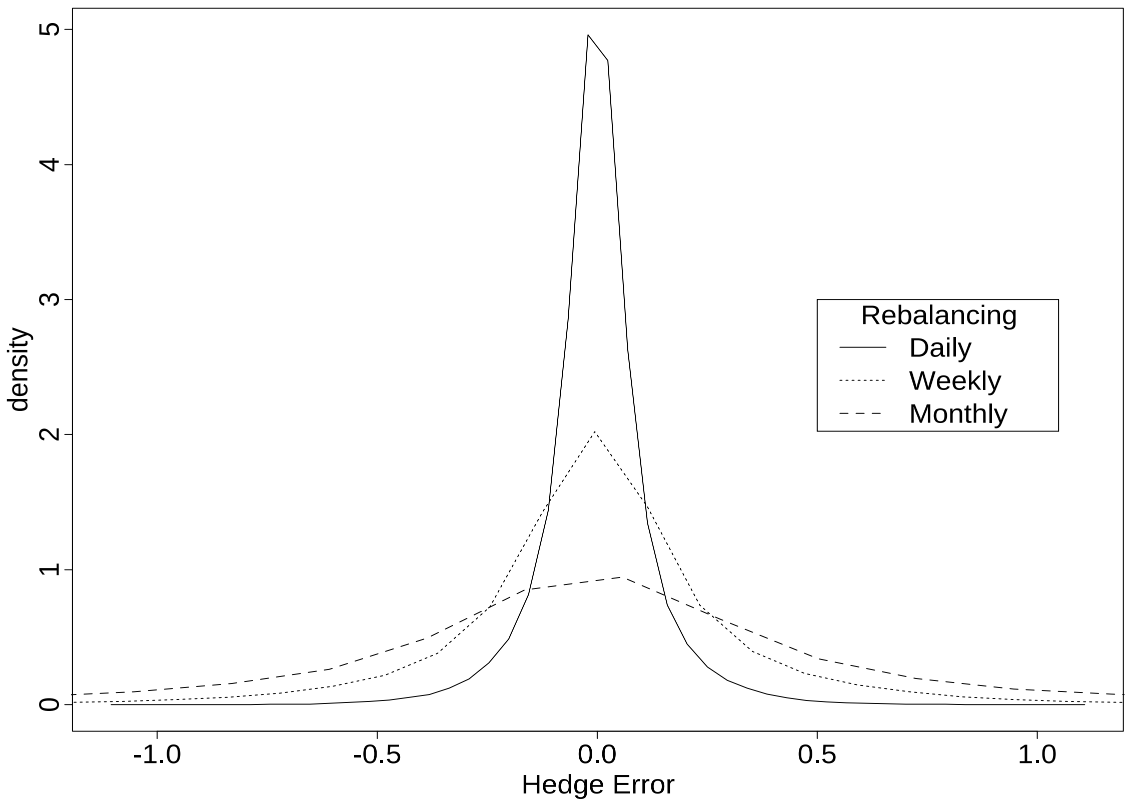

Figure 5.10: Density for Uncensored Hedging Errors of a Maturity Guarantee with a 2% Management Charge, Assuming Lognormal Investment Model

will still have a capital requirement for the hedging error of up to $2.39 (monthly hedging). The risk in the maturity guarantee has been traded for a certain payment with a small residual risk.

This example illustrates how hedging risk can limit the capital required to cover the risk in a portfolio. If the maturity guarantee and the hedge were parts of two different portfolios, and the risk measure was applied separately to each portfolio, the capital requirement would not account for their offsetting nature. Thus there is great advantage to using a holistic approach to measuring risk capital.

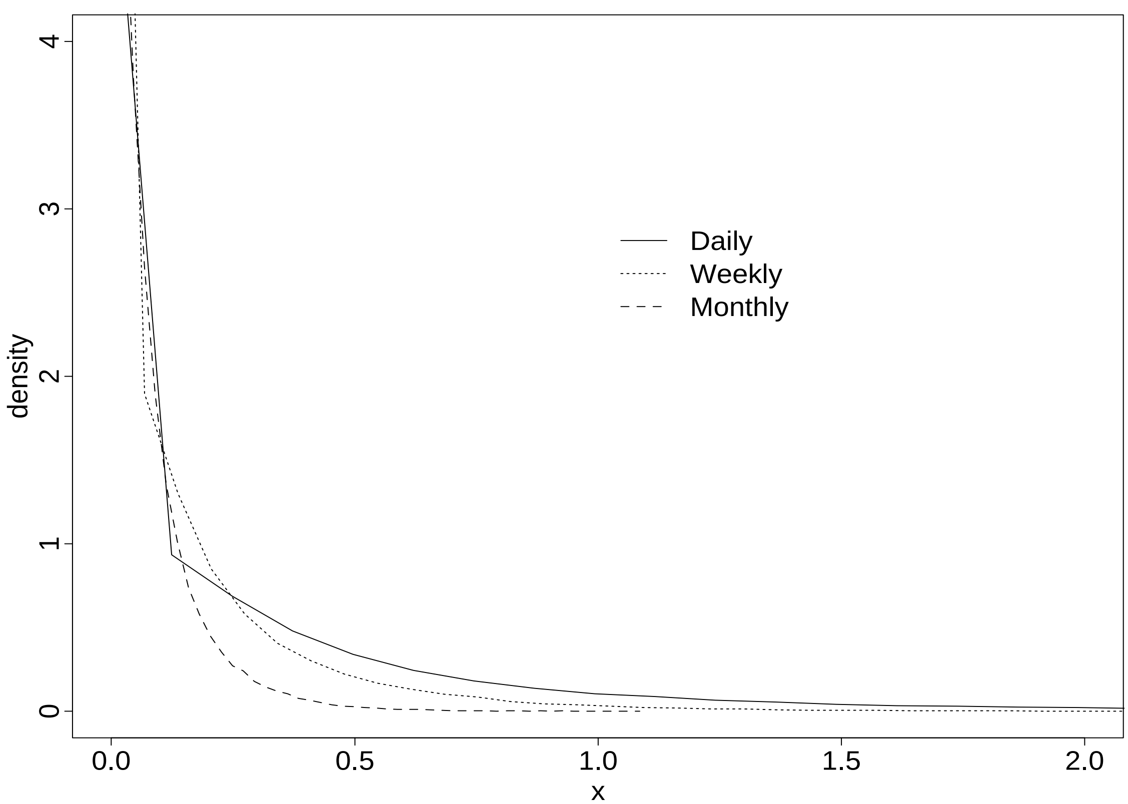

Figure 5.11: Density for Censored Hedging Errors for a Maturity Guarantee with a 2% Management Charge, Assuming Lognormal Investment Model

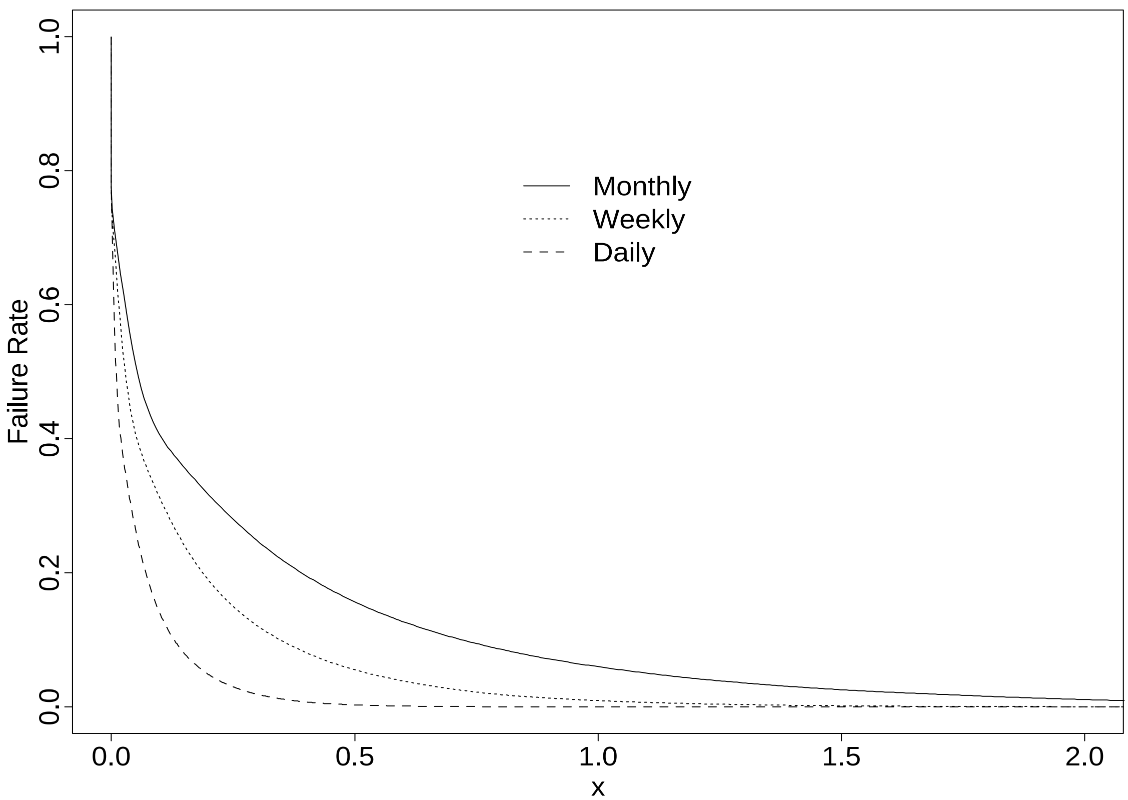

Figure 5.12: Smoothed Failure Rates for Maturity Guarantee Hedging Errors

Chapter 6

Review, Conclusions and Areas for Further Research

6.1 Review and Conclusions

The main purposes of this dissertation have been (i) to illustrate the insufficiency of the value-at-risk methodology for determining capital requirements and (ii) to provide a new coherent risk measure that can be applied to loss distributions in a way that coincides with the underlying purpose for having capital requirements.

Coherent risk measures have properties that are consistent with popular ideas of risk aversion, including first and second stochastic dominance. Value-at-risk is not a coherent risk measure, and fails three of the five coherency requirements. In contrast, all concave distorted risk measures are coherent. There is a distortion function that replicates VaR, however it is not concave, and is not coherent. The PH-transform and the dual power transform are two coherent distorted risk measures that have been introduced by Wang for applications in premium principles. Both the PH-transform and the dual power transform are members of the beta family of distribution functions. We have shown that the beta distribution function is a concave distortion function over a specific range of parameters, with the PH-transform and the dual power transform as special cases. The Kullback-Leibler mean information was used to compare parameters for the beta transform with parameters for the PH-transform and the dual power transform. Using second order stochastic dominance, we were able to prove that strictly concave distortion functions preserve second order stochastic dominance whereas general concave distortion functions only preserve this weakly. This leads to a preference for the beta distorted risk measure over the CTE and all other piecewise linear concave distortion functions.

Tail analysis was used to determine that the transformed failure rate in the tail of the distribution became heavier tailed, but always maintained the increasing or decreasing failure rate property. The Beta risk measures were applied to three portfolios, a portfolio of maturity guarantees and a portfolio of annuity rate guarantees, and a combined portfolio of maturity and annuity rate guarantees. Even though the maturity guarantees have an upper bound to the loss, the annuity rate guarantee loss distribution has a lighter tail than the maturity guarantees. This is evident by the kurtosis of the two portfolios, and from the comparison of the two portfolios in the mean equated combined portfolio example. Using two techniques for allocating risk capital, we illustrated how the Beta risk measure can be used to determine the division of capital requirements among portfolios. Lastly, we considered hedging a portfolio of maturity guarantees and illustrated how the required capital would be reduced in these circumstances. Throughout these illustrations, the Beta($\frac{1}{4}$, 4) risk measure provided a good estimate of the risk capital that should be held for each portfolio.

6.2 Areas for Further Research

6.2.1 Censored vs. Uncensored Loss Distribution

Using a censored distribution leads to discrepancies between time zero pricing of risks and time zero risk measure values. As well, for risk measures using the censored loss distribution, the translation invariant property of coherency generally applies only to the censored risk. In order to remove this discrepancy and to have the translation invariant property apply to the uncensored risk, a method used by Wang (1999) should be investigated. This method allows the loss random variable to be negative (uncensored) and uses a dual distortion of g, $g^*(u) = 1 - g(1-u)$, to transform the negative losses. In other words, for any risk X, with ddf $S_X(x)$, the risk measure $\rho_g^*(X)$ is defined by,

$$\rho_g^*(X) = \int_{-\infty}^{0} g[S_X(x)] - 1 \;\; dx + \int_0^{\infty} g[S_X(x)] \;\; dx \tag{6.1}$$

If g is a concave distortion function, then g^* is a convex distortion function. For the PH-transform $g(x) = x^{\frac{1}{\gamma}}$, the dual distortion $g^*(x) = 1 - (1-x)^{\kappa}$, is the dual power transform. As well, for the beta distortion,

$$g(x) = \int_0^x \frac{1}{\beta(\frac{1}{\gamma}, \kappa)} t^{\frac{1}{\gamma}-1}(1-t)^{\kappa-1} dt \tag{6.2}$$

the dual distortion is

$$g^*(x) = \int_0^{1-x} \frac{1}{\beta(\frac{1}{\gamma}, \kappa)} (1-t)^{\frac{1}{\gamma}-1} t^{\kappa-1} dt, \tag{6.3}$$

which is also Beta. If $\kappa = \gamma$, then g^* is Beta with inverted parameters. This risk measure is affected by the size of the possible gains, and may have limited application in capital adequacy, however there may be more appropriate application for this risk measure in valuation and pricing aspects of insurance and finance.

6.2.2 Extreme Value Theory

We have used extreme value theory(Bassi, Embrechts and Kafetzaki, 1997) to help with our understanding of percentile-VaR and the dual power risk measure, however we have not considered the application of this theory to the general Beta risk measure.

6.2.3 Dynamic Risk Measures

This dissertation considers single period risk measures. To generalize these risk measures to the multiperiod case, properties of dynamic risk measures will have to be investigated. Tan Wang (1999) develops a set of properties that are desirable for dynamic risk measures which may help to suggest a method to generalize the beta risk measure.

6.2.4 Mixture Problems

Another desirable property that has not been incorporated in this thesis is that a risk measure should have an extra risk loading for parameter uncertainty. This does not lead to a contradiction of the subadditivity rule, but pertains to mixture problems in insurance. The PH-distortion risk measures for $\gamma > 1$ satisfy this property (Wang, 1995a), which may be able to be extended to the Beta distortion risk measures with $\gamma > 1$ and $\kappa > 1$. Applications of this property may include risks with limited available history, or in modeling liquidity risk or credit risk.

6.2.5 Allocating Risk Capital Among Portfolios

The marginal and proportional methods for capital allocation among portfolios were used in Chapter 5. Defining correlation structures between portfolios, or subdividing portfolios into independent subportfolios, may facilitate this process. As well, another method using conditional allocation could be incorporated based on the following relation:

Let $\rho_g(.)$ be any coherent risk measure, then for any censored risks X and Y we have,

$$\begin{aligned} \rho(X+Y) &\leq \rho_g(X) + \rho_g(Y) & (6.4)\\ &= E_Y[\rho_g(X|Y)] + E_X[\rho_g(Y|X)] & (6.5)\\ &= \rho_U[\rho_g(X|Y)] + \rho_U[\rho_g(Y|X)] & (6.6) \end{aligned}$$

where U is the Uniform distorted risk measure.

6.2.6 Allocating Risk Capital Among Risk Factors Within One Portfolio

In this application, we are considering dividing the risk capital among risk factors. Examples of risk factors already discussed are: the short-term interest rate, the long-term interest rate, the fund accumulation rate, the inflation rate, the mortality rate, and the lapse rate. Once the relevant risk factors have been identified, we would like to determine the size or amount of risk that relates to each risk factor. To illustrate, given a portfolio X, let Q be a vector of risk factors, so that $Q = (q_1, q_2, ..., q_k)$, using risk measure ρ, we would like to divide the total risk capital $\rho(X)$ into portions relating to the risk capital required due to the risk from each risk factor, $\rho(X; q_i)$, $i = 1, ..., k$. First we assume that the relevant risk factors are independent and discuss how this effects capital allocation, then we introduce dependence between the factors.

Independent Risk Factors

If risk factors are independent, then changes in one factor do not effect the value of other factors, and the cumulative distribution of risk factors can be factored into a product of their marginal distributions. Depending on the risk measure used and the construction of the portfolio, having independent risk factors can lead to different results. To illustrate, let Q be a vector of risk factors, so that $Q = (q_1, q_2, ..., q_k)$, since the q_i's are independent, the decumulative distribution function for Q can be factored, so that

$$F_Q(X) = F_{q_1}(X)\ F_{q_2}(X)...\ F_{q_k}(X) \tag{6.7}$$

If the portfolio can also be divided into subportfolios that only depend on one risk factor, so that $X = X_{q_1} + X_{q_2} + ... + X_{q_k}$, then by using an additive risk measure, for example $\rho(X) = E[X]$, we would have

$$\rho_U(X) = E[X] = \int_0^\infty X dF_Q(x) \tag{6.8}$$

$$= \int_0^\infty X_{q_1} + X_{q_2} + ... + X_{q_k} d(F_{q_1}(X)\ F_{q_2}(X)...\ F_{q_k}(X)) \tag{6.9}$$

$$= \int_0^\infty X_{q_1} dF_{q_1}(X) + \int_0^\infty X_{q_2} dF_{q_2}(X) + ... + \int_0^\infty X_{q_k} dF_{q_k}(X) \tag{6.10}$$

$$= \rho(X; q_1) + \rho(X; q_2) + ... + \rho(X; q_k) \tag{6.11}$$

Since $E[X]$ is an additive risk measure, for independent risks with independent subportfolios, we have

$$\rho(X) = \sum_{i=1}^{k} \rho(X; q_i) \tag{6.12}$$

Examples of additive risk models:

Example 6.2.1 Historical Price Sensitivity: *Let* $\rho(X)$ *be distributed as*

$$\partial X_Q = \sum_{i=1}^{k} \frac{dX_{q_i}}{dq_i} \quad \partial q_i. \tag{6.13}$$

This implies that $\rho(X; q_i)$ *is distributed as*

$$\frac{dX_{q_i}}{dq_i} \quad \partial q_i \tag{6.14}$$

$$\text{and} \quad \rho(X; Q) = \sum_{i=1}^{k} \rho(X; q_i) \tag{6.15}$$

This is similar to the method suggested by the 1993 revision of the Basel Capital Accord, where the four risk factors considered were interest rate risk, exchange rate risk, commodity risk and equity risk, if these risks are assumed independent.

Example 6.2.2 The Analytical Method: *The analytic model assumes that the portfolio can be broken into smaller units, each unit depending on one variable. In this case:*

$$X = \sum_{i=1}^{k} X_{q_i}. \tag{6.16}$$

Using $\rho(X) = Var(X)$ *the variance, and using the assumption that the risk factors are independent, let*

$$\rho(X; Q) = Var_Q(X) = Var_Q(\sum_{i=1}^{k} X_{q_i}) = \sum_{i=1}^{k} Var_{q_i}(X_{q_i}) = \sum_{i=1}^{k} \rho(X_{q_i}) \tag{6.17}$$

Additive models provide a natural algorithm to determine the relative importance or the sensitivity of the portfolio to each risk factor, by comparing individual risk measure values. The higher the risk measure value for an individual risk factor, the higher the exposure and the relative importance of that risk factor. Non-additive models are not as easily decomposed. Assuming that the risk factors are known to be independent, in special cases a transformation can produce an additive model; however this is generally not the case.

Dependent Risk Factors

A common technique used to compute a risk measure in practice is to use the simulation approach. Often it is assumed that the joint distribution of the risk factors follows a multivariate normal model (Frees and Valdez, 1998). This facilitates the calculation and analysis of the risk measure; however, it limits the risk measure to only evaluating risk related to second order dependencies in risk factors. Risk factors are each assumed to follow a normal distribution, and are related to other risk factors only through a correlation factor. The risk related to the change in price of a financial investment could depend on an underlying risk factor through the skewness or kurtosis of the distribution of the risk factor, or on the change in volatility. As well, dependencies between risk factors may be insufficiently modeled by a correlation. Derivative securities often provide examples of the first problem. Using the multivariate normal assumption, only delta and gamma risks are evaluated. Methods in finance that are used to value derivatives often consider the theta and vega risks of a portfolio, the risk related to the change in price of the investment with respect to time, and volatility. Using the simulation approach, it is possible to use any form for the joint density function, and more complex relationships between risk factors can be implemented.

An alternative option to assuming multivariate normal risk factor distributions is to assume some relation between the marginal and joint risk factor distributions using a copula (Frees and Valdez, 1998). Preliminary investigations into this area have identified that copulas are very useful in modeling dependence of random variables. They provide relevant information about the dependence structure of a multidimensional random vector. Thus copulas merit further attention and may prove useful for our study of multiple sub-portfolios and for our study of portfolios with multiple risk factors. Frees and Valdez (1998) illustrates two methods to specify a family of copulas for bivariate data, however copulas for higher dimension multivariate data are not always convenient to identify.

References

Artzner, P. (1997). Thinking coherently. *Risk Magazine*, 10(11):68–71.

Artzner, P. (1999). Application of coherent risk measures to capital requirements in finance. *North American Actuarial Journal*, 3(2):11–25.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1996). A characterization of measures of risk. Presented at University of Waterloo.

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

Barlow, R. and Proschan, F. (1965). *Mathematical Theory of Reliability.* John Wiley, New York.

Bassi, F., Embrechts, P., and Kafetzaki, M. (1997). A survival kit on quantile estimation. Presented at University of Waterloo.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., and Nesbit, C. J. (1997). *Actuarial Mathematics.* Society of Actuaries, Schaumburg, Illinois, second edition.

Boyle, P. P. and Hardy, M. R. (1998). Reserving for maturity guarantees: Two approaches. *Insurance: Mathematics and Economics*, pages 113–127.

Buchen, P. W. and Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31(1):143–159.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory.* John Wiley and Sons Inc.

Deprez, O. and Gerber, H. (1985). On convex principles of premium calculation. *Insurance: Mathematics and Economics*, 3(4):179–190.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *The Journal of Derivatives*, 4(3):7–48.

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modeling Extremal Events for insurance and Finance.* Springer.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.

Gerber, H. U. (1979). *An Introduction to Mathematical Risk Theory.* S.S. Huebner Foundation for Insurance Education, Philadelphia, Pennsylvania.

Goovaerts, DeVylder, F., and Haezendonck, J. (1984). *Insurance Premiums.* North Holland, The Netherlands.

Hardy, M. (1998). Maturity guarantees for segregated fund contracts; hedging and reserving. *Institute for Insurance and Pensions Research.*

Hogg, R. V. and Craig, A. T. (1995). *Introduction to Mathematical Statistics.* Prentice Hall, Englewood Cliffs, New Jersey, fifth edition.

Hogg, R. V. and Klugman, S. A. (1984). *Loss Distributions.* Wiley, New York.

Huber, P. (1997). A review of Wilkie's stochastic asset model. *British Actuarial Journal*, 3(1):181–210.

Hull, J. C. (1997). *Options, Futures, and Other Derivatives.* Prentice Hall, Upper Saddle River, New Jersey, third edition.

Kaas, R., van Heerwarden, A., and Goovaerts, M. (1994). *Ordering of Actuarial Risks.* Caire, Brussels.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models: From Data to Decisions.* John Wiley and Sons, INC., Toronto, Ontario.

McQueen, R. (1996). *Who Killed Confederation Life?* McClelland and Stewart Inc.

Morgan, J. P. (1995). Introduction to riskmetrics. Technical report, Guaranty Trust Company: Risk Management Services, riskmetrics@jpmorgan.com, New York.

Müller, A. (1996). Ordering of risks: A comparative study via stop-loss transforms. *Insurance: Mathematics and Economics*, 17:215–222.

van Heerwaarden, A. (1991). *Ordering of Risks.* Thesis Publishers Amsterdam, Amsterdam.

Wang, S. (1995a). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54.

Wang, S. (1995b). Premium calculation by transforming the layer premium density. Technical Report 95-18, University of Waterloo: Institute of Insurance and Pension Research.

Wang, S. (1995c). The price of risk: An actuarial/economic model. Technical Report 95-06, University of Waterloo: Institute of Insurance and Pension Research.

Wang, S. (1996a). Ordering of risks under PH-Transforms. *Insurance: Mathematics and Economics*, 18(2):109–114.

Wang, S. (1996b). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(1):71–92.

Wang, S. (1996c). Risk measures with applications in insurance ratemaking and actuarial valuation. Technical report, University of Waterloo; Course Notes.

Wang, S. (1998). An actuarial index of the right-tail risk. *North American Actuarial Journal*, 2(2):88–101.

Wang, S. and Young, V. R. (1998). Risk-adjusted credibility premiums using distorted probabilities. *Scandanavian Actuarial Journal*, 1998(2):143–165.

Wang, S. S. (1999a). A class of distortion operators for pricing financial and insurance risks. Technical report, SCOR Reinsurance Company.

Wang, S. S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2):173–183.

Wang, T. (1999b). A characterization of dynamic risk measures. Technical report, University of British Columbia.

Wilkie, A. (1995). More on a stochastic asset model for actuarial use. *British Actuarial Journal*, 1(5):777–911.

Wirch, J. L. (1999). Raising value at risk. *North American Actuarial Journal*, 3(2):106–115.

Wirch, J. L. and Hardy, M. R. (2000). A synthesis of risk measures. *Insurance: Mathematics and Economics*.