

Estimation and allocation of
insurance risk capital

by

Hyun Tae Kim

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2007

©Hyun Tae Kim 2007

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

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

Hyun Tae Kim

Acknowledgements

When properly managed, risks can be opportunities. In my life, looking back, those risks have turned to be blessings; it just took a fair time to realize that.

First of all I thank my wife Jinsun (Julia) for her love, support, and patience during my Ph.D. studies. Ever since we got married she has been my stronghold and the closest friend. Our beloved children, Christopher and Hannah, have been very nicely behaving; there will be one less student in my family now, but I am sure they wouldn't mind. My parents have been my role models ever since I started understanding things in life; they taught me by examples. My parents-in-law played a special role in my life as well.

I am deeply indebted to Professor Mary Hardy, my supervisor, who knows exactly how to train and support her apprentice. Her knowledge, insight, passion, and patience have been my life-time lessons; I could not ask for a better mentor as a student. Special thanks go to Professor Harry Panjer, my master program's supervisor, who has supported me over various obstacles since I arrived at UW; Professor Gord Willmot for his support and kindness during my studies; Professor Paul Marriott generously spent much time with me in reviewing and discussing the thesis material, which substantially improved not only the quality of the final thesis, but also my ability to think critically. Professor Tony Wirjanto at Economics went through my thesis cover to cover and provided valuable suggestions as well as a list of typos in the thesis; Professor Jed Frees, the external examiner, made valuable suggestions on various spots in the thesis which I had missed.

I am glad to know Professor David Matthews, the chair of the department, who has done great things for me. Professor Ken Seng Tan always cheered up my spirit whenever I felt blue. Professor Rob Brown kindly supported me during my master's

program with an exciting project. I also thank graduate officers of the department for their supports. Students in our department are lucky to have such remarkable support staff. Mary Lou Dufton, the graduate secretary of the department, knows my school life better than my wife does, and I owe her cookies and chocolates beyond numbering. I have had a lot of help from the expertise of Lucy Simpson, Gwen Sharp, Joan Hatton, Anissia Anniss, and Amy Aldous. I also like to thank Linda Lingard, who now is retired, for her special care for me during my early days at UW.

There are many students that I am thankful of. Joonghee, So-Yeun, Chang-kee, Yunhee, Chongsun and other Korean students have shared much of my memories on the 6th floor of the MC building. Thanks also go to Kai, Johnny, Ali, Bill, George and Diego for their friendships. My office mates Lilia and Ivan made our office bigger and brighter than it actually is.

Chapter 5 of the thesis was possible thanks to the data that GGY provided for me. I am very grateful to David Gilliland who allowed me to use the AXIS software and Wes Leong for his technical support and effort he spared for me. I would also like to acknowledge the research support from our department, the Institute for Quantitative Finance and Insurance, Faculty of Mathematics, Natural Sciences and Engineering Research Council of Canada, the Society of Actuaries.

Finally, and most of all, I thank the Lord who have made all these things possible and led me to where I stand today.

“The riddles of God are more satisfying than the solutions of man.”

Gilbert K. Chesterton

Introduction to the Book of Job, 1907

Abstract

Estimating tail risk measures such as Value at Risk (VaR) and Conditional Tail Expectation (CTE) is a vital component in financial and actuarial risk management. The CTE is a preferred risk measure, due to coherence and a widespread acceptance in actuarial community. In particular we focus on the estimation of the CTE using both parametric and nonparametric approaches.

In parametric case the conditional tail expectation and variance are analytically derived for the exponential distribution family and its transformed distributions.

For small i.i.d. samples the exact bootstrap (EB) and the influence function are used as nonparametric methods in estimating the bias and the the variance of the empirical CTE. In particular, it is shown that the bias is corrected using the bootstrap for the CTE case. In variance estimation the influence function of the bootstrapped quantile is derived, and can be used to estimate the variance of any bootstrapped L-estimator without simulations, including the VaR and the CTE, via the nonparametric delta method. An industry model are provided by applying theoretical findings on the bias and the variance of the estimated CTE.

Finally a new capital allocation method is proposed. Inspired by the allocation of the solvency exchange option by Sherris (2006), this method resembles the CTE allocation in its form and properties, but has its own unique features, such as manager-based decomposition. Through a numerical example the proposed allocation is shown to fail the no undercut axiom, but we argue that this axiom may not be aligned with the economic reality.

Contents

1	Introduction	1
1.1	Introduction and Motivation	1
1.2	Challenges	5
1.3	Risk measure as required capital	7
1.4	The Conditional Tail Expectation	13
1.5	Modelling Dependency	15
1.6	Capital Allocation	17
1.7	Overview of Thesis	21
2	Conditional tail moments of the EF and its related distributions	23
2.1	Introduction	23
2.2	The Exponential Family	25
2.3	Conditional tail moments of the EF	29
2.3.1	When generating function exists	29
2.3.2	When no generating function exists	36
2.4	Transformed Distributions	43
2.5	Conclusion	50

3	Bias correction of risk measures using the bootstrap	51
3.1	Introduction and motivation	51
3.2	Estimating VaR and CTE using finite sample	54
3.3	Bias of sample estimates of VaR and CTE	56
3.4	The bootstrap	60
3.4.1	Overview	60
3.4.2	Bootstrapping L-estimators	63
3.5	Simulations	70
3.5.1	Examples	70
3.5.2	Estimating the 99% Quantile risk measure	71
3.5.3	Estimating the 95% CTE	76
3.6	A practical guideline	79
3.7	Concluding remarks	82
4	Variance estimation of bootstrapped risk measures	85
4.1	Introduction	85
4.2	The bootstrap	88
4.2.1	Improvement using recursion	92
4.3	Nonparametric delta method	95
4.4	IF of bootstrapped L-estimator	101
4.5	Numerical example	109
4.6	Concluding remarks	111
5	A simulation study	113
5.1	Introduction	113
5.2	The model	116

5.3	Simulation backgrounds	119
5.4	Analysis	120
5.5	Alternative model	122
5.6	Model and parameter uncertainty	130
5.7	Concluding remarks	133
6	A new capital allocation method	135
6.1	Introduction	135
6.2	Value vs. Risk	137
6.2.1	Two measures	137
6.2.2	Valuation perspective	138
6.2.3	Solvency perspective	142
6.3	Allocation review	143
6.4	A new allocation	149
6.5	Properties of the new allocation	154
6.5.1	Fairness of the allocation	155
6.5.2	Further decomposition	156
6.6	Numerical example	161
6.7	No undercut axiom	165
6.8	Concluding remarks	170
7	Discussion and future work	171
7.1	Summary and discussion	171
7.2	Future ideas	174
7.2.1	CTE Comparison: parametric vs. nonparametric	174
7.2.2	The Shifted Power Transform	174

7.2.3	Bootstrap after transformation	176
7.2.4	A semi-parametric method for CTE estimation	176
7.2.5	Fitting the GPD for Tail Region	177
7.2.6	Allocation via Bootstrap	178

List of Tables

1.1	US RBC framework for life insurer	3
1.2	Canada MCCSR framework	4
3.1	99% Quantile estimators for the Lognormal example	73
3.2	99% Quantile estimators for the RSLN2 example	74
3.3	99% Quantile estimators for the Pareto example	75
3.4	95% CTE estimators for the lognormal example	77
3.5	95% CTE estimators for the RSLN-2 example	78
3.6	95% CTE estimators for the Pareto example	78
3.7	Application of MSE test to the three examples, 99% CTE.	83
4.1	Variance of the EB CTE 95% for three examples	110
5.1	Basic statistics based on the full data (Unit: CDN dollar)	120
5.2	Simulation result of the Seg. fund model	123
6.1	Ordinary balance sheet of an insurer	140
6.2	Entitlements of shareholders and policy holders at time 1	140
6.3	Economic balance sheet of an insurer	141
6.4	Entitlements of shareholders and policy holders at time 1 for line i . .	149

6.5	Economic balance sheet of an insurer with premium inflow	151
6.6	Valuations of company A and B	163
6.7	Valuations of the merged company ANB	163
6.8	Capital allocations to each line of company ANB	164
6.9	Different decompositions of the allocated capital (ρ is set at CTE 95%)	166
6.10	Capital allocations of company A, before and after adding Seg fund 3.	167

List of Figures

3.1	The weights of the EB for several order statistics when $n = 100$. . .	64
3.2	Comparisons of the weights: HD vs. EB estimators when $n = 100$. . .	66
4.1	IF of mean, quantile, CTE for $U(0, 1)$, $\alpha = 80\%$	99
4.2	IF of quantile: Empirical (top) vs. EB (bottom) for $U(0, 1)$, $\alpha = 80\%$	106
4.3	IF of CTE: ordinary CTE (top) vs. EB (bottom) for $U(0, 1)$, $\alpha = 80\%$	108
5.1	Cell assumption screen (top) and cash flow projection (bottom) . . .	118
5.2	Histogram of GMMB loss based on the full data	121
5.3	Standardized Q-Q plot after log transform	125
5.4	Sample mean excess function of the data	127
5.5	Fitted GPD for the data above threshold	128

Chapter 1

Introduction

1.1 Introduction and Motivation

Equipped with advances in quantitative techniques and computing technology, the methodology for determining an appropriate level of required capital or risk capital of a company has recently received much attention and has been actively investigated by both academics and practitioners. The importance of this topic lies in the fact that risk capital is a cushion against unexpected loss that can lead to insolvency. A company needs to hold adequate capital in order to keep the probability of insolvency as low as possible, while staying competitive in the capital market. Quantifying the degree of adequacy and determining the capital amount is thus an important question to be answered in practical risk management. In the financial industry the parties interested in the determination of risk capital include the management of insurers and banks, regulators, rating agencies, and investors. Policyholders are also concerned with the capital level of insurers.

Currently different countries use different methods to regulate the capital level of insurers and each different method has its own characteristics. In North America, the regulatory required capital formulas take the bottom-up approach, meaning that the total capital is computed by combining risk capital for each line, with certain combining methods to reflect diversification of risks. For example the US has the Risk Based Capital (RBC) system where the combining method is based on the *square-root formula*. Table 1.1 shows the RBC formula for US life insurers. The formula varies across different insurance classes, but all adopt similar square root formulas; see the report by the Academy joint RBC task force of AAA Academy Joint RBC Task Force (2002) for details on this. Also refer to Butsic (1993) or Butsic (1994) for the developments of the original RBC formula for P&C insurers and Feldblum (1996) for a general discussion.

From a mathematical perspective the RBC formula recognizes the dependence of component risks through the square root rule but it is based on specific simplifying parametric assumptions; in particular it employs a normal loss assumption along with the (co)variance as the underlying risk measure. These assumptions are criticized as unrealistic in that the loss distribution often has fatter tails than the normal distribution and the covariance can only measure *linear* relationships whereas in the real world we tend to observe more complicated dependencies. Also each risk component is calculated by multiplying a risk charge factor to the nominal value of each asset¹; this does not effectively consider the tail region of risks where extreme

¹For example, interest risk component C3a in Table 1.1 is computed by policy reserves multiplied by prescribed factors. These factors are expressed in percentages and vary depending on the risk characteristics of underlying product.

<p>C0: Insurance affiliate investment and (non-derivative) off-balance sheet risk</p> <p>C1cs: Invested common stock asset risk</p> <p>C1o: Invested asset risk, plus reinsurance credit risk except for assets in C1cs</p> <p>C2: Insurance risk</p> <p>C3a: Interest rate risk</p> <p>C3b: Health provider credit risk</p> <p>C4a: Business risk - guaranty fund assessment and separate account risks</p> <p>C4b: Business risk - health administrative expense risk</p> $RBC = C0 + C4a + \sqrt{(C1o + C3a)^2 + (C1cs)^2 + (C2)^2 + (C3b)^2 + (C4b)^2}$

Table 1.1: US RBC framework for life insurer

losses tend to occur. Since its first introduction the RBC system has evolved to reflect risks more realistically and it is expected that stochastic modeling will be required in the next version of RBC system, so called C-3 Phase II for variable annuity guarantees; see, e.g., the report of Life Capital Adequacy Subcommittee (2005) of the American Academy of Actuaries.

Canada uses the “Minimum Continuing Capital and Surplus Requirements” (or MCCR in short) to determine the solvency margin of insurers. The MCCR of a life insurer is determined by applying factors for each of five risk components to specific assets or liabilities and by adding the results (see Table 1.2 for the five components). Then it is compared to the insurer’s capital which is composed of two tiers; core capital and supplementary capital. For details, refer to the MCCR guideline published

<ol style="list-style-type: none">1. Asset default risk (C-1 risk)2. mortality/morbidity/lapse risks3. Interest margin pricing risk4. Changes in interest rate environment risk (C-3 risk)5. Segregated funds guarantees risk <p>Required Capital = 1+2+3+4+5</p>

Table 1.2: Canada MCCR framework

by the Office of the Superintendent of Financial Institutions (2003) of Canada. The OSFI is currently using stochastic modelling to determine the capital requirements of the segregated fund guarantees. Stochastic modelling methods for guarantees of insurance and investment products including segregated funds are discussed in Hardy (2003).

A survey of different risk capital systems for other countries can be found in KPMG (2002) or Chapter 10 of Atkinson and Dallas (2000).

International efforts in developing a set of rules for accounting and capital requirements for insurance companies have been initiated by the actuarial and accounting professions and insurance regulators; similar efforts are made on the banking industry as well. One of the basic ideas is to develop a new framework that measures the inherent risks of companies more realistically, by moving to stochastic model-

ing. By a stochastic model we mean that the dynamic of a company is captured by a quantitative model, rather than from traditional factor-based or accounting-based methods. These approaches have been described by the Basel Committee on Banking Supervision, the International Association of Insurance Supervisors, the International Organization of Securities Commissions, and the International Accounting Standards Board, to name a few. The Society of Actuaries (SoA) has formed the Risk Management Task Force and the risk management section to develop risk management resources for actuaries in both traditional and non-traditional fields. A recent book by McNeil *et al.* (2005) provides an excellent introduction to the financial risk regulatory framework of the past and the future, as well as an extensive coverage on quantitative methods used in risk management.

1.2 Challenges

There are generic challenges around the required capital determination under stochastic modeling. Let us suppose a multi-line insurer has n different business lines with random losses denoted by X_1, X_2, \dots, X_n , respectively. It is a convention throughout this thesis unless otherwise specified that random loss (or variable) covers one period, i.e., the realization of loss occurs at the end of period, with the period defined to be a suitable time frame for business cycle and planning, such as one year. Then the aggregate loss of the company is $S = X_1 + X_2 + \dots + X_n$. Statistically the required capital can be expressed as a functional of the loss, or $\rho(S)$. Typically, assuming loss is represented by positive number, the required capital will be a number in the right tail region of the distribution to keep $Pr(S > \rho(S))$ small, so that the company's solvency is assured at some acceptable level. The first challenge is to define a suitable

functional $\rho()$ that can be reliable and consistent in some sense. We will formally define the risk measure and give a further discussion on its desirable properties later in this chapter.

Next, we should be able to properly model the aggregate loss S . Because the dynamic of S depends on the distribution of random vector (X_1, X_2, \dots, X_n) , or the multivariate loss distribution, the dependency structure plays a critical role in understanding the behavior of the aggregate loss. There have been two major directions on the parametric side to solve this problem. The first one is to use a class of simple multivariate distributions that is flexible enough to be applied to long-tailed and skewed data. The elliptical distribution class and mixtures of, such as the normal, distributions would be examples. The other approach is to split the multivariate distribution into the dependency structure and the marginal, and model the former separately. This is commonly known as the copula method. Additional difficulty arises from the fact that there is a demand for a more accurate modeling in the tail region, rather than the part around the centre of the distribution. If one brings in time as another variable, because much of market information is time-indexed, the problem extends to multivariate process.

Another interesting problem is the allocation of the capital to each business unit. Determining the allocated capital for a line is similar to determining the total capital of company in a sense that both try to do the same thing for a given business entity; one for a business unit, the other for the whole company. They can be however quite different because capital allocation is an *ex post* process, given the total capital. Given the total capital of an insurer, $\rho(S)$, the capital allocation determines how to

split this into smaller pieces to be assigned to each business unit. Mathematically,

$$\rho(S) = \sum_1^n AC_i,$$

where AC_i is the allocated capital for line i . Thus the capital allocation is a rule to determine each AC_i given $\rho(S)$. As in the risk measure case, there have been active research on allocation methods that are reasonable and consistent with some criteria. We elaborate this topic further in Section 1.6 and Chapter 6.

In the remaining sections in this chapter we provide a brief discussion on each of these challenging issues, to prepare for the developments throughout this thesis.

1.3 Risk measure as required capital

A risk measure, based on a loss random variable, is designed to tell how risky one's business (or random loss) is. In particular it is expressed by a number representing the risk embedded in the given random variable, or loss. Mean, higher moments, or quantiles are simple examples. Actuaries have long used various risk measures in setting adequate premium to charge for insurable risks to cover both expected and unexpected loss; see, e.g., Bühlmann (1980) and Wang *et al.* (1997). Recently risk measures have received attention for another reason, that is, for setting the required capital of a financial company. The capital is defined as the assets less the liabilities and serves as the cushion to protect the company from bankruptcy in adverse business events. So the task of computing an adequate premium for a loss is logically equivalent to that of determining an adequate capital amount for a company, if the company's operation is characterized by a random variable, even though there will be other economic or practical considerations in its setting. Mathematically a risk measure or

required capital of a loss random variable S is defined as a function that maps the random loss to a real number.

Definition 1.1 (e.g., McNeil *et al.* (2005), pp.238) *A risk measure or risk capital² of random variable S is a mapping ρ :*

$$\rho : S \rightarrow \mathbf{R}. \quad (1.1)$$

In the actuarial context, it is natural to think a risk measure as a mapping from a nonnegative random loss to a nonnegative real value. Since there is no mathematical restriction on $\rho(S)$ there can be many specific risk measures satisfying the above definition. For example, see Dhaene *et al.* (2006) and references therein for a survey of different risk measures currently used and their properties. Here we present some popular ones. One could use the expected loss with some additional padding as a risk measure, that is,

$$\rho(S) = E(S) + k\sqrt{Var(S)},$$

with a suitable k prescribed by management. When S is multivariate normal this leads to a similar framework to the current RBC system. Another example is the *Dutch risk measure*, defined by

$$\rho(S) = E(S) + kE[(S - \alpha E(S))_+], \quad 0 \leq k \leq 1, \alpha \geq 1,$$

which originated from the Dutch premium principle; see Kaas *et al.* (1994).

The most two popular and widely-accepted risk measures in current risk management endeavor are the quantile and the conditional tail expectation (CTE). The

²Some authors, e.g., Dhaene *et al.* (2006), define the risk capital as $\rho(S) - P(S)$ where $P(S)$ is the value of liability. We are content with our definition as long as there is no confusion.

α -th quantile, also called the Value at Risk at confidence level α (or VaR_α) in finance literature, is defined by

$$\rho(S) = Q_\alpha(S) = \inf\{x \in \mathbf{R} | P[X \leq x] \geq \alpha\}. \quad (1.2)$$

In practice α is a number close to 1, such as $\alpha = 0.95$. When the random variable is absolutely continuous this reduces to $Q_\alpha(S) = F_S^{-1}(\alpha)$.

The conditional tail expectation (CTE), also known as the tail conditional expectation (TCE), expected shortfall (ES), or TailVaR (TVaR)³, has received much attention recently as a preferred coherent risk measure in the sense of Artzner *et al.* (1999). The CTE at confidence level α is defined as the expected value of the random loss when the loss exceeds the threshold quantile:

$$\rho(S) = \text{CTE}_\alpha(S) = E[S | S > Q_\alpha(S)], \quad (1.3)$$

when S is continuous at $Q_\alpha(S)$ ⁴. While the list of different risk measures is not exhaustive, some risk measures are said to be better than others based on certain criteria.

In this regard, recent academic research has revealed many valuable concepts and extended our knowledge on risk measures and capital requirement. For example, Wang *et al.* (1997) discussed desirable properties of risk measures in connection with

³Dhaene *et al.* (2006) give slightly different definitions for the ES and the TVaR

⁴In the more general case the CTE with confidence level α is calculated as follows. Let $\beta' = \max\{\beta : Q_\alpha = Q_\beta\}$

$$\text{then } \text{CTE}_\alpha(X) = \frac{(1 - \beta') E[X | X > Q_\alpha] + (\beta' - \alpha) Q_\alpha}{1 - \alpha}$$

This complication is automatically managed when the CTE is estimated by simulation.

risk pricing. Most notably Artzner *et al.* (1999) proposed a set of axioms to define the coherence of risk measures for a single period of time frame; later Artzner *et al.* (2006) extended the criteria to multi-period of time and Hardy and Wirth (2004) proposed the Iterated CTE as an example of the multi-period model.

Definition 1.2 (Coherent risk measure; Artzner *et al.* (1999)) *A risk measure ρ satisfying the following four axioms is called coherent:*

1. *Subadditivity: For all random losses X and Y , $\rho(X + Y) \leq \rho(X) + \rho(Y)$*
2. *Monotonicity: If $X \leq Y$ for each outcome, then $\rho(X) \leq \rho(Y)$*
3. *Positive homogeneity: For positive constant b , $\rho(bX) = b\rho(X)$*
4. *Translation Invariance: For positive constant c , $\rho(X + c) = \rho(X) + c$*

The subadditivity axiom means that the required capital amount of sum of the losses are not bigger than the sum of required capital of each loss. The monotonicity axiom indicates the capital must be greater for a loss which is always bigger than another. Positive homogeneity requires that risk capital must be independent of a currency change. Translation invariance means that adding a sure risk increases the risk capital by the same amount. Coherence properties have been well accepted in academic community and provided an important basis for recent research on risk management. In practice the list of these axioms is used to examine a specifically chosen risk measure, rather than to produce a new specific risk measure.

It is useful to define the distortion function and the distortion risk measure to discuss the coherence of risk measures. The distortion risk measure was introduced

by Wang (1996) to the actuarial community. Let $g : [0, 1] \rightarrow [0, 1]$ be an increasing function with $g(0) = 0$ and $g(1) = 1$. Then the transform $g(F(x))$, where $F(x)$ is the distribution function of the random loss in question, defines a distorted probability distribution, and g is called a *distortion function*.

Definition 1.3 (Distortion risk measure) For a real-valued loss random variable X a distortion risk measure ρ_g is defined as the mean value under the distorted probability function $g(F(x))$:

$$\rho_g(X) = - \int_{-\infty}^0 [1 - g(\bar{F}(x))] dx + \int_0^{+\infty} g(\bar{F}(x)) dx, \quad (1.4)$$

where $\bar{F}(x) = 1 - F(x)$.

When X is non-negative, the first term of the right hand side disappears. An alternative expression of (1.4) provided g is differentiable is, due to, e.g., Jones and Zitikis (2003),

$$\rho_g(X) = \int_0^1 Q_{1-\alpha}(X) dg(\alpha) = \int_0^1 Q_\alpha(X) g'(1 - \alpha) d\alpha. \quad (1.5)$$

The distortion risk measure in general satisfies translation invariance, positive homogeneity, monotonicity, and additivity for comonotonic⁵ losses. Moreover, if g is concave, the risk measure becomes *coherent* by satisfying subadditivity as well; see Wirth and Hardy (2000) and Dhaene *et al.* (2006).

Using the distortion risk measure approach, we can easily express some popular risk measures through corresponding distortion function g . For instance, the distortion functions that correspond to VaR and CTE are, respectively,

⁵See (1.7) for definition

$$g^{VaR}(t) = \begin{cases} 1 & \text{if } t > 1 - \alpha \\ 0 & \text{if } t < 1 - \alpha \end{cases}$$

and

$$g^{CTE}(t) = \begin{cases} 1 & \text{if } t > 1 - \alpha \\ \frac{t}{1-\alpha} & \text{if } t < 1 - \alpha \end{cases}$$

For the CTE case, as an example, one can obtain $CTE_\alpha(X)$ for a nonnegative loss X using $g^{CTE}(t)$ and (1.4):

$$\begin{aligned} \int_0^\infty g^{CTE}(\bar{F}(x))dx &= \int_0^{Q_\alpha} 1dx + \int_{Q_\alpha}^\infty \frac{1-F(x)}{1-\alpha}dx \\ &= Q_\alpha + \frac{1}{1-\alpha} \int_{Q_\alpha}^\infty (1-F(x))dx \\ &= \frac{1}{1-\alpha} \int_{Q_\alpha}^\infty xf(x)dx \\ &= E[X|X > Q_\alpha(X)] \end{aligned}$$

Wang (2000) proposed an alternative risk measure by arguing that both the VaR and the CTE give no considerations to losses below the threshold α , making risk managers overlook medium-level losses. The Wang transform (WT) risk measure is defined by

$$g^{WT}(t) = \Phi(\Phi^{-1}(t) - \Phi^{-1}(\alpha)),$$

where $\Phi(\cdot)$ is the distribution function of the standard normal. It is easy to see that the CTE and the WT are coherent whereas the VaR is not because g^{VaR} is not concave. Another observation is that the WT measure reflects the risk characteristics over the whole domain of the risk random variable as well as the tail region as suggested by his original paper.

1.4 The Conditional Tail Expectation

VaR has been particularly popular in the past decade partly due to its simplicity. Jorion (1997) offers a comprehensive discussion on VaR; it also has been employed by regulatory bodies including the Basel committee on Banking Supervision and the SEC, as well as some commercial risk management systems such as the RiskMetrics by J.P. Morgan and RAROC by Bankers Trust.

Despite its widespread use, recent developments on risk measures identified a problem associated with VaR. In particular it can be super-additive, which means that the VaR of the sum of component risks can be greater than the sum of VaRs of each component; see Wirch and Hardy (1999) for details. In contrast, the CTE, another simple tail measure, is known to be superior to VaR based on the coherency criteria, as shown in Artzner *et al.* (1999). The CTE has become a new standard in both regulation and risk management. For example, the Canadian statutory balance sheet provisions on segregated funds (known as variable annuities in the US) is now computed based on the CTE at 95%.

From an academic perspective however there still is much room for research on the CTE at various fronts. On the parametric side, for example, we need to know how to compute the CTE for a fitted distribution. If we cannot compute the CTE analytically then large-scale Monte Carlo simulations will be needed to estimate this risk measure accurately because we are concerned with the tail of loss distributions. Several papers derive the analytic CTE formulas for various distributions. Panjer (2002) developed CTE formula and its decomposition for the normal case, Landsman and Valdez (2003) generalized the CTE formula to elliptical distributions. Hardy (2003) provided CTE

formula for the regime-switching lognormal (RSLN) model. The CTE formulas for the exponential dispersion model and a multivariate gamma model are developed by Landsman and Valdez (2005) and Furman and Landsman (2005), respectively. Note that because the quantile threshold must be computed first to get the CTE, as indicated by its definition, simulations are inevitable for the quantile (or VaR) part if it is not available analytically.

On the nonparametric side we are concerned with accurate estimation of the CTE and its variability. Even though the empirical CTE will converge to the true one as one increases the sample size, actuaries often find themselves left with relatively small samples in practice. This might be due to the nature of the data such as rare natural disaster events. Small sample situations also arise even though parametric models are available if models are large and complicated. Risk management in life insurance requires stochastic simulations using an office model. If the company's portfolio consists of millions of policies, which is not unusual for large insurers, these simulations may be too slow to use a large number of scenarios; it is not uncommon that projecting future cash flows for one scenario takes about 30 minutes for a single block of product in practice, and aggregating several blocks takes proportionally longer⁶. The practical constraints on the feasible number of Monte Carlo simulations are well known in the actuarial industry (see Christiansen (1998) and Chueh (2002) for example), and consequently estimating tail risk measures from small samples is of interest to actuaries in practice. A more detailed explanation on this issue is given in the first section of Chapter 3 and 5. For small samples, there are statistical techniques available to estimate the CTE and its variance, where no parametric information on

⁶In Chapter 5 we support this evidence with a real industry model.

the parent distribution is required. The bootstrap or the kernel density estimation would be examples in this category.

The insurance regulatory framework is moving from a rule-based format to a principle-based one, and from book values to market values, making quantitative risk management techniques more relevant than ever. Considering its simplicity, coherency, and popularity the CTE seems set to become the standard risk measure in the insurance sector.

1.5 Modelling Dependency

In illustrating risk capital in the first section the aggregate loss has been defined as $S = X_1 + \dots + X_n$ where X_i represents the loss of i -th business line. No specific assumptions, such as independence or identical distribution has been made on the X_i 's. In practice they are hardly independent or identically distributed; rather we observe dependence among risks everyday in the financial world. For instance, different business units are exposed to the common factor of economic recession or inflation; two individuals, if married, can be subject to the same accident in insurance; and a single big event can have a huge impact on several different business units as the events of 9/11 showed us. Therefore both the marginals and the dependence among them needs to be studied in order to compute the risk capital, such as $\rho(S) = E[S|S > Q_\alpha(S)]$ under the CTE risk measure.

For a risk capital discussion, studying dependency among marginal risks has some

challenging aspects. In practice the joint distribution of component losses are unavailable in most cases while marginal distributions are relatively easy to obtain. Thus we need techniques that allows us to construct the joint distribution from known marginal distributions. One way to do this is to use copulas. Another issue is that we are generally more interested in the right tail region than we are in the middle region because insolvency occurs only when the aggregate loss exceeds a certain, usually fairly large, threshold. There are some parametric distribution classes where the CTE of the aggregate loss can be analytically expressed; see, e.g., Landsman and Valdez (2003) and Landsman and Valdez (2005). If the given data does not fit into this class one would rely on other methods such as copulas.

In this section several well-known mathematical concepts are presented including copulas, comonotonicity, and tail dependence for future discussions.

A n -dimensional copula is defined to be a cumulative distribution function (c.d.f.) $C : [0, 1]^n \rightarrow [0, 1]$ with uniform marginal distributions. If the joint distribution $F(X_1, \dots, X_n)$ and each marginal $F_i(X_i)$ is given, the dependence structure between X_i 's are described through the copula function C and the copula is independent of the marginal distributions:

$$F(X_1, \dots, X_n) = C(F_1(X_1), \dots, F_n(X_n)) \quad (1.6)$$

When the marginal distributions are continuous, C is unique, from Copulas have been proved to be an effective tool in understanding and solving dependence problems. Standard texts include Nelsen (1999) and Joe (1997); also see Embrechts *et al.* (2003) and Frees and Valdez (1998) for the application of copulas in risk management and insurance, respectively.

In connection with copulas, the comonotonicity concept is useful for dependence studies. We call an array of random variables (X_1, \dots, X_n) comonotonic if there exists a random variable Z and increasing functions $h_i(x)$ such that

$$(X_1, \dots, X_n) = (h_1(Z), \dots, h_n(Z)), \quad (1.7)$$

where the equality holds in distribution. When we consider the sum of (dependent) individual losses $S = \sum X_i$ the worst scenario is that the joint distribution is comonotonic. In other words, the risk is greatest when all the individual losses have a comonotonic copula. The research on comonotonicity is an active field in actuarial profession and is still expanding. Some recent papers include Dhaene *et al.* (2002b), Dhaene *et al.* (2002a), and Dhaene *et al.* (2006). The last paper proposes bounds (and therefore an approximation) for the dependent random sum.

1.6 Capital Allocation

Once a capital requirement has been established based on a risk measure, one of the following tasks is to allocate it back to each business unit or risk component. The allocation of the capital seems to be the natural reverse process of the total required capital calculation but it has its own challenges, mostly due to the nature of the dependence structures of the combined risks. In the business world, breaking down the total capital into its components can be of interest to the managers of each business line and the executives of the company, though it is not the primary concern of regulators. The main motivations for capital allocation are the following, which are mutually not exclusive:

Absolute comparison Since capital is costly to hold, it is natural to allocate it to each line to redistribute its cost across lines. This view implies different returns on capital for each line and the management is often able to answer if a line is worth to keep or not by comparing, say, the return on capital of a line to company's hurdle rate

Relative comparison As capital is defined as a tail risk measure of the whole company, one may wish to split this capital to assess the riskiness of each line's position and compare one to another.

Diversification study The allocation also can serve as a tool to study the diversification effect. For example, adding or dropping a line will affect all the other lines and the direction and the magnitude of the changes can give useful information on the dependency structure

Compensation scheme Allocation of capital may provide the basis of each manager's performance, which in turn is linked to their compensations

Pricing basis Insurers may want to use the allocation in pricing, even though there are other considerations in pricing such as market competition and company's strategic planning. A line with an excessive capital would have to produce a larger profit by increasing the product margin.

Probably the most prominent use of capital allocation is the calculation of return on capital (ROC) measured by line where the ROC is defined to be the ratio of the line's profit to the line's allocated capital. The ROC can be used in comparing relative profitability, e.g., which line is more profitable?, as well as in the absolute profitability of a line, e.g., is this line a value-adder? Compensations of line managers can be

linked to the ROC in the same context; a line manager who achieved a higher return will receive a larger bonus.

Consider again a multi-line insurer whose aggregate loss is given by $S = \sum_1^n X_i$. The allocation problem is the task to assign the total capital, denoted by $\rho(S)$, to each business line using a suitable rule. Mathematically, an allocation is a mapping from a n -dimensional random vector to a n -dimensional real-valued vector.

In responding to the need to construct sensible capital allocation methods, researchers have proposed a set of axioms that any desirable allocation method is expected to satisfy. The list of axioms themselves are however still under scrutiny because the capital allocation itself is a relatively new phenomenon. Different authors have proposed different sets of axioms, while some axioms appear repeatedly. For axiomatic approach to capital allocation see Denault (2001), Hesselager and Anderson (2002), and Kalkbrener (2005). Axioms below are adapted from Valdez and Chernih (2003), which in turn were elaborated from Denault (2001) and Hesselager and Anderson (2002).

Definition 1.4 (Fair allocation axioms) *Suppose that a company has n business lines and each line's loss is represented by X_i , $i = 1, \dots, n$, with the aggregate loss $S = \sum_1^n X_i$. If the allocated capital for line i , denoted by AC_i , satisfies the following axioms, the allocation rule is said to be fair.*

1. Full allocation *The allocated capitals add up to the total capital.*

$$\rho(S) = \sum_1^n AC_i \tag{1.8}$$

- 2. No undercut** Let $\{X'_1, X'_2, \dots, X'_k\}$ be an arbitrary subset of $\{X_1, X_2, \dots, X_n\}$, and AC'_i , $i = 1, \dots, k$, be the allocated capital based on this subset. Then

$$\sum_1^k AC'_i \leq \rho\left(\sum_1^k X'_i\right) \quad (1.9)$$

- 3. Symmetry** If two lines have the same allocated capital for any collection of lines within the company, those two lines should have the same allocated capital under the whole company.
- 4. Consistency** The allocated capital for one line should not depend on the level at which allocation occurs.

Axioms above are defined rather informally for a quick introduction, and the symmetry and the consistency can be confusing without further explanations. We provide more formal definitions with explanation in Chapter 6, because a new allocation method developed therein will be examined in detail against these axioms.

There are various specific allocation methods available; some satisfy all the axioms while others do not. For example, the CTE of the aggregate loss, $E[S|S > Q_\alpha(S)]$, can be decomposed in a simple additive manner:

$$E[S|S > Q_\alpha(S)] = \sum_1^n E[X_i|S > Q_\alpha(S)],$$

which was first observed by Overbeck (2000). Thus the allocated capital for line i is

$$AC_i = E[X_i|S > Q_\alpha(S)]$$

This allocation is also known to satisfy the first three axioms; see Panjer (2002) for its properties⁷. The reason why this allocation looks natural is that the allocation rule is inherent in the risk measure CTE. Actually this allocation is an example

⁷We show that it also satisfies the consistency axiom in Chapter 6.

of the Euler allocation principle, which also gives a similar decomposition for the aggregate VaR; see Section 6.3 of McNeil *et al.* (2005) and references therein. Explicit expression of the CTE allocation for parametric multivariate models have been derived by several authors. Panjer (2002) showed this for the multivariate normal case and related it to the CAPM. Landsman and Valdez (2003) provided the CTE allocation for the elliptical distribution class which contains the normal as a special case. Furman and Landsman (2005) and Cai and Li (2005) derived the CTE allocation formula for a multivariate gamma distribution and the Phase-type distribution class, respectively.

There are other allocation methods available. For instance, Wang (2002) recommended an allocation using the exponential tilting. More recently Valdez and Chernih (2003) extended Wang's idea to the elliptical distribution class. Note that all these methods, including the CTE allocation, depend on the property of the risk measure $\rho(\cdot)$. This is however not a mandate condition for an allocation and there is an allocation method that is independent of $\rho(\cdot)$ (the only information needed is the total capital amount), as explained in Goovaerts *et al.* (2005).

1.7 Overview of Thesis

In this chapter we have motivated the need for a good risk measure as a tool to determine the risk capital of financial entities. Recognizing the conditional tail expectation (CTE) as a preferred risk measure, we focus on this risk measure throughout the thesis. We have also observed that there are challenging and interesting topics around quantitative modeling, including copula and capital allocation. There are other tools

that can be, and are being, used for this purpose, such as extreme value theory (EVT) and financial time series models. We decided to forego their introduction in this chapter because they are not directly related to this thesis.

In Chapter 2 the conditional tail expectation and its higher order moments are derived for the exponential distribution family (EF) class and its transformed distributions.

Chapter 3 is devoted to nonparametric estimation of the CTE and VaR, and their biases using the bootstrap method, for small sample sizes. For variability of the estimated risk measure, Chapter 4 provides an alternative technique using the influence function of the bootstrapped L-estimator. Numerical examples are presented to illustrate the theoretical findings. Together these two chapters give us a new method to estimate the mean squared error for the bootstrapped CTE, and this is examined through an industry-scaled model in Chapter 5.

Inspired by the solvency exchange option idea, a new capital allocation method is proposed in Chapter 6. This new method is similar to the CTE in its form, but is independent of the risk measure used for the total capital determination. A closer look on the proposed method prompts a critique on one of the fair allocation axioms; in particular, we argue that the no undercut axiom may not be well-aligned with the economic reality.

Chapter 2

Conditional tail moments of the EF and its related distributions

2.1 Introduction

When a random variable X is continuous the conditional tail expectation (CTE) of $T(X)$, with threshold $Q_\alpha = F_X^{-1}(\alpha)$, is defined as

$$CTE_\alpha(T(X)) = E(T(X)|X > Q_\alpha) = \frac{\int_{Q_\alpha}^{\infty} T(x) f(x) dx}{\bar{F}(Q_\alpha)}, \quad (2.1)$$

when it exists, with $\bar{F}(x) = 1 - F(x)$. This form is more general than the usual CTE definition given in (1.3) and is employed here for later developments. $T(X)$ is assumed to be a monotone continuous function of X ; it does not need to be an increasing function in theory. For instance, setting $T(X) = X$ produces the usual CTE and $T(X) = X^k$ gives the k -th moments for integral k . For a discrete random variable X with probability mass function $f(x)$ taking values on $x = 0, 1, \dots$, the

CTE is similarly defined by

$$CTE_\alpha(T(X)) = E(T(X)|X > Q_\alpha) = \frac{\sum_{k=Q_\alpha+1}^{\infty} T(k)f(k)}{\sum_{k=Q_\alpha+1}^{\infty} f(k)}, \quad (2.2)$$

where quantile Q_α is an integer satisfying $Q_\alpha = \inf\{x|F_X(x) \geq \alpha\}$

The conditional tail variance of $T = T(X)$ and its covariance can be defined by

$$\begin{aligned} CTVar_\alpha(T) &= E[(T - CTE_\alpha(T))^2|X > Q_\alpha] \\ &= CTE_\alpha(T^2) - (CTE_\alpha(T))^2, \end{aligned}$$

and

$$\begin{aligned} CTCov_\alpha(T_i, T_j) &= E[(T_i - CTE_\alpha(T_i))(T_j - CTE_\alpha(T_j))|X > Q_\alpha] \\ &= CTE_\alpha(T_i T_j) - CTE_\alpha(T_i)CTE_\alpha(T_j) \end{aligned}$$

respectively, where each T, T_i and T_j is a function of X . The same quantities for discrete random variables can be defined similarly.

Knowing the analytic expression of the CTE for a given parametric distribution is important because otherwise one must rely on Monte Carlo simulation to compute this risk measure, which may be costly. Thus finding the conditional tail expectation for parametric distributions has been an active topic in risk management and actuarial science. Landsman and Valdez (2003) derived the CTE formula for elliptical distributions, which is a class containing the normal distribution. Later the same authors developed CTE formula for the exponential dispersion model (EDM) in Landsman and Valdez (2005). Hardy (2003) provided the CTE for the regime-switching lognormal (RSLN2) model, and the CTE for a multivariate gamma model

was developed by Furman and Landsman (2005). All the discussion in these papers, however, have been focused on the first moment, or CTE. The second tail moment was considered in Valdez (2004) as a variability index for tail area beyond quantiles. It can also be motivated through variance of the CTE. When the threshold is constant the CTE is closely linked to the mean excess loss, defined by $E[X - d|X > d]$, thus higher tail moments can be used in the pricing of insurance with a deductible; see Section 2.6 of Klugman *et al.* (1998).

This chapter proposes a new formula for the conditional tail moments for the exponential family (EF) class and its transformed distributions. Both the EF and the EDM extend the natural exponential family class, and they share many common distributions such as Normal, Gamma, Poisson, to name a few, there are members in the EF class but not in the EDM class, such as the Lognormal distribution. Together with the transformed distributions discussed in Section 4 we have conditional tail moments for a wide coverage of distributions used for actuarial modeling.

Section 2 introduces and summarizes some properties of the exponential family; in Section 3 formulas are derived for the conditional tail moments of the EF; conditional tail moments of other transformed distributions are illustrated in Section 4 through examples.

2.2 The Exponential Family

The exponential family (EF) of distributions is commonly used in statistical estimation and decision theory. It includes many familiar distributions such as the Normal, multivariate Normal, Gamma, Inverse Gaussian, Poisson, and Binomial distributions.

The lognormal distribution is also included in the exponential family. Standard texts on the EF include Casella and Berger (2003), Lehmann (1983), and Brown (1986). Following the notation in Lehmann (1983), the density of an s -parameter exponential family with parameter vector $\theta = (\theta_1, \dots, \theta_s)$ is defined by

$$f(x; \theta) = \exp \left\{ \sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta) \right\} c(x) \quad (2.3)$$

with respect to the Lebesgue measure and the counting measure for continuous and discrete distributions, respectively. The function $c(x)$ does not depend on parameters θ_i .

Often the density in (2.3) is converted to a canonical form through reparameterization, $\eta_i(\theta) = \eta_i$, to result in

$$f(x; \eta) = \exp \left\{ \sum_{i=1}^s \eta_i T_i(x) - A(\eta) \right\} c(x) \quad (2.4)$$

where $\eta = (\eta_1, \dots, \eta_s)$ is the natural parameter vector and $T = (T_1, \dots, T_s)$ is a natural sufficient statistic. $A(\eta)$ is a constant assumed to be differentiable. Note that $c(x)$ in (2.3) and (2.4) are identical. One of the properties of the canonical parametrization is that the set

$$\mathcal{H} = \left\{ \eta = (\eta_1, \dots, \eta_s) : \int_{-\infty}^{\infty} \exp \left[\sum_{i=1}^s \eta_i T_i(x) - A(\eta) \right] c(x) < \infty \right\}$$

is convex. (The integral is replaced by a summation for discrete cases) Also, it is known that if there are linear constraints among η_i 's or T_i 's, the number s in (2.4) can be reduced. For the theoretical development of the EF, however, the presentation (2.4) is most convenient when it is minimal, that is there should be no linear constraints among η_i 's or T_i 's. These are important properties for standard estimation procedures but are not our primary concern here. This chapter focuses on conditional

tail moments for distributions known to be in the EF. Let us take the normal distribution as an example.

Example 2.1 (Normal distribution) *The density of the normal distribution with mean μ and variance σ^2 is*

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < \infty,$$

or, in the canonical form,

$$f(x; \eta_1, \eta_2) = \exp \left\{ \eta_1 x^2 + \eta_2 x + \frac{\eta_2^2}{4\eta_1} + \frac{1}{2} \log(-2\eta_1) \right\} \frac{1}{\sqrt{2\pi}}, \quad (2.5)$$

with following parameterization:

$$\begin{aligned} \eta_1 &= -1/(2\sigma^2), \quad \eta_2 = \mu/\sigma^2; \quad T_1(X) = X^2, \quad T_2(X) = X; \quad c(x) = 1/\sqrt{2\pi}; \\ A(\eta) &= \mu^2/(2\sigma^2) + \log \sigma = -\frac{\eta_2^2}{4\eta_1} - \frac{1}{2} \log(-2\eta_1). \end{aligned}$$

The natural parameter space is given by $\mathcal{H} = \{\eta = (\eta_1, \eta_2) : \eta_1 < 0, -\infty < \eta_2 < \infty\}$, which is convex, and there is no linear constraint either between η_1 and η_2 , or between T_1 and T_2 .

When expressed in the canonical form, we have the following well known results of the EF; see e.g., Lehmann (1983).

1. The order of integration and differentiation can be switched:

$$\frac{\partial}{\partial \eta_i} \int f(x) dx = \int \frac{\partial}{\partial \eta_i} f(x) dx,$$

where $f(x) = f(x; \eta)$ defined in (2.4). In fact $\int f(x) dx$ has derivatives of all orders and these can be obtained by differentiating inside the integral sign.

2. For any $i = 1, \dots, s$ we have

$$E(T_i) = \frac{\partial}{\partial \eta_i} A(\eta), \quad (2.6)$$

which is finite whenever $A(\eta)$ is differentiable. Often one of T_i 's equals to X , so that we get $E(T_i) = E(X)$, but this is not always true, as seen in the Lognormal case.

3. For any $i, j = 1, \dots, s$

$$Cov(T_i, T_j) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} A(\eta). \quad (2.7)$$

If $i = j$ this reduces to

$$Var(T_i) = \frac{\partial^2}{\partial \eta_i^2} A(\eta).$$

4. Assuming the moment generating function (m.g.f.) exists for a given random loss in some neighborhood $\sum u_i^2 < \delta$, the mgf of T in the exponential family is $M_T(u) = e^{A(\eta+u) - A(\eta)}$ and the cumulant generating function (c.g.f.) is $K_T(u) = \log(M_T(u)) = A(\eta + u) - A(\eta)$.

For the normal distribution case it is easy to verify these properties using the canonical form in (2.5). For instance,

$$E(T_1) = E(X^2) = \frac{\partial}{\partial \eta_1} A(\eta) = \left(\frac{\eta_2}{2\eta_1} \right)^2 - \frac{1}{2\eta_1} = \mu^2 + \sigma^2,$$

$$E(T_2) = E(X) = \frac{\partial}{\partial \eta_2} A(\eta) = -\frac{\eta_2}{2\eta_1} = \mu,$$

and

$$Var(T_2) = Var(X) = \frac{\partial^2}{\partial \eta_2^2} A(\eta) = -\frac{1}{2\eta_1} = \sigma^2.$$

With some algebra we can also verify, for fixed σ , that

$$M_T(u) = e^{A(\eta+u) - A(\eta)} = \exp(\mu u + \sigma^2 u^2 / 2).$$

2.3 Conditional tail moments of the EF

2.3.1 When generating function exists

In this section we give a general formula to calculate the conditional tail moments of a member of the EF using the moment generating function (m.g.f.) and the cumulant generating function (c.g.f.) provided that they exist. Denote $T = (T_1, \dots, T_s)^t$ and $u = (u_1, \dots, u_s)$ in this section. We define the conditional tail m.g.f. and the conditional tail c.g.f. by, respectively,

$$M_T^*(u|d) = E(e^{uT}|X > d) \quad (2.8)$$

and

$$K_T^*(u|d) = \log E(e^{uT}|X > d), \quad (2.9)$$

where d is a given threshold. The conditional tail generating functions exist whenever the original generating functions do.

Lemma 2.2 *Suppose that X , a member of the EF defined in (2.4), has the conditional tail c.g.f. defined in (2.9). Then, for a given d , it is expressed by*

$$K_T^*(u|d) = A(\eta + u) - A(\eta) + \log \bar{F}(d; \eta + u) - \log \bar{F}(d; \eta). \quad (2.10)$$

Proof:

$$\begin{aligned} M_T^*(u|d) &= \frac{1}{\bar{F}(d; \eta)} \int_d^\infty e^{u_1 T_1 + \dots + u_s T_s} e^{\sum \eta_i T_i - A(\eta)} c(x) dx \\ &= \frac{e^{A(\eta+u) - A(\eta)}}{\bar{F}(d; \eta)} \int_d^\infty e^{\sum (\eta_i + u_i) T_i - A(\eta+u)} c(x) dx \\ &= M_T(u) \frac{\bar{F}(d; \eta + u)}{\bar{F}(d; \eta)}, \text{ or} \\ &= \frac{e^{A(\eta+u)} \bar{F}(d; \eta + u)}{e^{A(\eta)} \bar{F}(d; \eta)}. \end{aligned}$$

The result follows immediately by taking the logarithm of the last equality in the above equation. Q.E.D.

We can now generate the conditional tail moments of any order by repeated differentiation at $u = 0$. We focus on the first two conditional moments because it is hard to motivate the usage of higher moments. We cannot, however, apply this for distributions which do not have m.g.f.'s (e.g., the Lognormal distribution). In particular the *CTE* and *CTCov* (and thus *CTVar*) are determined by setting $d = Q_\alpha$,

$$\frac{\partial}{\partial u_i} K_T^*(u|Q_\alpha)|_{u=0} = CTE_\alpha(T_i)$$

and

$$\frac{\partial^2}{\partial u_i \partial u_j} K_T^*(u|Q_\alpha)|_{u=0} = CTCov_\alpha(T_i, T_j).$$

The following formulas provide explicit expressions for *CTE* and *CTCov* when the corresponding conditional tail c.g.f.'s exist.

Lemma 2.3 *Suppose that X is a member of the EF and has a conditional tail c.g.f.. Then the CTE of $T_i(X)$ is given by*

$$E[T_i|X > Q_\alpha] = \frac{\partial}{\partial \eta_i} A(\eta) + \frac{\partial}{\partial \eta_i} \log \bar{F}(x; \eta)|_{x=Q_\alpha}$$

Proof: Using the identity

$$\frac{\partial}{\partial u_i} A(\eta + u) = \frac{\partial}{\partial \eta_i} A(\eta + u),$$

we have

$$\begin{aligned} \frac{\partial}{\partial u_i} K_T^*(u|d)|_{u=0} &= \frac{\partial}{\partial u_i} [A(\eta + u) - A(\eta) + \log \bar{F}(d; \eta + u) - \log \bar{F}(d; \eta)]|_{u=0} \\ &= \frac{\partial}{\partial \eta_i} A(\eta) + \frac{\partial}{\partial \eta_i} \log \bar{F}(d; \eta) \end{aligned}$$

The assertion follows by setting $d = Q_\alpha$. Q.E.D.

We choose expression $\frac{\partial}{\partial \eta_i} \log \bar{F}(x; \eta)|_{x=Q_\alpha}$ instead of $\frac{\partial}{\partial \eta_i} \log \bar{F}(Q_\alpha; \eta)$ to make it clear that the differentiation should be done before evaluating at $x = Q_\alpha$. If the order is switched one gets a totally different result because Q_α is generally a function of η .

Using the identity in (2.6) one could express the CTE alternatively as

$$E[T_i | X > Q_\alpha] = E(T_i) + \frac{\partial}{\partial \eta_i} \log \bar{F}(x; \eta)|_{x=Q_\alpha}$$

The result is similar to the CTE formula of the exponential dispersion model (EDM) developed by Landsman and Valdez (2005). They treated the quantile as a constant in the proof of Theorem 1, but that is incorrect because a quantile is a function of the unknown parameters. Their result holds nevertheless. Also their proof of Theorem 1 can be shortened by defining the conditional m.g.f. of the EDM, which can be obtained easily, for both reproductive and additive form.

If we define the loss as the left side of the distribution, which is a standard practice in banking industry, the corresponding CTE formula becomes

$$E[T_i | X < Q_\alpha] = E(T_i) + \frac{\partial}{\partial \eta_i} \log F(x)|_{x=Q_\alpha}.$$

In fact the trimmed mean can be similarly derived for the whole EF class.

Lemma 2.4 *Suppose X is a member of the EF and $K_T^*(u|d)$ exists for X . Then $CTCov(T_i, T_j)$ can be expressed as*

$$CTVar_\alpha(T_i, T_j) = \frac{\partial}{\partial \eta_j} \frac{\partial}{\partial \eta_i} A(\eta) + \frac{\partial}{\partial \eta_j} \frac{\partial}{\partial \eta_i} \log \bar{F}(x; \eta) \Big|_{x=Q_\alpha}.$$

Proof:

$$\begin{aligned} \frac{\partial^2}{\partial u_i \partial u_j} K_T^*(u|d) \Big|_{u=0} &= \frac{\partial^2}{\partial u_j \partial u_i} [A(\eta + u) - A(\eta) + \log \bar{F}(d; \eta + u) - \log \bar{F}(d; \eta)] \Big|_{u=0} \\ &= \frac{\partial}{\partial \eta_j} \left[\frac{\partial}{\partial \eta_i} A(\eta + u) + \frac{\partial}{\partial \eta_i} \log \bar{F}(d; \eta + u) \right] \Big|_{u=0} \\ &= \frac{\partial}{\partial \eta_j} \frac{\partial}{\partial \eta_i} A(\eta) + \frac{\partial}{\partial \eta_j} \frac{\partial}{\partial \eta_i} \log \bar{F}(d; \eta). \end{aligned}$$

Again setting $d = Q_\alpha$ gives the result. If $i = j$, the formula reduces to the conditional tail variance. Q.E.D.

As for the first moment, there is an alternative expression for the conditional tail covariance using (2.7):

$$CTVar_\alpha(T_i, T_j) = Cov(T_i, T_j) + \frac{\partial}{\partial \eta_j} \frac{\partial}{\partial \eta_i} \log \bar{F}(x; \eta) \Big|_{x=Q_\alpha}.$$

Example 2.5 (Normal distribution (continued)) *From (2.5), we take the derivative with regard to η_2 to get $CTE(X|X > Q_\alpha)$ since $T_2 = X$. First note that $\frac{\partial}{\partial \eta_2} f(x) = (x + \frac{\eta_2}{2\eta_1})f(x)$ for the normal density $f(x)$. To calculate $\frac{\partial}{\partial \eta_2} \log \bar{F}(x)$,*

$$\begin{aligned} \frac{\partial}{\partial \eta_2} \bar{F}(x) &= \frac{\partial}{\partial \eta_2} \int_x^\infty f(t) dt = \int_x^\infty \frac{\partial}{\partial \eta_2} f(t) dt \\ &= \int_x^\infty (t + \eta_2/(2\eta_1)) f(t) dt \\ &= \frac{1}{2\eta_1} \int_x^\infty (2\eta_1 t + \eta_2) f(t) dt \\ &= \frac{1}{2\eta_1} f(t) \Big|_x^\infty \\ &= -\frac{1}{2\eta_1} f(x). \end{aligned}$$

Thus

$$\begin{aligned}\frac{\partial}{\partial \eta_2} \log \bar{F}(x) \Big|_{x=Q_\alpha} &= \frac{-1}{2\eta_1} f(Q_\alpha) / \bar{F}(Q_\alpha) \\ &= \sigma^2 h_X(Q_\alpha),\end{aligned}$$

where $h_X(x)$ is the hazard function of $N(\mu, \sigma^2)$. Therefore for the normal distribution

$$CTE_\alpha(T_2) = CTE_\alpha(X) = \mu + \sigma^2 h_X(Q_\alpha),$$

or in the standardized form, $\mu + \sigma h_Z(z_\alpha)$, where $Z = (X - \mu) / \sigma$ and $z_\alpha = (Q_\alpha - \mu) / \sigma$.

This result agrees with Panjer (2002) and Landsman and Valdez (2003).

For the CTVar,

$$\begin{aligned}\frac{\partial^2}{\partial \eta_2^2} \log \bar{F}(x) &= \frac{\partial}{\partial \eta_2} \left(\frac{\partial}{\partial \eta_2} \log \bar{F}(x) \right) \\ &= \frac{\partial}{\partial \eta_2} \left(\frac{-1}{2\eta_1} f(x) / \bar{F}(x) \right) \\ &= -\frac{1}{2\eta_1} \frac{\partial}{\partial \eta_2} \left(f(x) / \bar{F}(x) \right) \\ &= -\frac{1}{2\eta_1} \frac{\frac{\partial}{\partial \eta_2} f(x) \bar{F}(x) - f(x) \frac{\partial}{\partial \eta_2} \bar{F}(x)}{\bar{F}(x)^2} \\ &= -\frac{1}{2\eta_1} \frac{[(x + \frac{\eta_2}{2\eta_1})d(x)] \bar{F}(x) - f(x) [\frac{-1}{2\eta_1} f(x)]}{\bar{F}(x)^2} \\ &= -\frac{1}{2\eta_1} \left[(x + \frac{\eta_2}{2\eta_1}) h_X(x) + \frac{1}{2\eta_1} h_X(x)^2 \right] \\ &= \sigma^2 \left[(x - \mu) h_X(x) - \sigma^2 h_X(x)^2 \right],\end{aligned}$$

and from (2.4)

$$\begin{aligned}CTVar_\alpha(X) &= Var(X) + \frac{\partial^2}{\partial \eta_2^2} \log \bar{F}(x) \Big|_{x=Q_\alpha} \\ &= \sigma^2 \left[1 + (Q_\alpha - \mu) h_X(Q_\alpha) - \sigma^2 h_X(Q_\alpha)^2 \right].\end{aligned}$$

If the standardized version is preferred,

$$CTVar_\alpha(X) = \sigma^2 \left[1 + z_\alpha h_Z(z_\alpha) - h_Z(z_\alpha)^2 \right],$$

where $Z = (X - \mu)/\sigma$ and $z_\alpha = (Q_\alpha - \mu)/\sigma$.

Example 2.6 (Exponential distribution) Consider $X \sim \text{Exp}(\beta)$, with density $f(x; \beta) = \beta^{-1} \exp(-x/\beta)$. Due to its memoryless property, it is clear that $CTE_\alpha(X) = Q_\alpha + E(X) = Q_\alpha + \beta$ and $CTVar_\alpha(X) = \text{Var}(X) = \beta^2$. The density, expressed in the canonical form in (2.4), is

$$f(x; \eta) = \exp(\eta x + \log(-\eta)),$$

where $T(X) = X$, $\eta = -1/\beta$, $c(x) = 1$ and $A(\eta) = -\log(-\eta)$. Using $F(x) = 1 - \exp(\eta x)$,

$$CTE_\alpha(X) = \frac{\partial}{\partial \eta} A(\eta) + \frac{\partial}{\partial \eta} \log \bar{F}(x; \eta) \Big|_{x=Q_\alpha} = -1/\eta + Q_\alpha = \beta + Q_\alpha$$

and

$$CTVar_\alpha(X) = \frac{\partial^2}{\partial \eta^2} A(\eta) + \frac{\partial^2}{\partial \eta^2} \log \bar{F}(x; \eta) \Big|_{x=Q_\alpha} = 1/\eta^2 = \beta^2,$$

as expected.

Example 2.7 (Poisson distribution) The Poisson probability function with mean λ is $f(x) = e^{-\lambda} \lambda^x \frac{1}{x!}$, $x = 0, 1, \dots$, which rearranges to $f(x) = \exp(x \log \lambda - \lambda) \frac{1}{x!} = \exp(x\eta - e^\eta) \frac{1}{x!}$. Hence the canonical form is obtained by parametrization $\eta = \log \lambda$.

The mean is of course $A(\eta)' = e^\eta = \lambda$. Then from basic calculus, we have

$$\begin{aligned}
\frac{\partial}{\partial \eta} \bar{F}(x) &= \frac{\partial}{\partial \eta} \left[\sum_{k=x+1}^{\infty} \exp(k\eta - e^\eta) \frac{1}{k!} \right] \\
&= \sum_{k=x+1}^{\infty} \exp(k\eta - e^\eta) \frac{1}{k!} \times (k - e^\eta) \\
&= \sum_{k=x+1}^{\infty} \exp(k\eta - e^\eta) \frac{1}{(k-1)!} - e^\eta \sum_{k=x+1}^{\infty} \exp(k\eta - e^\eta) \frac{1}{k!} \\
&= e^\eta \sum_{k=x+1}^{\infty} \exp((k-1)\eta - e^\eta) \frac{1}{(k-1)!} - e^\eta \sum_{k=x+1}^{\infty} \exp(k\eta - e^\eta) \frac{1}{k!} \\
&= e^\eta \bar{F}(x-1) - e^\eta \bar{F}(x) \\
&= e^\eta f(x).
\end{aligned}$$

Noting that there is no explicit form for the distribution function of the Poisson, we define a discrete hazard function by $h(x) = f(x)/\bar{F}(x)$, $x = 0, 1, \dots$. Then

$$\frac{\partial \log \bar{F}(x)}{\partial \eta} = \frac{\frac{\partial \bar{F}(x)}{\partial \eta}}{\bar{F}(x)} = \frac{e^\eta f(x)}{\bar{F}(x)} = e^\eta h(x).$$

The CTE of the Poisson with integer threshold d is then

$$\begin{aligned}
E[X|X > d] &= E(X) + \left. \frac{\partial \log \bar{F}(x)}{\partial \eta} \right|_d \\
&= e^\eta + e^\eta h(d) \\
&= \lambda + \lambda h(d).
\end{aligned}$$

This is identical to the normal case when $\mu = \sigma^2 = \lambda$. For the second moment we first note that, from its canonical density form,

$$\frac{\partial}{\partial \eta} f(x) = (x - e^\eta) f(x).$$

Then

$$\begin{aligned}
\frac{\partial^2}{\partial \eta^2} \log \bar{F}(x) &= \frac{\partial}{\partial \eta} \left[\frac{\partial \log \bar{F}(x)}{\partial \eta} \right] \\
&= \frac{\partial}{\partial \eta} \left[\frac{e^\eta f(x)}{\bar{F}(x)} \right] \\
&= \frac{[e^\eta f(x) + e^\eta \frac{\partial}{\partial \eta} f(x)] \bar{F}(x) - e^\eta f(x) \frac{\partial}{\partial \eta} \bar{F}(x)}{(\bar{F}(x))^2} \\
&= \frac{[e^\eta f(x) + e^\eta (x - e^\eta) f(x)] \bar{F}(x) - e^\eta f(x) e^\eta f(x)}{(\bar{F}(x))^2} \\
&= (xe^\eta + e^\eta - e^{2\eta}) \frac{f(x)}{\bar{F}(x)} - e^{2\eta} \left(\frac{f(x)}{\bar{F}(x)} \right)^2 \\
&= (xe^\eta + e^\eta - e^{2\eta}) h(x) - e^{2\eta} h(x)^2 \\
&= (x\lambda + \lambda - \lambda^2) h(x) - \lambda^2 h(x)^2
\end{aligned}$$

Finally we have

$$\begin{aligned}
CTVar_d(X) &= \lambda + (d\lambda + \lambda - \lambda^2) h(d) - \lambda^2 h(d)^2 \\
&= \lambda [1 + (d + 1 - \lambda) h(d) - \lambda h(d)^2],
\end{aligned}$$

which is a similar, but not identical, form to the normal case.

While the first conditional moments of all these examples are also given in Landsman and Valdez (2005) because they belong to the exponential dispersion family as well, the expressions for the moments of the second order are new.

2.3.2 When no generating function exists

In this subsection we prove that the same results hold for a member of the EF even when $K_T^*(u|d)$ does not exist. Thus the following two lemmas are more general results

than Lemma 2.3 and 2.4, but their proofs are longer because no generating function is assumed. For notational simplicity we suppress η and set $f(x; \eta) = f(x)$ and $F(x; \eta) = F(x)$ throughout this subsection.

Lemma 2.8 *Suppose X is a member of the EF given in (2.4). Then the CTE of $T_i(X)$ is given by*

$$E[T_i | X > Q_\alpha] = \frac{\partial}{\partial \eta_i} A(\eta) + \frac{\partial}{\partial \eta_i} \log \bar{F}(x) \Big|_{x=Q_\alpha}$$

Proof:

First it should be noted that the threshold does not have to be a quantile and it can be any function of $\eta = (\eta_1, \dots, \eta_s)$ or a constant. We start with

$$\int_d^\infty f(x) dx = \bar{F}(d), \quad (2.11)$$

and take the derivative with respect to η_i of both sides of this equation, assuming d is a constant.

Because the order of integration and differentiation can be switched for the EF, the left hand side becomes

$$\begin{aligned} \frac{\partial}{\partial \eta_i} \int_d^\infty f(x) dx &= \int_d^\infty \frac{\partial}{\partial \eta_i} f(x) dx \\ &= \int_d^\infty \frac{\partial}{\partial \eta_i} \log f(x) f(x) dx, \end{aligned}$$

leading to a new equation

$$\int_d^\infty \frac{\partial}{\partial \eta_i} \log f(x) f(x) dx = \frac{\partial \bar{F}(d)}{\partial \eta_i}. \quad (2.12)$$

If both sides are divided by $\bar{F}(d)$,

$$\frac{\int_d^\infty \frac{\partial}{\partial \eta_i} \log f(x) f(x) dx}{\bar{F}(d)} = \frac{1}{\bar{F}(d)} \frac{\partial \bar{F}(d)}{\partial \eta_i},$$

which in turn, by definition of the CTE, becomes

$$E\left(\frac{\partial}{\partial\eta_i} \log f(X)|X > d\right) = \frac{\partial}{\partial\eta_i} \log \bar{F}(d).$$

Now, because $\frac{\partial}{\partial\eta_i} \log f = T_i - \frac{\partial}{\partial\eta_i} A(\eta)$ for the exponential family defined in (2.4), the preceding equation reduces to

$$E\left(T_i - \frac{\partial}{\partial\eta_i} A(\eta)|X > d\right) = \frac{\partial}{\partial\eta_i} \log \bar{F}(d),$$

or

$$E(T_i|X > d) = \frac{\partial}{\partial\eta_i} A(\eta) + \frac{\partial}{\partial\eta_i} \log \bar{F}(d).$$

Finally setting $d = Q_\alpha$ completes the proof. Note that whenever $E(T_i)$ is finite, so is $E(T_i|X > Q_\alpha)$. The discrete case can be similarly proved. Q.E.D.

Lemma 2.9 *Suppose X is a member of the EF given in (2.4). Then $CTCov(T_i, T_j)$ can be expressed as*

$$CTCov_\alpha(T_i, T_j) = \frac{\partial^2}{\partial\eta_i \partial\eta_j} A(\eta) + \frac{\partial^2}{\partial\eta_i \partial\eta_j} \log \bar{F}(x)|_{x=Q_\alpha} \quad (2.13)$$

Proof:

The task here is to differentiate equation (2.12), this time with respect to η_j . The left side then becomes

$$\begin{aligned}
& \int_d^\infty \frac{\partial}{\partial \eta_j} \left[\frac{\partial}{\partial \eta_i} \log f(x) f(x) \right] dx \\
&= \int_d^\infty \frac{\partial^2}{\partial \eta_j \partial \eta_i} \log f(x) f(x) dx + \int_d^\infty \frac{\partial}{\partial \eta_j} \log f(x) \frac{\partial}{\partial \eta_i} \log f(x) f(x) dx \\
&= \int_d^\infty \frac{\partial^2}{\partial \eta_j \partial \eta_i} A(\eta) f(x) dx + \int_d^\infty (T_j - \frac{\partial}{\partial \eta_j} A(\eta))(T_i - \frac{\partial}{\partial \eta_i} A(\eta)) f(x) dx \\
&= -\frac{\partial^2}{\partial \eta_j \partial \eta_i} A(\eta) \int_d^\infty f(x) dx + \frac{\int_d^\infty (T_j - \frac{\partial}{\partial \eta_j} A(\eta))(T_i - \frac{\partial}{\partial \eta_i} A(\eta)) f(x) dx}{\bar{F}(d)} \bar{F}(d) \\
&= -\frac{\partial^2}{\partial \eta_j \partial \eta_i} A(\eta) \bar{F}(d) + E[(T_j - \frac{\partial}{\partial \eta_j} A(\eta))(T_i - \frac{\partial}{\partial \eta_i} A(\eta)) | X > d] \bar{F}(d) \\
&= -\bar{F}(d) \left(\frac{\partial^2}{\partial \eta_j \partial \eta_i} A(\eta) - E[(T_j - \frac{\partial}{\partial \eta_j} A(\eta))(T_i - \frac{\partial}{\partial \eta_i} A(\eta)) | X > d] \right),
\end{aligned}$$

where $\mu_j = \frac{\partial}{\partial \eta_j} A(\eta) = E(T_j)$, as given in (2.6).

Hence taking the derivative of (2.12) with respect to η_j yields

$$-\bar{F}(d) \left(\frac{\partial^2}{\partial \eta_j \partial \eta_i} A(\eta) - E[(T_j - \mu_j)(T_i - \mu_i) | X > d] \right) = \frac{\partial^2}{\partial \eta_j \partial \eta_i} \bar{F}(d),$$

or

$$E[(T_j - \mu_j)(T_i - \mu_i) | X > d] = \frac{\partial^2}{\partial \eta_j \partial \eta_i} A(\eta) + \frac{1}{\bar{F}(d)} \frac{\partial^2}{\partial \eta_j \partial \eta_i} \bar{F}(d).$$

By setting $d = Q_\alpha$, we have

$$CTE_\alpha[(T_j - \mu_j)(T_i - \mu_i)] = \frac{\partial^2}{\partial \eta_j \partial \eta_i} A(\eta) + \frac{1}{\bar{F}(Q_\alpha)} \frac{\partial^2}{\partial \eta_j \partial \eta_i} \bar{F}(x) \Big|_{x=Q_\alpha}. \quad (2.14)$$

As the final step,

$$\begin{aligned}
CTCov_\alpha(T_i, T_j) &= CTE_\alpha[(T_i - CTE_\alpha(T_i))(T_j - CTE_\alpha(T_j))] \\
&= CTE_\alpha[(T_i - \mu_i - \frac{\partial}{\partial \eta_i} \log \bar{F}(x)|_{Q_\alpha})(T_j - \mu_j - \frac{\partial}{\partial \eta_j} \log \bar{F}(x)|_{Q_\alpha})] \\
&= CTE_\alpha[(T_i - \mu_i)(T_j - \mu_j)] - \frac{\partial}{\partial \eta_i} \log \bar{F}(x)|_{Q_\alpha} \times \frac{\partial}{\partial \eta_j} \log \bar{F}(x)|_{Q_\alpha} \\
&= \frac{\partial^2}{\partial \eta_i \partial \eta_j} A(\eta) + \frac{\partial^2}{\partial \eta_j \partial \eta_i} \log \bar{F}(x) \Big|_{x=Q_\alpha}.
\end{aligned}$$

The last line is justified by (2.14) and the following equation:

$$\frac{\partial^2}{\partial \eta_j \partial \eta_i} \log \bar{F}(x) = -\frac{\partial}{\partial \eta_i} \log \bar{F}(x) \times \frac{\partial}{\partial \eta_j} \log \bar{F}(x) + \frac{1}{\bar{F}(x)} \frac{\partial^2}{\partial \eta_j \partial \eta_i} \bar{F}(x).$$

Note that when $i = j$ we recover the conditional tail variance formula

$$CTVar_\alpha(T_i) = Var(T_i) + \frac{\partial^2}{\partial \eta_i^2} \log \bar{F}(x) \Big|_{x=Q_\alpha}. \quad (2.15)$$

Q.E.D.

The conditional tail moments of the Lognormal distribution are interesting for two reasons. First, $T(X) \neq X$ and therefore it does not belong to the exponential dispersion family discussed in Landsman and Valdez (2005). Secondly there is no m.g.f. available. The Lognormal however is a member of the EF and plays an important role in risk management and finance because of its link to geometric Brownian motion in asset price modeling.

Example 2.10 (Lognormal distribution) *The density of the Lognormal distribution in canonical form is*

$$f(x) = \exp \left\{ \eta_1 (\log x)^2 + \eta_2 \log x + \frac{\eta_2^2}{4\eta_1} + \frac{1}{2} \log(-2\eta_1) \right\} \frac{1}{x\sqrt{2\pi}},$$

with the following parameterization:

$$\eta_1 = -1/(2\sigma^2), \eta_2 = \mu/\sigma^2; \quad T_1(X) = (\log X)^2, T_2(X) = \log X; \quad c(x) = \frac{1}{x\sqrt{2\pi}};$$

$$A(\eta) = \mu^2/(2\sigma^2) + \log \sigma = -\frac{\eta_2^2}{4\eta_1} - \frac{1}{2} \log(-2\eta_1),$$

$$\frac{\partial}{\partial \eta_2} A(\eta) = -\frac{\eta_2}{2\eta_1} = \mu, \text{ and}$$

$$\frac{\partial^2}{\partial \eta_2^2} A(\eta) = -\frac{1}{2\eta_1} = \sigma^2.$$

Now using Lemma 2.8 and Lemma 2.9 we can obtain $CTE_\alpha(T_1) = E((\log X)^2|X > Q_\alpha)$ or $CTE_\alpha(T_2) = E(\log X|X > Q_\alpha)$. Let us try $CTE(T_2)$ for this example.

To start with, it can be easily verified that $f'(x) = \frac{1}{x}(2\eta_1 \log x + \eta_2 - 1)f(x)$ and $\frac{\partial}{\partial \eta_2} f(x) = (\frac{\eta_2}{2\eta_1} + \log x)f(x)$ for the Lognormal density. To calculate $\frac{\partial}{\partial \eta_2} \log \bar{F}(x)$, first

$$\begin{aligned} \frac{\partial}{\partial \eta_2} \bar{F}(x) &= \int_x^\infty \frac{\partial}{\partial \eta_2} f(t) dt \\ &= \int_x^\infty (\log t + \eta_2/(2\eta_1)) f(t) dt \\ &= \frac{1}{2\eta_1} \int_x^\infty (2\eta_1 \log t + \eta_2) f(t) dt \\ &= \frac{1}{2\eta_1} \left[\int_x^\infty (2\eta_1 \log t + \eta_2 - 1) f(t) dt + \int_x^\infty f(t) dt \right] \\ &= \frac{1}{2\eta_1} \left[\int_x^\infty t f'(t) dt + \bar{F}(x) \right] \\ &= -\frac{1}{2\eta_1} x f(x). \end{aligned}$$

Thus

$$\left. \frac{\partial}{\partial \eta_2} \log \bar{F}(x) \right|_{Q_\alpha} = \left. \frac{\frac{\partial}{\partial \eta_2} \bar{F}(x)}{\bar{F}(x)} \right|_{Q_\alpha} = \frac{-\frac{1}{2\eta_1} Q_\alpha F(Q_\alpha)}{\bar{f}(Q_\alpha)} = \sigma^2 Q_\alpha h_X(Q_\alpha),$$

where $h_X(x)$ is the hazard function of $LN(\mu, \sigma^2)$. Therefore we have $CTE_\alpha(T_2) = CTE_\alpha(\log X) = \mu + \sigma^2 y h_X(x)$.

For the second moment,

$$\begin{aligned}
\frac{\partial^2}{\partial \eta_2^2} \log \bar{F}(x) &= \frac{\partial}{\partial \eta_2} \left(\frac{\partial}{\partial \eta_2} \log \bar{F}(x) \right) \\
&= \frac{\partial}{\partial \eta_2} \left(\frac{-x}{2\eta_1} F(x)/\bar{F}(x) \right) \\
&= -\frac{x}{2\eta_1} \frac{\partial}{\partial \eta_2} \left(F(x)/\bar{F}(x) \right) \\
&= -\frac{x}{2\eta_1} \frac{\frac{\partial}{\partial \eta_2} F(x) \bar{F}(x) - F(x) \frac{\partial}{\partial \eta_2} \bar{F}(x)}{\bar{F}(x)^2} \\
&= -\frac{x}{2\eta_1} \frac{[(\log x + \frac{\eta_2}{2\eta_1})F(x)] \bar{F}(x) - F(x) [\frac{-x}{2\eta_1} F(x)]}{\bar{F}(x)^2} \\
&= -\frac{x}{2\eta_1} \left[(\log x + \frac{\eta_2}{2\eta_1}) h_X(x) + \frac{x}{2\eta_1} h_X(x)^2 \right] \\
&= \sigma^2 \left[(\log x - \mu) x h_X(x) - \sigma^2 x^2 h_X(x)^2 \right],
\end{aligned}$$

and from (2.15)

$$\begin{aligned}
CTVar_\alpha(\log X) &= Var(\log X) + \frac{\partial^2}{\partial \eta_2^2} \log \bar{F}(x) \Big|_{Q_\alpha} \\
&= \sigma^2 \left[1 + (\log Q_\alpha - \mu) Q_\alpha h_X(Q_\alpha) - \sigma^2 Q_\alpha^2 h_X(Q_\alpha)^2 \right].
\end{aligned}$$

There are two comments for this specific example. First, unlike many cases with one of T_i 's being equal to X , the formulas lead to $CTE(\log X)$ or $CTE((\log X)^2)$, instead of $CTE(X)$ or $CTE(X^2)$. This is somewhat disappointing because we are usually more interested in the conditional tail moments of X rather than a function of X . In the next section $CTE(X^k) = E(X^k | X > Q_\alpha)$ for the lognormal will be calculated using a different method.

Secondly although direct calculation has been used to get $CTE(T_2)$ and $CTVar(T_2)$, there is an easier way to calculate these quantities in this example as following.

Since $W = \log X \sim N(\mu, \sigma^2)$,

$$\begin{aligned} CTE(T_2) &= E(\log X | X > Q_\alpha) = E(\log X | \log X > \log Q_\alpha) \\ &= CTE_{\log Q_\alpha}(W) \\ &= \mu + \sigma^2 h_W(\log Q_\alpha) \\ &= \mu + \sigma^2 Q_\alpha h_X(Q_\alpha). \end{aligned}$$

$E((\log X)^2 | X > Q_\alpha)$ can be determined in a similar fashion. Actually this type of technique provides an efficient way to determine the conditional tail moments for various distributions and will be further illustrated in the next section.

2.4 Transformed Distributions

Suppose we are interested in the conditional tail moments of Y , where Y is a function of X . In other words $Y = g(X)$ for some function g . If we assume g to be monotonic and increasing, then

$$\begin{aligned} \alpha &= Pr[Y \leq Q_\alpha(Y)] = Pr[g(X) \leq Q_\alpha(Y)] \\ &= Pr[X \leq g^{-1}(Q_\alpha(Y))] = Pr[X \leq Q_\alpha(X)]. \end{aligned}$$

The last equality holds because the probability should equal α for X as well. Thus the quantile is preserved under the transformation in the sense that $g^{-1}(Q_\alpha(Y)) =$

$Q_\alpha(X)$. Using this the conditional tail moment of Y is easily computed¹ as

$$\begin{aligned} CTE_\alpha(Y^k) &= E[Y^k|Y > Q_\alpha(Y)] \\ &= E[g(X)^k|X > Q_\alpha(X)] \\ &= CTE_\alpha(g(X)^k), \quad k = 1, 2, \dots, \end{aligned}$$

provided the moments exist. If g is a monotonic decreasing function, then

$$E[Y^k|Y > Q_\alpha(Y)] = E[g(X)^k|X < Q_{1-\alpha}(X)].$$

This simple tool is useful to compute the CTEs of various distributions transformed from the EF class where the CTEs are easily determined. For example, the Lognormal is a distribution transformed from the normal distribution; the transformed gamma distribution is from the gamma distribution; the Pareto distribution (with known location parameter) is from the shifted exponential distribution, to name a few. See the appendix of Klugman *et al.* (1998) for a survey of these transformed distributions.

In this section selected examples in this type of distributions are provided for illustration purposes.

Example 2.11 (Linear transformation) *When $Y = g(X) = aX + b$, $a > 0$, the*

¹We use $E[Y^k|Y > Q_\alpha(Y)]$ rather than $E[Y^k|X > Q_\alpha(X)]$ because often Y itself is a familiar random variable.

conditional tail moments of Y are calculated as follows:

$$\begin{aligned}
CTE_\alpha(Y^k) &= E((aX + b)^k | aX + b > Q_\alpha(Y)) \\
&= E\left(\sum_{i=0}^k \binom{k}{i} a^i X^i b^{k-i} | X > \frac{Q_\alpha(Y) - b}{a}\right) \\
&= \sum_{i=0}^k \binom{k}{i} a^i b^{k-i} E(X^i | X > Q_\alpha(X)) \\
&= \sum_{i=0}^k \binom{k}{i} a^i b^{k-i} CTE_\alpha(X^i), \quad k = 1, 2, \dots
\end{aligned}$$

If X is a standardized random variable with no parameter involved (e.g., standard normal), this equation is convenient in that one standard table of $CTE(X^i)$ can be used to calculate $CTE(Y^k)$ for any fixed a and b .

Example 2.12 (Lognormal distribution-revisited) To determine the conditional tail moments of $Y \sim LN(\mu, \sigma^2)$, consider $X = \log Y$. It follows that $X \sim N(\mu, \sigma^2)$ and

$$\begin{aligned}
CTE_\alpha(Y^k) &= E(Y^k | Y > Q_\alpha(Y)) = E(e^{kX} | e^X > Q_\alpha(Y)) \\
&= E(e^{kX} | X > \log Q_\alpha(Y)) = E(e^{kX} | X > Q_\alpha(X)).
\end{aligned}$$

This is directly related to the conditional tail m.g.f. of X defined in (2.8), that is, it can be calculated as follows:

$$\begin{aligned}
E(e^{uX} | X > d) &= \frac{1}{\bar{F}(d)} \int_d^\infty e^{ux} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\bar{F}(d)} \int_d^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu-\sigma^2 u)^2}{2\sigma^2}\right) dx \times \exp\left(\mu u + \frac{\sigma^2}{2} u^2\right) \\
&= \frac{\bar{F}(d - \sigma^2 u)}{\bar{F}(d)} \exp\left(\mu u + \frac{\sigma^2}{2} u^2\right).
\end{aligned}$$

Since $X \sim N(\mu, \sigma^2)$ the conditional tail moments of $Y = e^X$ are therefore given by

$$\begin{aligned} CTE_\alpha(Y^k) &= E(e^{kX} | X > Q_\alpha(X)) \\ &= \frac{\bar{F}(Q_\alpha(X) - \sigma^2 k)}{\bar{F}(Q_\alpha(X))} \exp\left(\mu k + \frac{\sigma^2}{2} k^2\right) \\ &= \frac{\bar{\Phi}(z_\alpha - \sigma k)}{\bar{\Phi}(z_\alpha)} \exp\left(\mu k + \frac{\sigma^2}{2} k^2\right), \quad k = 1, 2, \dots, \end{aligned}$$

where Φ is the c.d.f. of $N(0, 1)$, and $z_\alpha = (Q_\alpha(X) - \mu)/\sigma$. In particular the first two moments are

$$\begin{aligned} CTE_\alpha(Y) &= E(Y | Y > Q_\alpha(Y)) \\ &= \frac{\bar{\Phi}(z_\alpha - \sigma)}{\bar{\Phi}(z_\alpha)} \exp\left(\mu + \frac{\sigma^2}{2}\right) \end{aligned}$$

and

$$\begin{aligned} CTVar_\alpha(Y) &= E(Y^2 | Y > Q_\alpha(Y)) - (CTE_\alpha(Y))^2 \\ &= \frac{\bar{\Phi}(z_\alpha - 2\sigma)}{\bar{\Phi}(z_\alpha)} \exp(2\mu + 2\sigma^2) - \left[\frac{\bar{\Phi}(z_\alpha - \sigma)}{\bar{\Phi}(z_\alpha)} \exp\left(\mu + \frac{\sigma^2}{2}\right) \right]^2. \end{aligned}$$

Example 2.13 (Pareto distribution) *The distribution function of Pareto is given by*

$$F_Y(y) = 1 - \left(\frac{c}{y}\right)^a, \quad y \geq c,$$

and the density function, expressed in the canonical form, becomes

$$\begin{aligned} f(y) &= \frac{ac^a}{y^{a+1}} \\ &= \exp(a \log c + \log a - (a + 1) \log y) \end{aligned}$$

Thus the Pareto distribution belongs to the EF when c is known², but it suffers from the same shortcoming as the lognormal distribution; that is, the formula will produce $CTE(T) = CTE(\log Y)$, not $CTE(Y)$.

Now consider the distribution function of $Y = g(X) = e^X$. It turns out that

$$\begin{aligned} F_X(x) &= Pr(\log Y \leq x) = Pr(Y \leq e^x) \\ &= 1 - c^a e^{-xa}, \quad X \geq \log c, \end{aligned}$$

which is $Exp(1/a)$ shifted by $\log c$. Therefore for any $Q_\alpha(Y) \geq c$,

$$\begin{aligned} CTE_\alpha(Y^k) &= E(Y^k | Y > Q_\alpha(Y)) \\ &= E(e^{kX} | X > Q_\alpha(X)) \\ &= \int_{Q_\alpha(X)}^{\infty} e^{kx} ac^a e^{-xa} dx / \bar{F}_X(Q_\alpha(X)) \\ &= \frac{ac^a}{a-k} e^{-(a-k)Q_\alpha(X)} / c^a e^{-aQ_\alpha(X)} \\ &= \frac{a}{a-k} e^{kQ_\alpha(X)} \\ &= \frac{ac^k}{a-k} (1-\alpha)^{-k/a}, \quad k = 1, 2, \dots \end{aligned}$$

The $CTVar$ can be obtained by using $CTVar_\alpha(Y) = CTE_\alpha(Y^2) - CTE_\alpha(Y)^2$.

Example 2.14 (Generalized Pareto distribution) *The Generalized Pareto distribution (GPD) has been widely studied in connection with the extreme value theory in insurance and finance. The distribution of GPD is defined as*

$$F_Y(y) = 1 - \left[1 + \frac{\xi}{\beta} y \right]_+^{-1/\xi}, \quad \beta > 0.$$

²When a distribution's support depends on unknown parameters the distribution fails to be an EF.

So the domain is $y \geq 0$ if $\xi > 0$, or $0 \leq y \leq -\beta/\xi$ if $\xi < 0$; when $\xi = 0$ this reduces to the exponential distribution with mean β . Note that the GDP is not a member of the EF because its support depends on the parameters. Since $X = \log(1 + \frac{\xi}{\beta}Y)$ has the exponential distribution with mean ξ , the conditional tail moments of GPD, for $\xi > 0$, are given by

$$\begin{aligned}
CTE_\alpha(Y^k) &= E(Y^k | Y > Q_\alpha(Y)) \\
&= E\left[\left(\frac{\beta}{\xi}(e^X - 1)\right)^k \mid \frac{\beta}{\xi}(e^X - 1) > Q_\alpha(Y)\right] \\
&= \left(\frac{\beta}{\xi}\right)^k E[(e^X - 1)^k | X > Q_\alpha(X)] \\
&= \left(\frac{\beta}{\xi}\right)^k \frac{1}{\bar{F}_X(Q_\alpha(X))} \int_{Q_\alpha(X)}^\infty (e^x - 1)^k \frac{1}{\xi} e^{-x/\xi} dx \\
&= \left(\frac{\beta}{\xi}\right)^k \frac{1}{1 - \alpha} \int_{Q_\alpha(X)}^\infty (e^x - 1)^k \frac{1}{\xi} e^{-x/\xi} dx.
\end{aligned}$$

Using a binomial expansion $(e^x - 1)^k = \sum_{j=0}^k \binom{k}{j} e^{jx} (-1)^{k-j}$, the integration term of the above equation becomes

$$\begin{aligned}
\int_{Q_\alpha(X)}^\infty (e^x - 1)^k \frac{1}{\xi} e^{-x/\xi} dx &= \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \frac{1}{\xi} \int_{Q_\alpha(X)}^\infty e^{-(1/\xi - j)x} dx \\
&= \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \frac{1}{\xi} \frac{1}{1/\xi - j} \left[1 - e^{-(1/\xi - j)x}\right]_{Q_\alpha(X)}^\infty \\
&= \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \frac{1}{1 - j\xi} (1 - \alpha)^{1-j\xi}.
\end{aligned}$$

Note that we need $1/\xi - j > 0$ for all $j = 0, 1, \dots, k$ to get the finite value for the integration. This is equivalent to $\xi < 1/k$.

Finally we have

$$CTE_\alpha(Y^k) = \left(\frac{\beta}{\xi}\right)^k \frac{1}{1 - \alpha} \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \frac{(1 - \alpha)^{1-j\xi}}{1 - j\xi}, \quad \xi < 1/k.$$

For the first moment,

$$\begin{aligned} CTE_\alpha(Y) &= \left(\frac{\beta}{\xi}\right) \frac{1}{1-\alpha} \sum_{j=0}^1 \binom{1}{j} (-1)^{1-j} \frac{(1-\alpha)^{1-j\xi}}{1-j\xi} \\ &= \frac{\beta}{\xi} \left(\frac{(1-\alpha)^{-\xi}}{1-\xi} - 1 \right), \quad \xi < 1. \end{aligned}$$

For the variance, with a little algebra, we get

$$CTVar_\alpha(Y) = \frac{\beta^2(1-\alpha)^{-2\xi}}{(1-2\xi)(1-\xi)^2}, \quad \xi < 1/2.$$

Example 2.15 (Weibull distribution) *The distribution function of Weibull distribution is*

$$F_Y(y) = 1 - \exp[-\left(\frac{y}{\beta}\right)^\tau], \quad \beta, \tau > 0. \quad (2.16)$$

The Weibull distribution does not belong to the EF since in the exponential term it has $(\frac{y}{\beta})^\tau$ that cannot be broken into ηT form (parameter multiplied by a statistic) as in the densities of the EF. The distribution of $X = (Y/\beta)^\tau$, however, has the exponential distribution with mean 1, recognized directly from (2.16). Therefore

$$\begin{aligned} CTE_\alpha(Y^k) &= E(Y^k | Y > Q_\alpha(Y)) \\ &= E(\beta^k X^{k/\tau} | \beta X^{1/\tau} > Q_\alpha(Y)) \\ &= E(\beta^k X^{k/\tau} | X > \left(\frac{Q_\alpha(Y)}{\beta}\right)^\tau) \\ &= \beta^k E(X^{k/\tau} | X > Q_\alpha(X)) \\ &= \beta^k \frac{\int_{Q_\alpha(X)}^\infty x^{k/\tau} e^{-x} dx}{1-\alpha} \\ &= \beta^k \frac{\Gamma(1+k/\tau)(1-\Gamma(1+k/\tau; (-\log(1-\alpha)/\beta)^\tau))}{1-\alpha}, \quad k = 1, 2, \dots, \end{aligned}$$

where the gamma and the incomplete gamma functions are respectively defined by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

and

$$\Gamma(\alpha; z) = \frac{1}{\Gamma(\alpha)} \int_0^z t^{\alpha-1} e^{-t} dt.$$

2.5 Conclusion

In this chapter the simple formulas to calculate the first two moments of the conditional tail measure have been proposed for the exponential family. These formulas are based on the canonical form of the density and can be obtained by differentiating with respect to the corresponding canonical parameter. Also the conditional tail moment (and cumulant) generating function has been derived for the exponential family, which can be used to produce the moments of any order.

Another technique to get these moments for distributions transformed from the exponential family has been illustrated, which can help calculate conditional tail moments otherwise considered difficult to handle. Since the exponential family contains a wide range of distributions its conditional tail moment formulas derived in this chapter can be useful for risk management purposes.

Chapter 3

Bias correction of risk measures using the bootstrap

3.1 Introduction and motivation

Among other risk measures introduced in the first chapter the VaR in (1.2) and the CTE in (1.3) are the most used ones in practice for the required capital determination. In particular, the CTE has been widely accepted in the insurance field, having been prescribed both by the Canadian Institute of Actuaries for Segregated Fund Contracts (CIA, 2001) and by the American Academy of Actuaries in its ‘C3 Phase 2’ report (AAA, 2005).

There has been an considerable work on estimating the VaR and the CTE in both finance and actuarial science side with different methods and under different contexts. Landsman and Valdez (2003), Landsman and Valdez (2005), Cai and Li (2005), and the previous chapter of the thesis considered CTE estimation in certain parametric families. Commercial products such as RiskMetrics, KMV, and CreditMetrics

provide the VaR value using internal models. Gouriéroux *et al.* (2000) and Scaillet (2004) use a kernel method to estimate the tail measures, which can be classified as a semi-parametric method. Extreme value theory is an alternative technique to estimate the tail risk measures; see, e.g., Chapter 7 of McNeil *et al.* (2005). On the non-parametric side, Dowd (2001) used order statistics to estimate the VaR. If we use the mean squared error (MSE) as the standard tool to assess the estimator's uncertainty, both the bias and the variance matter. Recent papers by Inui *et al.* (2005), Bao and Ullah (2004), and Berkowitz and O'Brien (2002) address the topic of measuring the bias of the estimated VaR.

Using a parametric approach, estimating these measures, like other statistical modelings, involves the model error and parameter error, where these uncertainty would gradually decrease as sample size gets larger in general. There is however another important consideration arising in practice. Insurance companies typically have an in-house or commercial model implemented for cash flow validations and predictions based on various product variables (e.g., mortality, lapse, tax, and other expenses) and economic factors (e.g., interest rate, stock price, and inflation) fueled from past historical data. The model is often very complicated so that it requires a significant time¹ to simulate cash flows for each scenario, especially with a massive business volume, leaving the managers with a small sample size to estimate the VaR or the CTE of a whole business block. Since extreme events are rare, estimating these tail risk measures gets even harder. This sampling constraint is well known in industry and recognized by the effort to use variance reduction techniques (see

¹It is not uncommon that simulating cash flows for each scenario takes 30 minutes for a single block of business in practice.

Glasserman *et al.* (2000) for the VaR; Manistre and Hancock (2005) for the CTE) or to find so called representative scenarios in hope to reduce the number of simulations in actuarial applications; see, e.g., Christiansen (1998) and Chueh (2002) for details.

To this end we consider a situation where the tail risk measure is to be estimated from a small i.i.d.² sample of a distribution using a Monte Carlo simulation. In particular, in this chapter, we investigate the magnitude and direction (positive or negative) of the bias, and assess the advantages and disadvantages of using bootstrap techniques to correct the estimates for bias, with no reference to the parent distribution. The issue of the bias can be a generic statistical problem but it also has a particular implication in connection to the insurance regulation. The current Canadian regulation, as specified in the report on segregated fund investment guarantees by CIA Task Force (2002), prescribes the CTE 95% as the measure to determine the segregated fund investment guarantee's required capital using the Monte Carlo simulation. In detail the Total Gross Capital Requirement (TGCR) computes the aggregate CTE by adding the estimated CTE values for each exposure (policy). The report also recognizes that this approach would overstate the true value because the CTE is subadditive, that is, the sum of the CTEs is larger than the CTE of the sum. If there is bias in the estimated CTE for each contract, however, it will accumulate (because it is systematic) over a large portfolio to make the sum substantially biased. This effect then further extends to the whole industry as well, which could be a concern of the regulator. Coupled with the small sample size, the bias of tail risk measures thus is an interesting problem with practical implications for actuaries.

This chapter starts with defining the empirical VaR and CTE and noting that they

²independent and identically distributed.

are generally biased. We then introduce the bootstrap method as a nonparametric tool to estimate (and potentially correct) the bias, along with a new finding that, unlike the VaR, the empirical CTE is systematically biased. Numerical examples show that while the CTE bias may be corrected, doing so may result in an increase in the variance. At the end of the chapter a practical algorithm is proposed, which systematically selects the estimate with the smallest MSE among the empirical, the bootstrapped, and the bias-corrected CTE estimates.

3.2 Estimating VaR and CTE using finite sample

Consider a random variable X and its c.d.f.³ F , which is assumed to be continuous. Suppose that the Monte Carlo simulation generates an i.i.d. random sample $\mathbf{X} = (X_1, \dots, X_n)$ from F . The ordered sample is denoted by $(X_{(1)}, \dots, X_{(n)})$.

For the sample quantile, \hat{Q}_α , there are several suggestions for possible estimators. The simplest candidate is $X_{(r)}$ where $(r-1)/n < \alpha \leq r/n$. A slightly more sophisticated approach uses an interpolation between $X_{(r-1)}$ and $X_{(r)}$ if $(r-1)/n < \alpha < r/n$. Both of these estimators are biased, in finite samples, but are asymptotically unbiased.

Another adjustment is the use of $X_{(\alpha(n+1))}$, using some form of interpolation if $\alpha(n+1)$ is not an integer. For more discussion of these estimators, See for example, Hyndman and Fan (1996) and Klugman *et al.* (1998).

Given the same random sample $\mathbf{X} = (X_1, \dots, X_n)$ the sample CTE estimate for confidence level α is generally taken as:

$$\widehat{CTE}_\alpha = \frac{1}{n(1-\alpha)} \sum_{i=[n\alpha]+1}^n X_{(i)}, \quad (3.1)$$

³cumulative distribution function.

where the $X_{(i)}$ is the i -th ordered value of \mathbf{X} , and $[\cdot]$ is the floor function.

These estimators for the CTE and VaR take the form of a linear combination of order statistics which is commonly called L -estimators (assuming any interpolation required is also linear).

When the quantile estimator is biased, the corresponding CTE estimator is also biased because in general

$$CTE = E_X[X|X > Q_\alpha] \neq E_{\widehat{Q}_\alpha}[E_X(X|X > \widehat{Q}_\alpha)] = E[\widehat{CTE}]. \quad (3.2)$$

The bias will tend to zero as the sample size gets larger since \widehat{Q}_α is a consistent estimator of the true quantile, but will materially affect the accuracy of the CTE estimate for small sample sizes.

We illustrate the general problem with a simple analytic example.

Example 3.1 Consider a uniform random variable $X \sim U(0, 1)$. Suppose that we are interested in α such that $n\alpha$ is an integer. The true VaR here is α , while the two simple estimators are $X_{(n\alpha)}$ and $X_{(n\alpha+1)}$.

Now, for a sample size of n the expected value of r -th ordered value is given by $r/(n+1)$, $1 \leq r \leq n$.

If we set $\widehat{Q}_\alpha = X_{(n\alpha)}$, the bias is $-\alpha/(n+1)$. On the other hand if $\widehat{Q}_\alpha = X_{(n\alpha+1)}$ the bias is $(1-\alpha)/(n+1)$. In this example the latter choice yields much smaller bias in absolute terms in the right tail region, where α is close to 1.

This example shows that the direction and the magnitude of VaR bias depends on the choice of estimating function, the location of the quantile (for example, how big is α ? and, which tail?), as well as the underlying distribution shape and the sample size. For non-negative loss random variable, negative bias of VaR could result

in inadequacy of reserve or capital, and positive bias could cause inefficient use of capital.

Example 3.2 For the same uniform distribution $X \sim U(0,1)$ the true CTE value is $(1 + \alpha)/2$. Assuming again that $n\alpha$ is an integer the expected value of the sample CTE estimate is

$$\begin{aligned} E\left[\frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^n X_{(i)}\right] &= \frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^n E[X_{(i)}] \\ &= \frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^n \frac{i}{n+1} \\ &= \frac{n(1+\alpha)+1}{2(n+1)}. \end{aligned}$$

So, the bias is $-\frac{\alpha}{2(n+1)}$, meaning that the sample CTE tends to underestimate the true CTE in this example.

We will see in the following section that the negative bias of the empirical CTE, given in (3.1), is a general observation.

3.3 Bias of sample estimates of VaR and CTE

We first focus on estimates using a single sample value $X_{(r)}$. We define the following two estimates and name them respectively the lower side estimate and the upper side estimate:

$$\widehat{Q}_L(\alpha) = X_{(r)}, \quad \text{if } (r-1)/n < \alpha \leq r/n \quad (3.3)$$

$$\widehat{Q}_U(\alpha) = X_{(r)}, \quad \text{if } (r-1)/n \leq \alpha < r/n \quad (3.4)$$

These are identical except when $n\alpha$ is integer. For example if $n = 100$ and $\alpha = 0.95$, $\widehat{Q}_L(0.95) = X_{(95)}$ whereas $\widehat{Q}_U(0.95) = X_{(96)}$.

As we saw in the Uniform example, we may find a better estimator to lie between the low side and high side. There are many versions of estimators based on both the low side and high side of the sample values, as discussed in Hyndman and Fan (1996). Here we choose the one recommended by them which is:

$$\widehat{Q}_{HF}(\alpha) = (1 - \gamma)X_{(g)} + \gamma X_{(g+1)}, \quad (3.5)$$

where $g = [(n + 1/3)\alpha + 1/3]$ and $\gamma = (n + 1/3)\alpha + 1/3 - g$, derived from the approximation of the incomplete beta function ratio. A slightly modified version of this estimate is also found in Klugman *et al.* (1998), termed the smoothed quantile estimate.

There are estimators for the quantile that use more than two values of $X_{(j)}$. Harrell and Davis (1982) proposed:

$$\widehat{Q}_{HD}(\alpha) = \sum_{j=1}^n w_j X_{(j)}, \quad (3.6)$$

where

$$w_j = \frac{\int_{(j-1)/n}^{j/n} t^{(n+1)\alpha-1} (1-t)^{(n+1)(1-\alpha)-1} dt}{B[(n+1)\alpha, (n+1)(1-\alpha)]}$$

with $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$. This estimate is actually the exact bootstrap estimate of $E(X_{((n+1)\alpha}))$, even for non-integer $(n+1)\alpha$, as noted by Hutson and Ernst (2000). We discuss the exact bootstrap in more detail in the following section. Mausser (2001) shows that $\widehat{Q}_{HD}(\alpha)$ performs better than $\widehat{Q}_U(\alpha)$ for the marginal

VaR determination of some financial asset portfolios.

It is interesting to examine the bias of a single order statistic (that is, $X_{(r)}$ for integer r) against the true quantile when $n\alpha$ is integer. For example, is $E(X_{(95)})$ larger or smaller than the true quantile $Q(0.95)$? Sometimes this type of question can be tackled using the quantile bounds. There are different bounds available for the expected value of one order statistic expressed as a function of parent population's quantile; see Section 4.4 of David (1981) and references therein. Those bounds are nonparametric but additional information, such as convexity of the distribution function, often leads to better bounds. A few useful bounds for an i.i.d. sample can be derived by the c-ordering equivalence as follows, where $F(x) = \Pr[X \leq x]$ is the c.d.f. of X and n is the sample size.

1. $F[E(X_{(r)})] \leq (\geq) \frac{r}{n+1}$ if F is convex (concave)
2. $F[E(X_{(r)})] \leq (\geq) \frac{r-1}{n}$ if $1/F$ is concave (convex)
3. $F[E(X_{(r)})] \leq (\geq) \frac{r}{n}$ if $1/(1-F)$ is convex (concave)

It is possible for a distribution to satisfy more than one criterion above. These bounds can serve as an informal guideline if applied to the empirical distribution. For example if $1/F$ is concave, $E(X_{(n\alpha+1)}) \leq Q(\alpha)$ and $E(X_{(n\alpha)}) \leq Q(\alpha - 1/n)$, indicating that $\widehat{Q}_L(\alpha) = X_{(n\alpha)}$ is a bad choice since it makes the bias worse. For many common distributions, including the Normal, the Exponential, the Pareto, and the Gamma, for certain parameters, the $(n\alpha)$ th order statistic can be shown to be negatively biased, or $E(X_{(n\alpha)}) \leq Q(\alpha)$, by the convexity of $1/(1-F)$. While this underestimation of VaR is observed in many fat-tailed financial data there are examples with positive bias.

Consider a special case of the Inverse Weibull, also known as the Fréchet distribution, with c.d.f.

$$F(x) = \exp[-x^{-a}], \quad a > 0, \quad x > 0. \quad (3.7)$$

It can be shown that $1/(1-F)$ is concave in the right tail region for any $0 < a < 1$, thus $E(X_{(n\alpha)}) \geq Q(\alpha)$ when α is close to 1. This implies that the VaR based on historical data may actually exceed the true VaR for some very fat-tailed distributions. We finally note that the recent result on the bias of the VaR estimator by Inui *et al.* (2005) is equivalent to the first criterion above. Defining the VaR as a left side tail risk measure, they proved that $E[X_{((n+1)\alpha)}]$ converges to $Q(\alpha)$ from below, or $E[X_{((n+1)\alpha)}] \leq Q(\alpha)$ in the left tail area where F is convex. However, for nonnegative distributions whose c.d.f.s are concave, such as the exponential distribution, the result does not hold.

Turning to the CTE, we have a result that the sample CTE estimate is always negatively biased. The CTE is a form of trimmed mean – where we trim the lower $n\alpha$ values of the sample of n losses. Rychlik (1998) showed that for any identically (but not necessarily independently) distributed random sample (X_1, \dots, X_n) the trimmed mean has the following upper bound:

$$E \left[\frac{1}{k+1-j} \sum_{i=j}^k X_{(i)} \right] \leq \frac{n}{n+1-j} \int_{(j-1)/n}^1 Q(\alpha) d\alpha,$$

Where $Q(\alpha)$ is the α -quantile of the distribution of X . Plugging in $k = n$ and $j = n\alpha + 1$, assuming $n\alpha$ is an integer, gives the upper bound for the sample CTE

$$E \left[\frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^n X_{(i)} \right] \leq \frac{1}{1-\alpha} \int_{\alpha}^1 Q(x) dx = E[X|X > Q(\alpha)].$$

Interestingly, if we set $j = k = n\alpha + 1$ the upper bound for expected value of the sample VaR, $\widehat{Q}_U(\alpha)$, becomes CTE_{α} , for an identically distributed sample. There

is also a lower bound available, which is not considered here; see the reference for details.

We have shown in this section that common estimators for quantile and CTE risk measures are biased in general. One method for estimating and correcting for bias is through the bootstrap technique.

3.4 The bootstrap

3.4.1 Overview

The bootstrap methodology is particularly useful for non-parametric statistical inference. It has been widely applied by financial practitioners and actuaries as well as statisticians. For a comprehensive treatment, see standard textbooks such as Hall (1992), Efron and Tibshirani (1993), Shao and Tu (1995) or Davison and Hinkley (1997).

The core idea of the bootstrap is to create pseudo-samples by resampling (with replacement) from the original sample. The relationship of the pseudo-samples to the original sample replicates many features of the relationship of the original sample to its underlying distribution.

The basic procedure of the bootstrap can be sketched as follows. Suppose we have an i.i.d. random sample $\mathbf{X} = (X_1, \dots, X_n)$ from an unknown distribution, with c.d.f. F , and we are interested in a parameter $t(F)$ such as a quantile or CTE. Pseudo-samples are generated by sampling, with replacement, generating a new sample of the same size n as the original, from the empirical distribution function (e.d.f.) \hat{F} . The generated sample, denoted by \mathbf{X}^* , is called a bootstrap sample; the capital letter states that this too is a random sample, but from the e.d.f., indicated by superscript $*$.

The statistic of interest using this generated sample then is denoted by $T^* = T(\mathbf{X}^*)$. We repeat the exercise R times for R different bootstrap samples $\mathbf{X}_1^*, \dots, \mathbf{X}_R^*$, each of size n . From each sample we generate the statistic of interest, that is, T_k^* from the k -th bootstrap sample, giving $T_j^* = T(\mathbf{X}_j^*)$, $j = 1, \dots, R$. Finally the bootstrap estimate of the statistic T is given by

$$E(T|\hat{F}) = E^*(T^*) \approx \frac{1}{R} \sum_{j=1}^R T_j^*, \quad (3.8)$$

following the notation in Davison and Hinkley (1997). Here the condition \hat{F} in the first term indicates that the expectation is taken with respect to the e.d.f.; the true parameter value is expressed by $E(T|F)$. It is sometimes possible to compute $E(T|\hat{F})$ analytically without actually performing the simulation. Usually, however, the resampling simulation, referred to the ordinary bootstrap (OB), is inevitable. In these cases the estimate is subject to sampling error. The difference between the true bootstrap estimate and the estimated one, called the resampling (simulation) error, decreases as the resampling size R gets larger. We denote $R^{-1} \sum T_j^*$ by \bar{T}^* and call this the standard bootstrap estimator.

The unknown bias $E(T|F) - t(F)$ is approximated by its bootstrap estimate $B = E(T|\hat{F}) - t(\hat{F})$, thus the bootstrap bias estimate under R resamplings is

$$B_R = \bar{T}^* - t(\hat{F}) = \bar{T}^* - T. \quad (3.9)$$

Note that the bootstrap bias B_R converges to the true bias B as $R \rightarrow \infty$. If this is achieved the remaining uncertainty is only attributed to the original statistical error, i.e., to the fact that the empirical \hat{F} does not perfectly represent the true F . The statistical error can be reduced when the sample size n gets larger or one has more

information on F such as its parametric properties, neither of which might be feasible in practice.

We know that the estimator (e.g. for VaR or CTE) T is, in general, biased for finite sample sizes, so we could use a new bias-corrected estimator

$$t(\hat{F}) - B_R = 2T - \bar{T}^* \quad (3.10)$$

In practice, there is a trade-off between the improvement in the bias through using bias correction and the variability of the estimate in bootstrap. When the efficiency of the estimator is measured by the mean squared error (MSE) which is the sum of the squared bias and the variance of the estimator, this means that the bias correction may reduce the contribution of the first term, but could increase the second. Whether the resulting MSE will be smaller after bias correction depends on the underlying distribution as well as on the estimator itself; see for example Jeske and Sampath (2003). We will show some examples in the next section.

While the bootstrapping procedure is straightforward, one should be aware that there are applications where the standard bootstrap can fail. For example, if the data is incomplete or dependent then it is known that the bootstrap won't work in general. In estimating tail risk measures, it is possible for the true VaR or CTE to lie beyond the data maximum; e.g., CTE 99% estimate cannot be obtained from a sample of size 100. To prevent this problem we assume that the tail risk estimates are less than the data maximum, making the bootstrap a sensible tool for our purpose. Also bootstrapping the tail side data does not pose problems under parametric simulations, as long as the model is correct, because rare events in this case are due to the model's nature, and not from undetected errors.

3.4.2 Bootstrapping L-estimators

For the quantile and CTE risk measures we can utilize results available on bootstrap estimation of L-estimators. Throughout this subsection we assume the statistic of interest, T , is an L -estimator, that is a linear combination of order statistics:

$$T = \sum_{i=1}^n c_i X_{(i)}. \quad (3.11)$$

It is evident that the sample VaR and the CTE are special cases of this form with an appropriate selection of the coefficients c_i , $1 \leq i \leq n$. Hutson and Ernst (2000) derived the formula for the exact bootstrap (EB) mean and variance of the L -estimator. The bootstrap is exact in the sense that the resampling error is completely eliminated in the procedure; this is equivalent to bootstrapping at $R = \infty$. We show here how to apply the exact bootstrap to the quantile measure.

Theorem 3.3 (Hutson and Ernst (2000)) *The exact bootstrap (EB) of the estimate of $E(X_{(r)}|F)$, $1 \leq r \leq n$ is*

$$E(X_{(r)}|\hat{F}) = \sum_{j=1}^n w_{j(r)} X_{(j)},$$

where

$$w_{j(r)} = r \binom{n}{r} \left[B\left(\frac{j}{n}; r, n - r + 1\right) - B\left(\frac{j-1}{n}; r, n - r + 1\right) \right],$$

and

$$B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt.$$

Some $w_{j(r)}$ values are presented in Figure 3.1.

The weights are spread around the sample estimate, with heavier weights around $X_{(r)}$ and gradually smaller weights for distant observations. The Harrell-Davis quantile estimator, introduced in Section 3, is equivalent to the EB of $E(X_{((n+1)\alpha})|\hat{F})$.

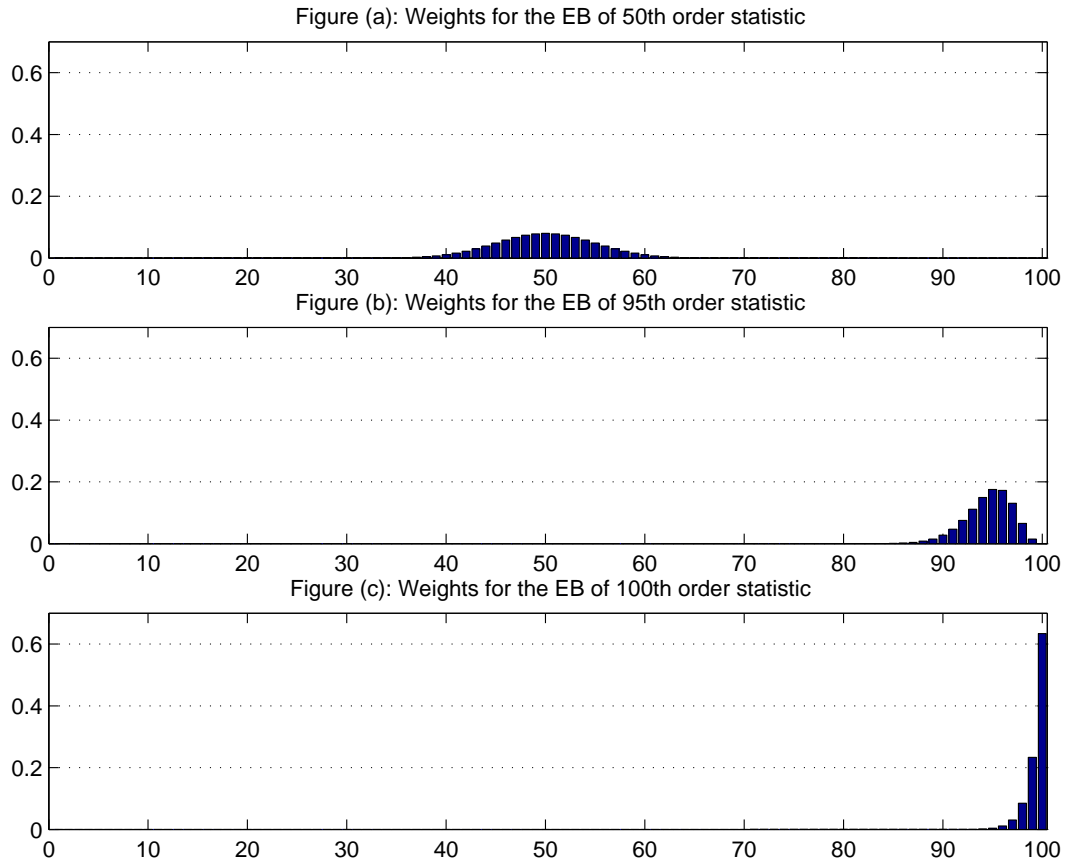


Figure 3.1: The weights of the EB for several order statistics when $n = 100$

Its weights can be compared to those of the EB at each order statistic. Figure 3.2 compares the 95% and 99% quantile weights. We can see from the figure that $\widehat{Q}_L^{EB} < \widehat{Q}_{HD} < \widehat{Q}_U^{EB}$ with \widehat{Q}_{HD} close to \widehat{Q}_U^{EB} .

There are several advantages of the EB over the OB with actual resamplings. The EB formula takes a simple analytic form and thus no simulations are involved. The simple form is easy to implement and significantly reduces the computing time compared to the OB. Since no resampling error is involved in the EB, we expect the the EB to be better than its OB counterpart in terms of variance on average. Finally the EB weights can be used for any samples with the same size because they are independent of the data. In the next section the EB is compare with the OB for both the VaR and the CTE using some parametric models.

Using Theorem 3.3 we now present several bootstrap-related quantities of the L -estimator in matrix form which is convenient in notation and useful for programming in matrix-based software such as Matlab. For a sample with size n , define $\mathbf{X}_{:n} = (X_{(1)}, \dots, X_{(n)})'$ and \mathbf{c} to be a column vector of size n . Then any L -estimator can be expressed as $T = \mathbf{c}'\mathbf{X}_{:n}$. For the CTE at confidence level α , for instance, we take

$$\mathbf{c} = (n(1 - \alpha))^{-1}(0, \dots, 0, 1, \dots, 1)'$$

with zeros for the first $n\alpha$ elements, to get the sample CTE estimate $\mathbf{c}'\mathbf{X}_{:n}$. Therefore the EB of T , an L -estimator, is

$$E(T|\widehat{F}) = E(\mathbf{c}'\mathbf{X}_{:n}|\widehat{F}) = \mathbf{c}'E(\mathbf{X}_{:n}|\widehat{F}) = \mathbf{c}'\mathbf{w}'\mathbf{X}_{:n}, \quad (3.12)$$

where the matrix $\mathbf{w} = \{w_{i(j)}\}_{i,j=1}^n$ comes from the EB weights for each element of $\mathbf{X}_{:n}$.

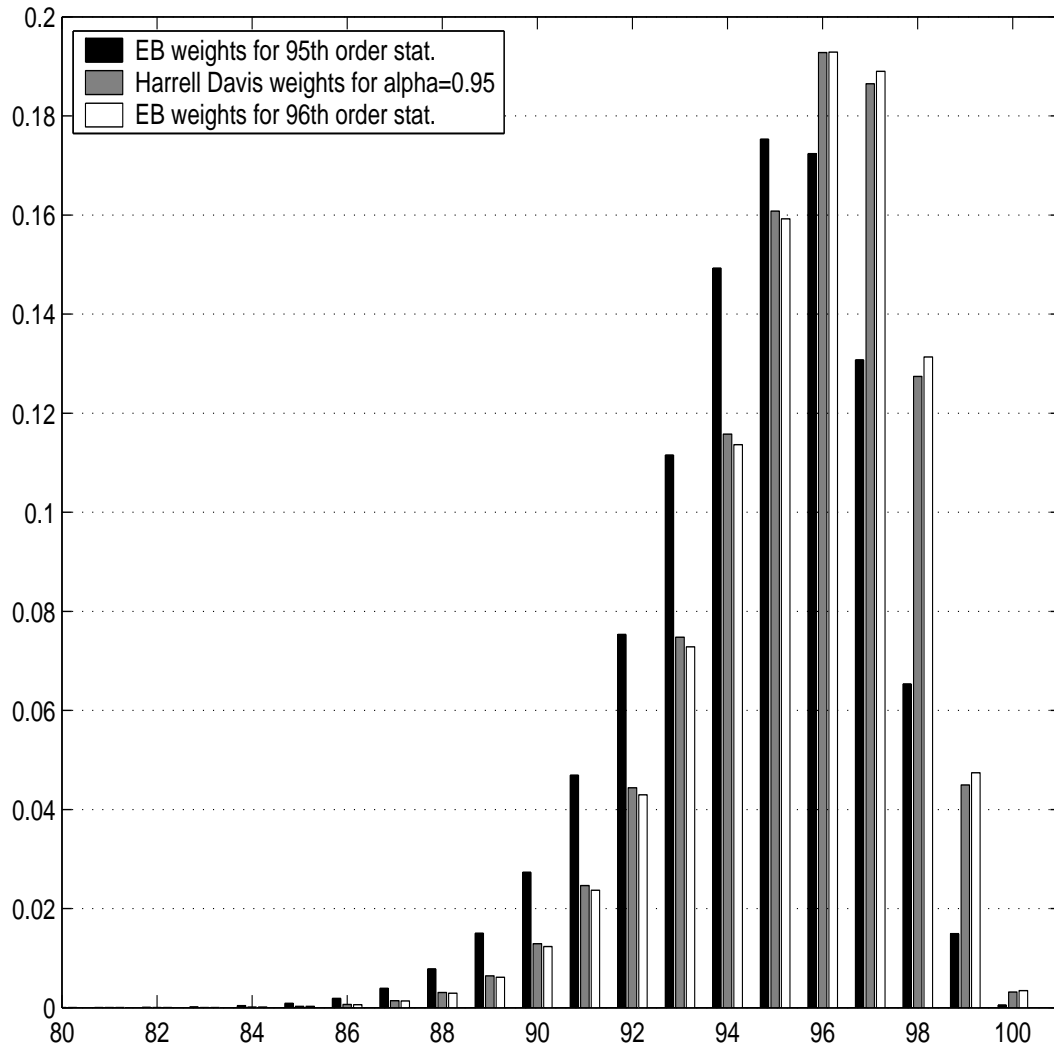


Figure 3.2: Comparisons of the weights: HD vs. EB estimators when $n = 100$.

Now the EB bias estimate is expressed by

$$B = E(T|\hat{F}) - t(\hat{F}) = \mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} - \mathbf{c}'\mathbf{X}_{:n} \quad (3.13)$$

Thus the bias-corrected estimator defined in (3.10) then is

$$t(\hat{F}) - B = \mathbf{c}'\mathbf{X}_{:n} - (\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} - \mathbf{c}'\mathbf{X}_{:n}) = \mathbf{c}'(\mathbf{2I} - \mathbf{w}')\mathbf{X}_{:n}. \quad (3.14)$$

Even though we have an analytic expression for the EB bias, we still need to see whether bias correction using the EB actually corrects the bias for the tail risk measures. The answer to this question is positive for the CTE as shown in the following theorem.

Theorem 3.4 *The empirical CTE estimator defined in (3.1) is always bigger than the EB of the CTE estimator, at any α such that $n\alpha$ is an integer. That is, mathematically,*

$$\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} < \mathbf{c}'\mathbf{X}_{:n},$$

for any given sample \mathbf{X} , where $\mathbf{c} = (n(1 - \alpha))^{-1}(0, \dots, 0, 1, \dots, 1)'$ with zeros for the first $n\alpha$ elements.

Proof: To start with, let us show that the weight matrix $\mathbf{w} = \{w_{i(j)}\}_{i,j=1}^n$ is doubly stochastic, meaning that the sum over each row and each column equal one. From Theorem 3.3, the (j, r) -th element of matrix \mathbf{w} is defined by

$$w_{j(r)} = r \binom{n}{r} \left[B\left(\frac{j}{n}; r, n - r + 1\right) - B\left(\frac{j-1}{n}; r, n - r + 1\right) \right]$$

We turn this into a different expression, which is

$$\begin{aligned} w_{j(r)} &= \frac{1}{B(r, n - r + 1)} \left[B\left(\frac{j}{n}; r, n - r + 1\right) - B\left(\frac{j-1}{n}; r, n - r + 1\right) \right] \\ &= \sum_{k=r}^n \binom{n}{k} \left(\frac{j}{n}\right)^k \left(1 - \frac{j}{n}\right)^{n-k} - \sum_{k=r}^n \binom{n}{k} \left(\frac{j-1}{n}\right)^k \left(1 - \frac{j-1}{n}\right)^{n-k} \end{aligned}$$

from a well known characteristic of the incomplete beta function ratio; see, e.g., Johnson *et al.* (1992). If we define two binomial random variables $Y_1 \sim \text{Bin}(n, \frac{j}{n})$ and $Y_2 \sim \text{Bin}(n, \frac{j-1}{n})$, the last expression becomes $Pr(Y_1 \geq r) - Pr(Y_2 \geq r)$. Thus the sum of j -th row elements of \mathbf{w} is

$$\begin{aligned} \sum_{r=1}^n w_{j(r)} &= \sum_{r=1}^n Pr(Y_1 \geq r) - \sum_{r=1}^n Pr(Y_2 \geq r) \\ &= E(Y_1) - E(Y_2) = n \frac{j}{n} - n \frac{j-1}{n} = 1. \end{aligned}$$

The sum over each column equals one by the definition of the weights, thus matrix \mathbf{w} has been proved to be doubly stochastic. Now put $\mathbf{c}'\mathbf{w}'\mathbf{X}_{\cdot\mathbf{n}} = \sum_{i=1}^n b_i X_{(i)}$ where $b_i = (n(1-\alpha))^{-1} \sum_{j=n\alpha+1}^n w_{i(j)}$. Note that for each i we have

$$0 < b_i < (n(1-\alpha))^{-1}$$

and

$$\sum_{i=1}^n b_i = 1,$$

because \mathbf{w} is doubly stochastic. Thus

$$\begin{aligned}
\mathbf{c}'\mathbf{X}_{:\mathbf{n}} - \mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} &= \sum_{i=n\alpha+1}^n (n(1-\alpha))^{-1}X_{(i)} - \sum_{i=1}^n b_i X_{(i)} \\
&= \sum_{i=n\alpha+1}^n (n(1-\alpha))^{-1}X_{(i)} - \sum_{i=1}^{n\alpha} b_i X_{(i)} - \sum_{i=n\alpha+1}^n b_i X_{(i)} \\
&= \sum_{i=n\alpha+1}^n [(n(1-\alpha))^{-1} - b_i]X_{(i)} - \sum_{i=1}^{n\alpha} b_i X_{(i)} \\
&> \sum_{i=n\alpha+1}^n [(n(1-\alpha))^{-1} - b_i]X_{(n\alpha)} - \sum_{i=1}^{n\alpha} b_i X_{(n\alpha)} \\
&= X_{(n\alpha)} \left[\sum_{i=n\alpha+1}^n [(n(1-\alpha))^{-1} - b_i] - \sum_{i=1}^{n\alpha} b_i \right] \\
&= X_{(n\alpha)} \left[\sum_{i=n\alpha+1}^n (n(1-\alpha))^{-1} - \sum_{i=1}^n b_i \right] = X_{(n\alpha)}[1 - 1] = 0.
\end{aligned}$$

We comment that the inequality in the middle of above expression holds for any random sample defined on the real line, not just positive ones. Q.E.D.

Note that this result also holds for the ordinary bootstrap with sufficiently large resampling size R because the EB is the limit value of the ordinary bootstrap.

Coupled with the result of Rychlik (1998) the above theorem gives

$$E(\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}}|F) < E(\mathbf{c}'\mathbf{X}_{:\mathbf{n}}|F) < CTE_\alpha(X),$$

for $\mathbf{c} = (n(1-\alpha))^{-1}(0, \dots, 0, 1, \dots, 1)'$ with zeros for the first $n\alpha$ elements. This implies that the bootstrap bias correction for CTE works in the right direction because the unknown bias $E(\mathbf{c}'\mathbf{X}_{:\mathbf{n}}|F) - CTE_\alpha(X)$ is estimated by $\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} - \mathbf{c}'\mathbf{X}_{:\mathbf{n}}$.

There is no similar result for the VaR case.

3.5 Simulations

3.5.1 Examples

Three different examples are used to compare the performance of various estimators for the VaR and the CTE using Monte Carlo simulations; the empirical, the OB with resamplings, the EB. For each bootstrap method the bias corrected estimator has also been computed. To assess the performance of the different estimators we repeat this with different generated samples and obtain the MSE values.

The first example is a 10-year European put option with the price return based on the Lognormal (LN) distribution. The initial price of the asset is set at \$100 and strike price is \$180, and the risk free rate is assumed 0.5% monthly effective. The LN parameters of the P-measure are $\mu = 0.00947$ and $\sigma = 0.04167$ which are derived from the monthly S&P 500 data during 1956-2001, as shown in Chapter 3 of Hardy (2003). Put options are often discussed in the cost of investment guarantees such as the segregated funds and variable annuities, where the strike price represents the guaranteed payment at maturity to customers. The put option here can be said to be at the money from insurer's perspective because the expected level of the fund in 10 years under the risk neutral measure is around \$182. We focus on the VaR and the CTE of the put option liability at different confidence levels. We assume no hedging here.

In the second example we consider the identical put option except that the underlying asset follows the Regime Switching Log-Normal distribution with two regimes (RSLN2) this time. See Hardy (2003) for details. The parameters are derived from the same S&P data: $\mu_1 = 0.0127, \mu_2 = -0.0162, \sigma_1 = 0.0351, \sigma_2 = 0.0691, p_{12} = 0.0468,$

and $p_{21} = 0.3232$. Since the left tail of the RSLN2 is fatter than the LN the risk measure associated with the guarantee cost is known to be significantly greater under the former model than those under the latter, and this is exactly what we observe in our simulations.

The final model is a fat-tailed Pareto distribution which has been popular in connection with extreme value theory in financial risk management, and is also used in property and casualty applications. The parameters are $\beta = 10$ and $\xi = 0.2$ where the c.d.f. of Pareto is

$$F(x) = 1 - \left(\frac{\beta}{\beta + \xi x} \right)^{1/\xi}, \quad x > 0, \quad (3.15)$$

following the notation of Manistre and Hancock (2005). The mean and the variance of this distribution are respectively 12.5 and 260.42 under our parameter choice. Note that the Pareto has fatter tail than the other two models.

In each model, we estimate the VaR using the \widehat{Q}_L , \widehat{Q}_U , \widehat{Q}_{HF} , and their various bootstrap versions using the OB with 100 resamplings as in (3.8), the EB as in (3.12), and the bias corrected ones given in (3.10) and (3.14) respectively. This results in fifteen different estimators. Also we included \widehat{Q}_{HD} in which case there is no bootstrapping. For the CTE estimation we compare the empirical estimate, the OB, the EB, with and without bias correction; in total five estimators.

3.5.2 Estimating the 99% Quantile risk measure

For the simulations we have simulated 20,000 samples of two sizes, 200 and 1000, from the given liability models. For each sample, we have calculated the lower quantile estimate, \widehat{Q}_L from equation (3.3), the upper estimate \widehat{Q}_U from equation (3.4), the HF

estimate, \widehat{Q}_{HF} from equation (3.5) and the HD estimate \widehat{Q}_{HD} from equation (3.6). In addition, for the first three measures, for each sample, we have calculated revised estimates using the ordinary bootstrap and the exact bootstrap, before and after bias correction.

The resulting values give the average bias and the average root mean square error, rMSE, associated with each measure. In Table 3.1, 3.2, and 3.3 show these values along with estimated standard errors⁴ of these averages expressed in percentage of the true values.

From these tables we note:

1. The tail measure from the sample is quite inaccurate even for sample size 1000, which would be on the high side in many actuarial applications.
2. The overall accuracy, as measured by the rMSE is improved by using the exact bootstrap, without bias correction. However we have no evidence that this would be true for quantiles in general.
3. Bias correction may increase the bias and substantially increase the rMSE – and in fact does so for the lognormal and RSLN models, for the \widehat{Q}_U and \widehat{Q}_{HF} . The usefulness of the bootstrap bias correction is limited because of the non-smoothness of the estimator, $\widehat{Q}_U = X_{(0.99n)}$, and further because we are near the bounds of the e.d.f. sample space. To illustrate more clearly, for the 99.5% quantile, \widehat{Q}_U would be the maximum value from the sample, which we expect

⁴Let $\hat{\theta}_i$ be the estimate from i -th sample, with $i = 1, 2, \dots, m$ (here we have $m = 20,000$). Then the variance of the average bias is estimated by $\hat{\sigma}^2/m$, where $\hat{\sigma}^2 = \sum(\hat{\theta}_i - \bar{\hat{\theta}})^2/(m-1)$ and the variance of the variance is estimated from the normal approximation $2(\hat{\sigma}^2)^2/(m-1)$. The same formulas are used for simulation result throughout Chapter 3 and 4.

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
\hat{Q}_L	-7.59%(0.12%)	16.7%(1.11%)	18.34%	-1.58%(0.06%)	7.91%(0.25%)	8.07%
\hat{Q}_L^{OB}	-9.34%(0.1%)	14.66%(0.85%)	17.38%	-1.94%(0.05%)	7.37%(0.22%)	7.62%
$\hat{Q}_L^{OB.bc}$	-5.84%(0.14%)	20.44%(1.66%)	21.26%	-1.21%(0.06%)	9.01%(0.32%)	9.09%
\hat{Q}_L^{EB}	-9.33%(0.1%)	14.55%(0.84%)	17.28%	-1.94%(0.05%)	7.32%(0.21%)	7.57%
$\hat{Q}_L^{EB.bc}$	-5.85%(0.14%)	20.36%(1.65%)	21.18%	-1.22%(0.06%)	8.96%(0.32%)	9.04%
\hat{Q}_U	4.69%(0.13%)	18.25%(1.32%)	18.84%	0.98%(0.06%)	8.05%(0.26%)	8.11%
\hat{Q}_U^{OB}	1.84%(0.11%)	15.75%(0.99%)	15.86%	0.58%(0.05%)	7.5%(0.22%)	7.52%
$\hat{Q}_U^{OB.bc}$	7.54%(0.16%)	22.81%(2.07%)	24.02%	1.38%(0.06%)	9.19%(0.34%)	9.29%
\hat{Q}_U^{EB}	1.84%(0.11%)	15.63%(0.97%)	15.74%	0.59%(0.05%)	7.45%(0.22%)	7.47%
$\hat{Q}_U^{EB.bc}$	7.55%(0.16%)	22.71%(2.05%)	23.93%	1.37%(0.06%)	9.14%(0.33%)	9.24%
\hat{Q}_{HF}	0.56%(0.12%)	16.93%(1.14%)	16.94%	0.12%(0.06%)	7.92%(0.25%)	7.92%
\hat{Q}_{HF}^{OB}	-1.92%(0.11%)	15.23%(0.92%)	15.35%	-0.27%(0.05%)	7.45%(0.22%)	7.45%
$\hat{Q}_{HF}^{OB.bc}$	3.04%(0.14%)	19.83%(1.56%)	20.06%	0.5%(0.06%)	8.84%(0.31%)	8.86%
\hat{Q}_{HF}^{EB}	-1.92%(0.11%)	15.13%(0.91%)	15.25%	-0.26%(0.05%)	7.4%(0.22%)	7.4%
$\hat{Q}_{HF}^{EB.bc}$	3.04%(0.14%)	19.74%(1.55%)	19.98%	0.5%(0.06%)	8.8%(0.31%)	8.81%
\hat{Q}_{HD}	1.72%(0.11%)	15.61%(0.97%)	15.71%	0.56%(0.05%)	7.45%(0.22%)	7.47%

True Value **39.7202**

Table 3.1: 99% Quantile estimators for the Lognormal example

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
\hat{Q}_L	-6.06%(0.09%)	13.25%(0.91%)	14.57%	-1.27%(0.04%)	6.29%(0.21%)	6.42%
\hat{Q}_L^{OB}	-7.58%(0.08%)	11.61%(0.7%)	13.87%	-1.57%(0.04%)	5.84%(0.18%)	6.05%
$\hat{Q}_L^{OB.bc}$	-4.55%(0.11%)	16.21%(1.36%)	16.83%	-0.97%(0.05%)	7.18%(0.27%)	7.24%
\hat{Q}_L^{EB}	-7.58%(0.08%)	11.53%(0.69%)	13.8%	-1.57%(0.04%)	5.81%(0.17%)	6.02%
$\hat{Q}_L^{EB.bc}$	-4.54%(0.11%)	16.14%(1.35%)	16.77%	-0.96%(0.05%)	7.15%(0.26%)	7.21%
\hat{Q}_U	3.5%(0.1%)	14.27%(1.06%)	14.69%	0.74%(0.05%)	6.41%(0.21%)	6.45%
\hat{Q}_U^{OB}	1.21%(0.09%)	12.33%(0.79%)	12.39%	0.42%(0.04%)	5.92%(0.18%)	5.94%
$\hat{Q}_U^{OB.bc}$	5.79%(0.13%)	17.76%(1.64%)	18.68%	1.05%(0.05%)	7.35%(0.28%)	7.42%
\hat{Q}_U^{EB}	1.22%(0.09%)	12.25%(0.78%)	12.31%	0.42%(0.04%)	5.89%(0.18%)	5.91%
$\hat{Q}_U^{EB.bc}$	5.78%(0.12%)	17.68%(1.62%)	18.6%	1.05%(0.05%)	7.31%(0.28%)	7.39%
\hat{Q}_{HF}	0.28%(0.09%)	13.31%(0.92%)	13.31%	0.06%(0.04%)	6.3%(0.21%)	6.3%
\hat{Q}_{HF}^{OB}	-1.75%(0.08%)	11.98%(0.74%)	12.1%	-0.25%(0.04%)	5.89%(0.18%)	5.89%
$\hat{Q}_{HF}^{OB.bc}$	2.31%(0.11%)	15.57%(1.26%)	15.74%	0.37%(0.05%)	7.07%(0.26%)	7.08%
\hat{Q}_{HF}^{EB}	-1.75%(0.08%)	11.91%(0.74%)	12.04%	-0.25%(0.04%)	5.86%(0.18%)	5.86%
$\hat{Q}_{HF}^{EB.bc}$	2.31%(0.11%)	15.49%(1.25%)	15.66%	0.37%(0.05%)	7.04%(0.26%)	7.05%
\hat{Q}_{HD}	1.12%(0.09%)	12.24%(0.78%)	12.29%	0.4%(0.04%)	5.89%(0.18%)	5.9%

True Value **51.8618**

Table 3.2: 99% Quantile estimators for the RSLN2 example

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
\hat{Q}_L	-5.86%(0.15%)	21.01%(3.34%)	21.81%	-1.36%(0.07%)	10.24%(0.79%)	10.33%
\hat{Q}_L^{OB}	-4.01%(0.14%)	20.03%(3.03%)	20.43%	-1.02%(0.07%)	9.7%(0.71%)	9.76%
$\hat{Q}_L^{OB.bc}$	-7.71%(0.18%)	26.03%(5.12%)	27.15%	-1.71%(0.08%)	11.58%(1.01%)	11.71%
\hat{Q}_L^{EB}	-3.99%(0.14%)	19.84%(2.97%)	20.23%	-1.01%(0.07%)	9.64%(0.7%)	9.7%
$\hat{Q}_L^{EB.bc}$	-7.73%(0.18%)	25.82%(5.04%)	26.95%	-1.71%(0.08%)	11.51%(1%)	11.64%
\hat{Q}_U	11.9%(0.22%)	30.65%(7.1%)	32.88%	1.99%(0.08%)	11.03%(0.92%)	11.2%
\hat{Q}_U^{OB}	13.38%(0.21%)	29.85%(6.74%)	32.71%	2.4%(0.07%)	10.42%(0.82%)	10.69%
$\hat{Q}_U^{OB.bc}$	10.43%(0.28%)	39.61%(11.86%)	40.96%	1.59%(0.09%)	12.53%(1.19%)	12.63%
\hat{Q}_U^{EB}	13.4%(0.21%)	29.63%(6.64%)	32.52%	2.41%(0.07%)	10.35%(0.81%)	10.63%
$\hat{Q}_U^{EB.bc}$	10.4%(0.28%)	39.4%(11.74%)	40.75%	1.58%(0.09%)	12.46%(1.17%)	12.56%
\hat{Q}_{HF}	5.92%(0.18%)	26.12%(5.16%)	26.78%	0.86%(0.08%)	10.64%(0.86%)	10.68%
\hat{Q}_{HF}^{OB}	7.52%(0.19%)	26.2%(5.19%)	27.26%	1.25%(0.07%)	10.17%(0.78%)	10.24%
$\hat{Q}_{HF}^{OB.bc}$	4.32%(0.22%)	31.48%(7.49%)	31.77%	0.48%(0.08%)	11.83%(1.06%)	11.84%
\hat{Q}_{HF}^{EB}	7.55%(0.18%)	26.02%(5.12%)	27.09%	1.25%(0.07%)	10.1%(0.77%)	10.18%
$\hat{Q}_{HF}^{EB.bc}$	4.3%(0.22%)	31.29%(7.4%)	31.58%	0.47%(0.08%)	11.76%(1.05%)	11.77%
\hat{Q}_{HD}	13.19%(0.21%)	29.49%(6.57%)	32.3%	2.37%(0.07%)	10.35%(0.81%)	10.61%

True Value **75.5943**

Table 3.3: 99% Quantile estimators for the Pareto example

to be positively biased for the models under consideration. But a bootstrap estimate of a sample maximum could never be positively biased.

4. Even in cases where the bias correction reduces the average bias, there is a concurrent increase in the standard error of the estimator for these examples.
5. The exact bootstrap (EB) is always more efficient than the ordinary bootstrap (OB) as we would expect, even though the improvement is marginal. The improvement is not statistically significant, but the systematic reduction in error is due to the elimination of bootstrap sampling volatility, thus it is strongly recommended to use the EB instead of the OB whenever possible.
6. The \widehat{Q}_{HF} and \widehat{Q}_{HF}^{EB} estimators perform well in all cases; if the estimator efficiency is important, this may be a good default selection.
7. Note that \widehat{Q}_{HD} performs well and often ranks the second best followed by \widehat{Q}_{HF} , but its performance is worsened when \widehat{Q}_{HF}^{EB} fails to be the best. As expected, we can see that $\widehat{Q}_L^{EB} < \widehat{Q}_{HD} < \widehat{Q}_U^{EB}$, and \widehat{Q}_{HD} is close to \widehat{Q}_U^{EB} throughout the simulations.

3.5.3 Estimating the 95% CTE

Following a similar process to the quantile results above, we simulated 20,000 samples of 200 values, and 20,000 samples each with 1,000 values. For each sample, we estimated the 95% CTE directly from the sample, (as the mean of the largest 5% of simulated loss values) and then again using the ordinary bootstrap (with 100 bootstrap replications) and the exact bootstrap, without and with bias correction.

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
\widehat{CTE}	-2.68%(0.12%)	16.89%(0.89%)	17.1%	-0.52%(0.05%)	7.42%(0.17%)	7.44%
\widehat{CTE}^{OB}	-5.38%(0.12%)	16.62%(0.86%)	17.47%	-1.06%(0.05%)	7.43%(0.17%)	7.5%
$\widehat{CTE}^{OB.bc}$	0.01%(0.12%)	17.36%(0.94%)	17.36%	0.02%(0.05%)	7.5%(0.18%)	7.5%
\widehat{CTE}^{EB}	-5.37%(0.12%)	16.55%(0.86%)	17.4%	-1.06%(0.05%)	7.39%(0.17%)	7.47%
$\widehat{CTE}^{EB.bc}$	0%(0.12%)	17.27%(0.93%)	17.27%	0.02%(0.05%)	7.46%(0.17%)	7.46%

True Value **31.2552**

Table 3.4: 95% CTE estimators for the lognormal example

The results are then averaged over the 20,000 simulations, and the final averages are shown in Tables 3.4, 3.5, and 3.6.

1. We notice that, as expected from Section 3.3 the sample estimates of the CTE are negatively biased, and the bias correction for the CTE does reduce the bias in all cases on average, unlike the quantile cases. However, in some cases the reduction in bias is outweighed by the resulting increase in variance, to give a slightly higher overall rMSE.
2. The use of the EB over the OB is again evident in these cases due to the elimination of the resampling error.
3. The case for using the bootstrap is not clear; the sample alone gives about the same accuracy as the bootstrapped variations, on average. In general, we find the EB with bias correction offers a very similar rMSE to the standard

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
\widehat{CTE}	-2.08%(0.09%)	12.68%(0.69%)	12.85%	-0.4%(0.04%)	5.62%(0.14%)	5.64%
\widehat{CTE}^{OB}	-4.17%(0.09%)	12.51%(0.67%)	13.18%	-0.83%(0.04%)	5.62%(0.14%)	5.68%
$\widehat{CTE}^{OB.bc}$	0.01%(0.09%)	13.01%(0.73%)	13.01%	0.02%(0.04%)	5.69%(0.14%)	5.69%
\widehat{CTE}^{EB}	-4.16%(0.09%)	12.44%(0.67%)	13.12%	-0.82%(0.04%)	5.6%(0.13%)	5.66%
$\widehat{CTE}^{EB.bc}$	0.01%(0.09%)	12.95%(0.72%)	12.95%	0.01%(0.04%)	5.65%(0.14%)	5.65%

True Value **42.9634**

Table 3.5: 95% CTE estimators for the RSLN-2 example

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
\widehat{CTE}	-1.32%(0.13%)	17.99%(2.06%)	18.03%	-0.33%(0.06%)	8.1%(0.42%)	8.11%
\widehat{CTE}^{OB}	-2.71%(0.13%)	17.75%(2.01%)	17.95%	-0.6%(0.06%)	8.11%(0.42%)	8.13%
$\widehat{CTE}^{OB.bc}$	0.08%(0.13%)	18.41%(2.16%)	18.41%	-0.06%(0.06%)	8.17%(0.43%)	8.17%
\widehat{CTE}^{EB}	-2.69%(0.12%)	17.67%(1.99%)	17.88%	-0.6%(0.06%)	8.07%(0.42%)	8.09%
$\widehat{CTE}^{EB.bc}$	0.06%(0.13%)	18.31%(2.14%)	18.31%	-0.06%(0.06%)	8.13%(0.42%)	8.13%

True Value **63.7853**

Table 3.6: 95% CTE estimators for the Pareto example

estimator, but with a smaller bias, which may be preferable. Also note that the bias-corrected one never achieves the lowest MSE, but there is no theoretical foundation to believe that this is true for all cases.

4. The situation in a practical context is that the risk manager has a single sample, and the only (non-parametric) way to estimate the bias to assess how much of a problem it might be is to bootstrap.

As a guideline as to whether to apply the bias correction, Efron and Tibshirani (1993) suggest that the ratio of the bootstrap bias estimate to the bootstrap standard error should be considered. If the ratio is bigger than 0.25 the bias correction is worth using.

If the bias is not large relative to the standard error, it may be worth using the EB estimate even without bias correction rather than the empirical estimator. We see from the tables that in some cases the EB estimate is preferred, at other times the empirical estimator has smaller rMSE. An interesting question is how an actuary with a single sample from an unknown underlying distribution should decide whether applying the exact bootstrap will improve the estimator of the CTE or not. To help with this decision we have developed a guideline which is described in the following section.

3.6 A practical guideline

In the previous section the true risk measure values was available in comparing different estimators' performances, but one would have only a single sample with no information on the true value. We propose a practical guideline for the CTE estimation that can be used to select the CTE estimator with the smallest MSE among three

candidates: the empirical one (\widehat{CTE}), the EB one (\widehat{CTE}^{EB}), and the bias-corrected one ($\widehat{CTE}^{EB.bc}$). The MSE for estimator $\hat{\theta}$ of a parameter θ is defined by

$$MSE(\hat{\theta}) = (E(\hat{\theta}) - \theta)^2 + Var(\hat{\theta}) \quad (3.16)$$

The proposed guideline is based on an approximation to the true MSE using one sample, which is given by

$$\widehat{MSE}(\hat{\theta}) = (\hat{\theta} - \tilde{\theta})^2 + \widehat{Var}(\hat{\theta}), \quad (3.17)$$

where $\tilde{\theta}$ is the benchmark that is considered to be close to θ and the variance estimate is computed by the ordinary resamplings. Based on findings in the previous two sections that the bias correction is theoretically valid for the CTE and numerical examples show that the bias-corrected estimate is fairly close to the true CTE on average, we set $\tilde{\theta} = \widehat{CTE}^{EB.bc} = \mathbf{c}'(\mathbf{2I} - \mathbf{w}')\mathbf{X}_{:n}$. For the empirical CTE estimator the approximated MSE therefore is given by

$$\widehat{MSE}(\mathbf{c}'\mathbf{X}_{:n}) = (\mathbf{c}'\mathbf{X}_{:n} - \mathbf{c}'(\mathbf{2I} - \mathbf{w}')\mathbf{X}_{:n})^2 + \mathbf{c}'\hat{\mathbf{V}}_{\mathbf{R}}(\mathbf{X}_{:n})\mathbf{c}, \quad (3.18)$$

where $\hat{\mathbf{V}}_{\mathbf{R}}(X_{:n})$ is the bootstrap estimate of the covariance matrix, $V(X_{:n})$ with $Cov(X_{(i)}, X_{(j)})$ for each element, using R resamplings. Similarly, for the EB estimator of CTE and the bias-corrected CTE estimator the approximation gives

$$\widehat{MSE}(\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n}) = (\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} - \mathbf{c}'(\mathbf{2I} - \mathbf{w}')\mathbf{X}_{:n})^2 + \mathbf{c}'\mathbf{w}'\hat{\mathbf{V}}_{\mathbf{R}}(\mathbf{X}_{:n})\mathbf{w}\mathbf{c}, \quad (3.19)$$

and

$$\widehat{MSE}(\mathbf{c}'(\mathbf{2I} - \mathbf{w}')\mathbf{X}_{:n}) = \mathbf{c}'(\mathbf{2I} - \mathbf{w}')\hat{\mathbf{V}}_{\mathbf{R}}(\mathbf{X}_{:n})(\mathbf{2I} - \mathbf{w}')\mathbf{c}, \quad (3.20)$$

respectively. Note that the bias is zero for the bias-corrected CTE estimator. So the guideline is simply to select the one that has the smallest approximate MSE among

three different CTE estimators for a given sample. Actually one could slightly modify the MSE approximation formula given in (3.17) by replacing the bias term $(\hat{\theta} - \tilde{\theta})^2$ by $(E(\hat{\theta}|\hat{F}) - \tilde{\theta})^2$ with the expectation taken with respect to the empirical distribution using the bootstrap. We did not take this form however because $\tilde{\theta}$ actually is an estimator from sample, rather than a true parameter, so both $\hat{\theta}$ and $\tilde{\theta}$ will have biases. Taking expectation on only $\hat{\theta}$ using the bootstrap therefore would distort the true bias. Since our goal is to provide an approximate rule of thumb, we keep the proposed guideline to avoid further theoretical difficulties.

To illustrate this, we generated 10,000 samples, each with 200 values of the the same three models to estimate 99% CTEs. For each sample, we computed (3.18), (3.19), and (3.20) and chose the estimate with the smallest value. Resampling size of 1000 was used to estimate the covariance matrix for each sample. The selected estimate is labeled ‘Mixed’ in Table 3.7. The tables illustrate that using the guideline gives an average rMSE that is near the lowest of the three, consistently being ranked in the second place throughout examples, without referring to the true CTE value. Because we generally assume the underlying model is unknown in practice, and only one sample is available, the proposed guideline can be useful to actuaries who estimate the CTE from a small sample generated from complicated industry-scaled models. In the Pareto case, the exact bootstrap offers a lower MSE, and in the LN and the RSLN case the empirical estimates are better.

Looking at the bias column we see that the mixed estimator tends to pick up the bias-corrected estimator quite often. This would be partly due to the fact that bias term of the bias-corrected one is set zero for the MSE approximation. This is somewhat different from the simulation results in the previous section where bias-corrected

estimator never achieved the lowest MSE value across the three examples. However, as explained earlier, we cannot exclude the bias-corrected one from the list of candidates based on several examples. Moreover choosing the bias-corrected one may have a good advantage in practice when similar CTE estimations are repeated on a large volume of business to produce their sum because the CTE bias will accumulate if not corrected.

3.7 Concluding remarks

We investigated the bias of two risk measures, the quantile and the CTE, in finite samples. For the quantile, different estimators are compared with bootstrapped and bias corrected bootstrapped ones. Simulations show that the exact bootstrap has definite advantage over the ordinary resampled bootstrap. The bootstrap bias correction however should not be applied to tail quantiles. The exact bootstrap offers a reasonably efficient estimator in many cases.

For the CTE we found the sample estimate is negatively biased from both theoretical prediction and numerical examples, and the bias correction works reasonably well for small sample size, though increase in variance may decrease the efficiency of the estimator with bias correction. In both VaR and CTE, the EB proves to be better than the OB in terms of the computational and theoretical efficiency, as expected.

Later we propose a practical guideline to help determine the optimal estimator by approximating the MSE among three: the empirical, the EB, and the bias-corrected one. In light of the current Canadian regulation, we can say that adding individually

CTE 99% in the Lognormal model				
CTE = 47.7281; sample size $n=200$				
Method	Estimate	Bias (s.e.)	Std (s.e.)	rMSE
Empirical	45.31	-2.41(0.07)	7.24(0.74)	7.63
EB	42.91	-4.82(0.07)	6.60(0.62)	8.17
EB.bc	47.72	0.00(0.08)	8.02(0.91)	8.02
Mixed	46.38	-1.35(0.08)	7.86(0.87)	7.98

CTE 99% in the RSLN put model				
CTE = 59.9989; sample size $n=200$				
Method	Estimate	Bias (s.e.)	Std (s.e.)	rMSE
Empirical	57.66	-2.34(0.07)	7.39(0.77)	7.75
EB	55.15	-4.85(0.07)	6.79(0.65)	8.34
EB.bc	60.16	0.16(0.08)	8.14(0.94)	8.14
Mixed	58.80	-1.20(0.08)	8.02(0.91)	8.11

CTE 99% in the Pareto model				
CTE = 106.993; sample size $n=200$				
Method	Estimate	Bias (s.e.)	Std (s.e.)	rMSE
Empirical	101.31	-5.68(0.34)	34.48(16.81)	34.94
EB	94.52	-12.47(0.30)	30.01(12.74)	32.50
EB.bc	108.10	1.10(0.39)	39.13(21.65)	39.14
Mixed	101.47	-5.52(0.34)	33.56(15.93)	34.01

Table 3.7: Application of MSE test to the three examples, 99% CTE.

estimated CTEs to produce the aggregate CTE will not be as conservative as one may think because the bias will systematically accumulate. However the exact offset between the bias effect and the diversification effect is largely unknown, because the latter depends on the dependency structure of different provisions.

Finally we comment that the examples in this chapter show that the variance is dominating in the MSE of tail risk measure and thus the bias correction exercise is just part of the whole story; in particular, the variance could easily compromise the risk reduction through the bias correction.

In the next chapter we focus on the variance estimation. In particular the variance of the bootstrapped L-estimator class is provided using the nonparametric delta method. It will be seen that the advantages of the EB of the L-estimator goes beyond the computational and theoretical efficiency; it provides a basis for influence function of the bootstrapped L-estimators.

Chapter 4

Variance estimation of bootstrapped risk measures

4.1 Introduction

Estimating the variance of a statistical estimator provides a logical way to assess its variability across various samples from the same distribution and, therefore, the uncertainty associated with individually computed estimates. For example if we have statistics $\hat{\theta}$ and $\hat{\sigma}_{\hat{\theta}}$, a standard error of $\hat{\theta}$, and a parameter θ such that $(\hat{\theta} - \theta)/\hat{\sigma}_{\hat{\theta}}$ converges to the standard normal in distribution as the sample size $n \rightarrow \infty$, then the approximate confidence interval for θ at α is given by

$$[\hat{\theta} - z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}]$$

according to the Central Limit Theorem together with Slutsky's Theorem.

In the previous chapter we considered estimating the bias of two tail risk mea-

asures, VaR and CTE, based on a small sized sample using the bootstrap method. This chapter is a continuation of the previous chapter, where we now assess the uncertainty of tail risk measures estimates by focusing on the variance. For an estimator relying on the tail region of a distribution, the variance is typically higher than an estimator from the center of a distribution. The estimated variance of a risk measure provides the information on the accuracy of the estimate, and also, in our context, on whether the number of scenarios is adequate to produce a reliable capital requirement.

We focus on the CTE as the statistic of interest, although the main result applies to any L-estimator, which is a linear combination of order statistics.

Under the parametric approach, for a given i.i.d. random sample, the variance of the ordinary CTE estimate $T = \mathbf{c}'\mathbf{X}_{:\mathbf{n}}$ is

$$\begin{aligned} Var(\mathbf{c}'\mathbf{X}_{:\mathbf{n}}) &= Var\left[\frac{1}{n(1-\alpha)} \sum X_i I(X_i > \widehat{Q}_\alpha)\right] \\ &= Var_{\widehat{Q}_\alpha}[E_X(T|\widehat{Q}_\alpha)] + E_{\widehat{Q}_\alpha}[Var_X(T|\widehat{Q}_\alpha)]. \end{aligned} \quad (4.1)$$

where $\widehat{Q}_\alpha = X_{(n\alpha)}$, the $(n\alpha)$ th order statistic, provided that $n\alpha$ is an integer. The outer expectation and the variance are taken with respect to the order statistic. The second term of (4.1) represents the additional uncertainty due to the variability of the order statistic; in principle this becomes larger where quantiles are hard to estimate. This expression also motivates the conditional tail moments of higher orders, as discussed in Chapter 2 for the exponential distribution family. While it is possible to express the variance in explicit forms for some distributions, in most cases numerical methods are used for the actual evaluation due to analytic difficulty.

In the nonparametric context estimating the variance of the CTE estimate is also

possible, which is the topic of this chapter. As discussed in Chapter 3, a risk manager may prefer the bootstrapped CTE over the empirical CTE for a given sample, after applying the proposed practical guideline. Therefore we are not only concerned with the variance of the empirical CTE, $Var(\mathbf{c}'\mathbf{X}_{:n})$, but also the variance of the bootstrapped CTE, $Var(\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n})$.

There are two popular nonparametric methods in variance estimation: the bootstrap and the nonparametric delta method using the influence function. For example, Manistre and Hancock (2005) use the nonparametric delta method to estimate the variances of the VaR and CTE. While the bootstrap is straightforward in estimating the variance of any estimate, there has been no discussion on its relative efficiency compared to the nonparametric delta method counterpart in the actuarial context. In this chapter both methods are discussed and compared to each other in a simulation study. Most notably the influence function of the bootstrapped quantile is derived, which in turn can be used to estimate any bootstrapped L-estimator including the bootstrapped CTE, or $Var(\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n})$. Note that for small samples the variance of any estimator will be different from that of the bootstrapped counterpart, even though both converge to the same quantity as the sample size increases.

For the simulation study we consider small sample sizes for this problem, say $n \leq 1000$, using different nonparametric methods. The sample size of less than 1000 is common in actuarial loss modelling and in operational or credit risk modelling.

4.2 The bootstrap

The simplest estimate of $Var(T)$, where $T = \mathbf{c}'\mathbf{X}_{:n}$, is the empirical one:

$$\frac{\sum_{i=n\alpha+1}^n (X_{(i)} - T)^2}{(n - n\alpha)(n - n\alpha - 1)} \quad (4.2)$$

This is an unbiased estimate of the variance, given the quantile \widehat{Q}_α , recognizing that T is the mean of an ordered sample $(X_{(n\alpha+1)}, X_{(n\alpha+2)}, \dots, X_{(n)})$ with size $n - n\alpha$. In the Monte Carlo simulation this quantity represents the last term $E[Var(T|\widehat{Q}_\alpha)]$ in (4.1). Hence the quantile variability component, $Var[E(T|\widehat{Q}_\alpha)]$, is systematically missing, which in turn makes estimate (4.2) always less than the true value because variance is nonnegative.

As in the case of estimating the CTE itself in the previous chapter, the bootstrap can be used to estimate the variance of the CTE. Using the resampling with replacement scheme, with R resamplings, the bootstrap estimate of the variance is given by

$$V_R = \sum_{j=1}^R [T_j^* - \overline{T}^*]^2 / (R - 1). \quad (4.3)$$

where \overline{T}_j^* is the CTE computed from j -th resample and \overline{T}^* is the bootstrap estimate of the CTE, or the average of \overline{T}_j^* , $1 \leq j \leq R$. Along with the bootstrap mean Hutson and Ernst (2000) provides the exact bootstrap (EB) result for the bootstrap variance of the L-estimator as follows.

Theorem 4.1 (Hutson and Ernst (2000)) *The exact bootstrap of estimate of $\sigma_{r:n}^2$*

and $\sigma_{rs:n}^2$, ($1 \leq r \leq n$ and $r < s$) are

$$\begin{aligned}\hat{\sigma}_{r:n}^2 &= \text{Var}(X_{r:n}|\hat{F}) = \sum_{j=1}^n w_{j(r)}(X_{j:n} - \hat{\mu}_{r:n})^2, \\ \hat{\sigma}_{rs:n} &= \text{Cov}(X_{r:n}, X_{s:n}|\hat{F}) = \sum_{j=2}^n \sum_{j=1}^{j-1} v_{ij(rs)}(X_{i:n} - \hat{\mu}_{r:n})(X_{j:n} - \hat{\mu}_{s:n}) \\ &\quad + \sum_{j=1}^n v_{jj(rs)}(X_{j:n} - \hat{\mu}_{r:n})(X_{j:n} - \hat{\mu}_{s:n})\end{aligned}$$

where the weights are

$$\begin{aligned}v_{ij(rs)} &= \int_{(j-1)/n}^{j/n} \int_{(i-1)/n}^{i/n} f_{rs:n}(u_r, u_s) du_r du_s, \\ v_{jj(rs)} &= \int_{(j-1)/n}^{j/n} \int_{(j-1)/n}^{u_s} f_{rs:n}(u_r, u_s) du_r du_s,\end{aligned}$$

and $f_{rs:n}(u_r, u_s) = {}_n C_{rs} u_r^{r-1} (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s}$ is the joint distribution of two uniform order statistics $U_{r:n}$ and $U_{s:n}$ with ${}_n C_{rs} = n!/(r-1)!(s-r-1)!(n-s)!$.

Hutson and Ernst (2000) also provide alternative expressions for the weights using the binomial series expansion of $(u_s - u_r)^{s-r-1}$. Then the weights become

$$\begin{aligned}v_{ij(rs)} &= {}_n C_{rs} \sum_{k=0}^{s-r-1} \binom{s-r-1}{k} (-1)^{s-r-1-k} \int_{(i-1)/n}^{i/n} u_r^{s-k-2} du_r \int_{(j-1)/n}^{j/n} u_s (1 - u_s)^{n-s} du_s \\ &= {}_n C_{rs} \sum_{k=0}^{s-r-1} \binom{s-r-1}{k} \frac{(-1)^{s-r-1-k}}{s-k-1} \left[\left(\frac{i}{n} \right)^{s-k-1} - \left(\frac{i-1}{n} \right)^{s-k-1} \right] \\ &\quad \times \left[B\left(\frac{j}{n}; k+1, n-s+1 \right) - B\left(\frac{j-1}{n}; k+1, n-s+1 \right) \right]\end{aligned}\tag{4.4}$$

and

$$v_{jj(rs)} = {}_n C_{rs} \sum_{k=0}^{s-r-1} \binom{s-r-1}{k} \frac{(-1)^{s-r-1-k}}{s-k-1} \left\{ B\left(\frac{j}{n}; s, n-s+1\right) - B\left(\frac{j-1}{n}; s, n-s+1\right) \right. \\ \left. - \left(\frac{j-1}{n}\right)^{s-k-1} \left[B\left(\frac{j}{n}; k+1, n-s+1\right) - B\left(\frac{j-1}{n}; k+1, n-s+1\right) \right] \right\}, \quad (4.5)$$

with

$$B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt.$$

We provide an alternative expression of the above formula in matrix form. Let $D(\mathbf{w}_r)$ be the diagonal matrix with (i, i) -th element equal to i -th element of $\mathbf{w}_r = (w_{1(r)}, w_{2(r)}, \dots, w_{n(r)})'$, the EB mean weight vector of the r -th order statistic defined in Theorem 3.3. Also denote the upper-triangular matrix $\{v_{ij(rs)}\}_{i \leq j}$ by $\mathbf{v}_{(rs)}$. Then

$$\begin{aligned} \text{Var}(X_{r:n} | \hat{F}) &= (\mathbf{X}_{:n} - \hat{\mu}_{(r)})' D(\mathbf{w}_r) (\mathbf{X}_{:n} - \hat{\mu}_{(r)}) \\ &= (\mathbf{X}_{:n} - \mathbf{w}'_r \mathbf{X}_{:n})' D(\mathbf{w}_r) (\mathbf{X}_{:n} - \mathbf{w}'_r \mathbf{X}_{:n}), \end{aligned}$$

for $r = 1, 2, \dots, n$, and

$$\begin{aligned} \text{Cov}(X_{r:n}, X_{s:n} | \hat{F}) &= (\mathbf{X}_{:n} - \hat{\mu}_{(r)})' \mathbf{v}_{(rs)} (\mathbf{X}_{:n} - \hat{\mu}_{(s)})' \\ &= (\mathbf{X}_{:n} - \mathbf{w}'_r \mathbf{X}_{:n})' \mathbf{v}_{(rs)} (\mathbf{X}_{:n} - \mathbf{w}'_s \mathbf{X}_{:n}), \end{aligned}$$

for $r < s$. Note that each (co)variance term of order statistic(s) is a quadratic form; $D(\mathbf{w}_r)$ and $\mathbf{v}_{(rs)}$ are independent of samples, so they are recyclable for other samples of the same size.

Now let us further define the EB covariance matrix of order statistics by:

$$\hat{\mathbf{V}}(\mathbf{X}_{:\mathbf{n}}) = \{Cov(X_{(r)}, X_{(s)}|\hat{F})\}_{i,j=1}^n. \quad (4.6)$$

Then the EB variance estimate of the empirical CTE estimate, the bootstrapped CTE in (3.12), and the bias-corrected CTE in (3.14) are, respectively,

$$Var(\mathbf{c}'\mathbf{X}_{:\mathbf{n}}|\hat{F}) = \mathbf{c}'\hat{\mathbf{V}}(\mathbf{X}_{:\mathbf{n}})\mathbf{c}, \quad (4.7)$$

$$Var(\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}}|\hat{F}) = \mathbf{c}'\mathbf{w}'\hat{\mathbf{V}}(\mathbf{X}_{:\mathbf{n}})\mathbf{w}\mathbf{c}, \quad (4.8)$$

and

$$Var(\mathbf{c}'(\mathbf{2I} - \mathbf{w}')\mathbf{X}_{:\mathbf{n}}|\hat{F}) = \mathbf{c}'(\mathbf{2I} - \mathbf{w}')\hat{\mathbf{V}}(\mathbf{X}_{:\mathbf{n}})(\mathbf{2I} - \mathbf{w})\mathbf{c}. \quad (4.9)$$

This implies that we can estimate the variance of any L-estimate, including the bootstrapped ones with no additional effort, as long as we have the EB covariance estimate matrix $\hat{\mathbf{V}}(\mathbf{X}_{:\mathbf{n}})$. As seen so far the matrix expression presents the formulas in a more appealing way; it is also highly beneficial in computation if the computer software, such as Matlab, handles matrices efficiently.

A closer look at this EB variance however prompts a computational issue which Hutson and Ernst (2000) did not discuss. First notice that $\hat{\mathbf{V}}(\mathbf{X}_{:\mathbf{n}})$ in (4.6) is a $n \times n$ matrix but computing each element of this matrix involves another $n \times n$ matrix. The total number of computations is of order $O(n^4)$ ¹; the computational burden increases exponentially as the sample size gets larger. Secondly if each $\mathbf{v}_{(rs)}$ is to be stored for future recycling, a huge storage capacity is required². Thirdly, even if these have been

¹For instance, when $n = 400$ the number of computations is about 1.28×10^{10} and the computing time takes more than a week in Matlab using a high-performance Unix machine.

²Hard disk space needs almost 6GB when $n = 200$; more than 200GB when $n = 400$ in Matlab.

stored for sample size n , one should recalculate all these matrices as the sample size changes. For these reasons we recommend approximating $\hat{\mathbf{V}}(\mathbf{X}_{\cdot n})$ by $\hat{\mathbf{V}}_{\mathbf{R}}(\mathbf{X}_{\cdot n})$ which is the estimate using the ordinary bootstrap approach with R bootstrap samples. In the following section we derive a recursive formula that reduces the computing time substantially.

4.2.1 Improvement using recursion

We have just seen that the EB variance computation gets impractical as one increases the sample size. One, of course, can rely on the ordinary bootstrap with resampling despite its less satisfactory accuracy. In this section two recursive formulas are presented which substantially reduce the EB variance computation.

Let us denote the k -th component inside the summation of (4.4) and (4.5) to be $\psi_{ij(r_s)}(k)$, $k = 0, 1, \dots, s - r - 1$, for given r and s ; i.e.,

$$\psi_{ij(r_s)}(k) = \binom{s-r-1}{k} (-1)^{s-r-1-k} \int_{\frac{i-1}{n}}^{\frac{i}{n}} u_r^{s-k-2} du_r \int_{\frac{j-1}{n}}^{\frac{j}{n}} u_s^k (1-u_s)^{n-s} du_s, \quad (4.10)$$

when $i < j$, and

$$\psi_{jj(r_s)}(k) = \binom{s-r-1}{k} (-1)^{s-r-1-k} \int_{\frac{i-1}{n}}^{\frac{i}{n}} u_r^{s-k-2} du_r \int_{\frac{j-1}{n}}^{u_s} u_s^k (1-u_s)^{n-s} du_s, \quad (4.11)$$

when $i = j$. Then we have the following lemma.

Lemma 4.2 *The EB variance weight matrices defined in (4.4) and (4.5) satisfy the following recursive formula:*

$$v_{ij(r+1,s)} = -\frac{s-r-1}{r} v_{ij(r_s)} + \frac{n C_{rs}}{s-r-1} \sum_{k=0}^{s-r-1} k \psi_{ij(r_s)}(k)$$

Proof: Applying a binomial coefficient relation

$$\binom{s-r-2}{k} = \binom{s-r-1}{k} - \frac{k}{s-r-1} \binom{s-r-1}{k}, \quad k = 0, 1, \dots, s-r-2,$$

to (4.10) and (4.11) leads to the following recursive form for $\psi_{ij(r,s)}$:

$$\psi_{ij(r+1,s)}(k) = -\psi_{ij(rs)}(k) + \frac{k}{s-r-1} \psi_{ij(rs)}(k), \quad k = 0, 1, \dots, s-r-2,$$

and $i \leq j$. Finally plugging this result back in (4.4) and (4.5), and rearranging yields the lemma. Q.E.D.

Simulation studies show that this recursive formulation reduces computing time by approximately 80% over computing with no recursion for sample size 400. Another recursion presented below involves $f_{rs:n}$ defined in Theorem 4.1.

Lemma 4.3 *Define ${}_nC_{rs} = n!/(r-1)!(s-r-1)!(n-s)!$. Then for*

$$f_{rs:n}(u_r, u_s) = {}_nC_{rs} u_r^{r-1} (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s},$$

the following recursive equation holds.

$$(s-r)f_{r-1,s:n} + (n-s+1)f_{r-1,s-1:n} + (r-1)f_{r,s:n} = n(n-s-1)f_{r-1,s-1:n-1} \quad (4.12)$$

Proof:

$$\begin{aligned}
\frac{f_{r-1,s;n}(u_r, u_s)}{{}_n C_{r-1,s}} &= u_r^{r-2} (u_s - u_r)^{s-r} (1 - u_s)^{n-s} \\
&= u_r^{r-2} (u_s - u_r)^{s-r-1} (u_s - u_r)^1 (1 - u_s)^{n-s} \\
&= u_r^{r-2} u_s (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s} - u_r^{r-1} (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s} \\
&= u_r^{r-2} u_s (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s} - \frac{f_{r,s;n}(u_r, u_s)}{{}_n C_{rs}} \\
&= -u_r^{r-2} (1 - u_s - 1) (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s} - \frac{f_{r,s;n}(u_r, u_s)}{{}_n C_{rs}} \\
&= -u_r^{r-2} (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s+1} \\
&\quad + u_r^{r-2} (u_s - u_r)^{s-r-1} (1 - u_s)^{n-s} - \frac{f_{r,s;n}(u_r, u_s)}{{}_n C_{rs}} \\
&= -\frac{f_{r-1,s-1;n}(u_r, u_s)}{{}_n C_{r-1,s-1}} + \frac{f_{r-1,s-1;n-1}(u_r, u_s)}{{}_{n-1} C_{r-1,s-1}} - \frac{f_{r,s;n}(u_r, u_s)}{{}_n C_{rs}},
\end{aligned}$$

which, setting $f_{r,s;n}(u_r, u_s) = f_{r,s;n}$,

$$\frac{f_{r-1,s;n}}{{}_n C_{r-1,s}} + \frac{f_{r-1,s-1;n}}{{}_n C_{r-1,s-1}} + \frac{f_{r,s;n}}{{}_n C_{rs}} = \frac{f_{r-1,s-1;n-1}}{{}_{n-1} C_{r-1,s-1}}, \quad (4.13)$$

which becomes (4.12). Q.E.D.

One can compute $v_{ij(rs)}$, $i < j$, from this recursion, by taking a series of integrations to both sides of (4.12) over each interval generated by $P_{n-1} \cup P_n$ with respect to both i and j , where

$$P_{n-1} = \left\{0, \frac{1}{n-1}, \frac{2}{n-1}, \dots, \frac{n-2}{n-1}, 1\right\}$$

and

$$P_n = \left\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$$

This scheme handles different intervals embedded in $v_{ij(rs)}$ for different n automatically. The formula allows both backward and forward recursion. For instance if all

$v_{ij(rs)}$, $r < s$, values are known for sample size n , we can compute $v_{ij(rs)}$ for sample size $n - 1, n - 2, \dots, 2, 1$, recursively. The forward recursion states that when all $v_{ij(rs)}$, $r < s$, values are known for sample size n , $v_{ij(rs)}$, for $n + 1$ can be computed by recursion provided that the first row of matrix $v_{ij(rs)}$ is specified for $n + 1$. We do not seek further developments in this direction because even though this recursion is computationally more efficient than non-recursive formula, its storage cost remains unchanged, which is impractical considering the price of current computer storage; we will leave this result for future generations to implement.

4.3 Nonparametric delta method

An alternative way to estimate the variance of an estimator is using the nonparametric delta method which employs the influence function of the estimator. Estimating variance through the nonparametric delta method is well known and can be found in standard texts such as Staudte and Sheather (1990) or Hampel *et al.* (1986).

Consider the von Mises expansion of any statistical functional $t(G)$ at F . The first order approximation is

$$t(G) \approx t(F) + \int L_t(x|F)dG(x). \quad (4.14)$$

Here the first derivative of t at F , L_t , is called the influence function (IF). The IF is a function of x given F and t , and defined by

$$L_t(x|F) = \lim_{\epsilon \rightarrow 0} \frac{t[(1 - \epsilon)F + \epsilon H_x] - t(F)}{\epsilon}. \quad (4.15)$$

with H_x is c.d.f. of a degenerated random variable at x , or commonly referred to the heaviside function. The IF measures the relative influence on $T(F)$ of a very small amount of contamination at x and also can be used to estimate the variance of T .

For our purpose of estimating the variance from the sample, we set $G = \hat{F}$, which is a choice that makes the approximation reasonably accurate as n increases, and the approximation becomes

$$t(\hat{F}) \approx t(F) + \int L_t(x|F)d\hat{F}(x) = t(F) + \frac{1}{n} \sum_{i=1}^n L_t(x_i|F) \quad (4.16)$$

Now by applying the central limit theorem, $T = t(\hat{F})$ has asymptotic normality:

$$t(\hat{F}) - t(F) \xrightarrow{d} N(0, v_L(F)), \quad (4.17)$$

where $v_L(F) = n^{-1} \text{Var}(L_t(X)) = n^{-1} \int L_t^2(x|F)dF(x)$.

Assuming no information on F in practice, we estimate this variance using the sample version:

$$v_L(\hat{F}) \equiv \frac{1}{n^2} \sum_{j=1}^n L_t^2(x_j|\hat{F}), \quad (4.18)$$

where x_j is the j -th observation of the sample. $L_t(x_j|\hat{F})$ is often called the empirical influence value.

Let us consider some well-known examples of IFs for our discussion. When $t(F) = \int x dF(x)$, the mean, it is easy to see that $t[(1 - \epsilon)F + \epsilon H_x] = \epsilon(x - t(F))$ so that the IF of the mean is $L_t(x|F) = x - t(F)$ from (4.15). The following is the IF of the α -th quantile:

Lemma 4.4 (e.g., Staudte and Sheather (1990)) *The influence function of $t(F) = F^{-1}(\alpha) = Q_\alpha$ is given by*

$$L_t(x|F) = \begin{cases} \frac{\alpha - 1}{f(Q_\alpha)}, & \text{if } x < Q_\alpha \\ 0, & \text{if } x = Q_\alpha \\ \frac{\alpha}{f(Q_\alpha)}, & \text{if } x > Q_\alpha \end{cases} \quad (4.19)$$

One comment on the IF of the quantile is that estimating the variance of any quantile gets difficult because its IF involves the density of the unknown distribution, making (4.18) hard to compute with samples. Although there are nonparametric tools available allowing us to estimate the density, such as kernel density estimation, it may not produce satisfactory values for tail regions, especially when the given sample is subject to skewness or (and) kurtosis, which is often true for financial and actuarial data. In practice therefore, using the nonparametric delta method to estimate the variance of VaR, a tail quantile risk measure, can be problematic.

The IF of the CTE is also available by the following lemma. The proof is provided here because the result was stated in Manistre and Hancock (2005) without proof. The derivation is essentially adapted from the IF of the trimmed mean as shown in, e.g., Section 3.2.2 of Staudte and Sheather (1990).

Lemma 4.5 *The influence function of $t(F) = (1 - \alpha)^{-1} \int_{Q_\alpha}^\infty x dF$ is given by*

$$L_t(x|F) = \begin{cases} \frac{x - \alpha Q_\alpha}{1 - \alpha} - t(F) & \text{if } Q_\alpha < x \\ Q_\alpha - t(F) & \text{if } Q_\alpha \geq x \end{cases} \quad (4.20)$$

Proof:

$$T(F) = \int_{F^{-1}(\alpha)}^\infty \frac{y}{1 - \alpha} dF(y) = \int_{F^{-1}(\alpha)}^\infty \frac{y}{1 - F(F^{-1}(\alpha))} dF(y) \quad (4.21)$$

Therefore by defining $F_\epsilon(y) = (1 - \epsilon)F(y) + \epsilon H(y - x)$, with $H(y - x) = H_x(y)$ being the heaviside function,

$$\begin{aligned}
T(F_\epsilon) &= \int_{F_\epsilon^{-1}(\alpha)}^{\infty} \frac{y}{1 - F_\epsilon(F_\epsilon^{-1}(\alpha))} dF_\epsilon(y) \\
&= \int_{F_\epsilon^{-1}(\alpha)}^{\infty} \frac{y}{1 - \alpha} dF_\epsilon(y) \\
&= \int_{F_\epsilon^{-1}(\alpha)}^{\infty} \frac{y}{1 - \alpha} d((1 - \epsilon)F(y) + \epsilon H(y - x)) \\
&= \frac{1 - \epsilon}{1 - \alpha} \int_{F_\epsilon^{-1}(\alpha)}^{\infty} y dF(y) + \frac{\epsilon}{1 - \alpha} \int_{F_\epsilon^{-1}(\alpha)}^{\infty} y dH(y - x) \\
&= \frac{1 - \epsilon}{1 - \alpha} \int_{F_\epsilon^{-1}(\alpha)}^{\infty} y dF(y) + \epsilon \frac{x}{1 - \alpha} I(F_\epsilon^{-1}(\alpha) < x)
\end{aligned}$$

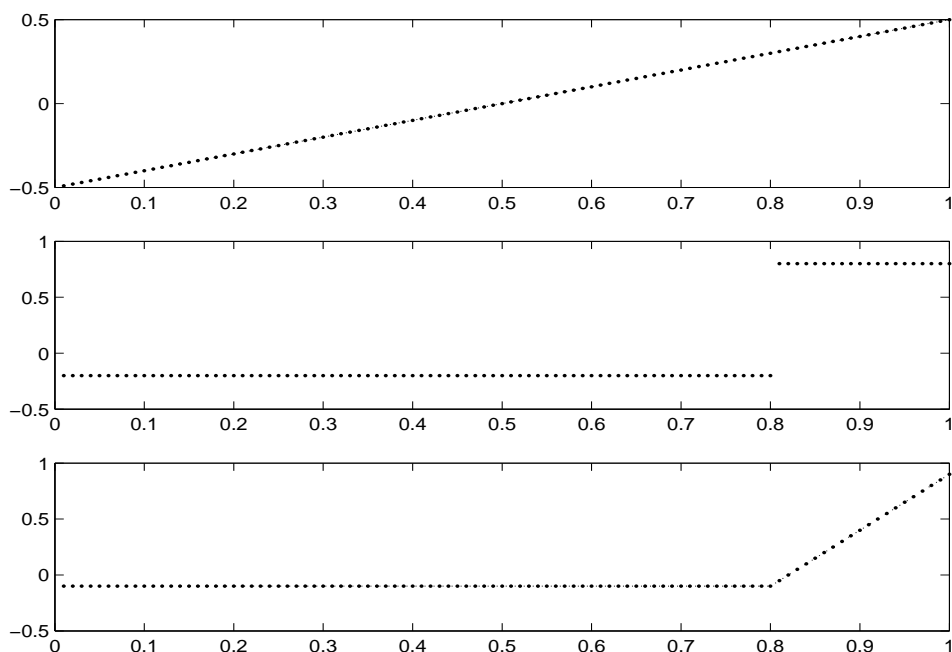
Differentiating with respect to ϵ gives

$$\begin{aligned}
\frac{\partial}{\partial \epsilon} T(F_\epsilon(y)) &= \frac{-1}{1 - \alpha} \int_{F_\epsilon^{-1}(\alpha)}^{\infty} y dF(y) - \frac{1 - \epsilon}{1 - \alpha} F_\epsilon^{-1}(\alpha) f(F_\epsilon^{-1}(\alpha)) \cdot \frac{\partial}{\partial \epsilon} F_\epsilon^{-1}(\alpha) \\
&\quad + \frac{x}{1 - \alpha} I(F_\epsilon^{-1}(\alpha) < x)
\end{aligned}$$

Finally taking $\epsilon = 0$ and plugging the quantile influence function into above equation completes the proof. Q.E.D.

Unlike the quantile case no density is involved in the IF of the CTE, so that we can easily estimate the variance either using (4.18) or the sample version of the following form:

$$\begin{aligned}
v_L(F) &= \frac{1}{n} \int L_t^2(x) dF(x) \\
&= \int_0^{Q_\alpha} (x - E[X|X > Q_\alpha])^2 dF + \int_{Q_\alpha}^{\infty} \left(\frac{x - \alpha Q_\alpha}{1 - \alpha} - E[X|X > Q_\alpha] \right)^2 dF(x) \\
&= \frac{\text{Var}(X|X > Q_\alpha) + \alpha(E[X|X > Q_\alpha] - Q_\alpha)^2}{n(1 - \alpha)},
\end{aligned}$$

Figure 4.1: IF of mean, quantile, CTE for $U(0, 1)$, $\alpha = 80\%$

as noted in Manistre and Hancock (2005). Numerically both methods will produce identical values.

For illustration, Figure 4.1 sketches the IFs of the mean, the quantile, and the CTE from top to bottom, respectively, for $U(0, 1)$ with $\alpha = 0.8$. Note that the graph of the CTE IF is a negative constant up to Q_α , followed by a linear function with positive slope. It is continuous but unbounded from above. This indicates that the CTE estimator is negatively influenced by a contamination at x when $x < Q_\alpha$; but its influence is constant, meaning that the CTE value becomes smaller if there is a small contamination at x but is not affected by the location of x as long as it is below the threshold. This is consistent with the definition of the CTE. On the other hand it is positively influenced by outliers exceeding the threshold; one infinite outlier will push

the value of CTE to infinity, unlike the VaR case where the IF function is bounded both from above and below, but is discontinuous at the quantile in question. In theory, if the IF is bounded and continuous the estimator will be qualitatively robust; however no estimators in the graph fall in this category.

Before closing this section we validate the similarity between the IF approach and the bootstrap approach in estimating the variance of an arbitrary statistic $T = t(\hat{F})$, using a linear approximation of the bootstrap estimator. This means that as sample size increases one would expect similar numbers from both methods. While this connection is already known (see, for example, Section 2.7.4 of Davison and Hinkley (1997)) we explain this because we will use both methods in the numerical example later in this chapter.

The bootstrap analogy to (4.16) can be expressed by:

$$t(\hat{F}^*) \approx t(\hat{F}) + \frac{1}{n} \sum_{i=1}^n L_t(x_i^* | \hat{F}) = t(\hat{F}) + \frac{1}{n} \sum_{i=1}^n f_i^* L_t(x_i | \hat{F}) \quad (4.22)$$

where \hat{F}^* is the e.d.f. of an bootstrap sample and x_i^* is its i -th element. f_i is the number of times that x_i^* equals x_i , for $1 \leq i \leq n$. Note that only f_i^* will vary on each resampling. By repeating resampling with replacement one can see that

$$Var(T | \hat{F}) \approx \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n Cov(f_i^*, f_j^* | \hat{F}) L_t(x_i | \hat{F}) L_t(x_j | \hat{F}), \quad (4.23)$$

where the left side is the bootstrap variance of T . Since f_i^* is multinomial for a given sample with probability of $1/n$ for each i , $Cov(f_i^*, f_j^* | \hat{F}) = 1 - 1/n$ if $i = j$ and $-1/n$

otherwise. Therefore a simpler expression is possible as follows.

$$\begin{aligned}
\text{Var}(T|\hat{F}) &\approx \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(f_i^*, f_j^*|\hat{F}) L_t(x_i|\hat{F}) L_t(x_j|\hat{F}) \\
&= \frac{1}{n^2} \left[\sum_{i=1}^n L_t^2(x_i|\hat{F}) - \frac{1}{n} \left(\sum_{i=1}^n L_t(x_i|\hat{F}) \right)^2 \right] \\
&\approx \frac{1}{n^2} \sum_{i=1}^n L_t^2(x_i|\hat{F}) \\
&= v_L(\hat{F})
\end{aligned}$$

Thus we observe that the bootstrap variance estimate $\text{Var}(T|\hat{F})$ and the empirical influence driven variance estimate $v_L(\hat{F})$ in (4.18) would produce close numbers as n increases.

Finally we note that while the IF approach illustrated in this section works well for the empirical CTE estimator, it cannot be applied directly to the bootstrapped or bias-corrected CTE estimators because these are not empirical estimates of some continuous functionals. In the next section we will elaborate further on this.

4.4 IF of bootstrapped L-estimator

Even though the variances of the empirical VaR and CTE are obtained through the IFs, given in (4.19) and (4.21) respectively, we found in the previous chapter that the EB estimators are often preferred to the empirical ones in terms of the mean squared error. So a natural question would be how to compute the variance of the ‘bootstrapped’ tail risk measure. In this chapter we answer this question and propose a method that can compute the variances of the EB L-estimators and the bias-corrected

L-estimators. In particular we first derive the IF of the exact bootstrap r -th order statistic using the fact that the weight vector $\mathbf{w}_r = (w_{1(r)}, \dots, w_{n(r)})$ is continuous and bounded. This result then directly extends to the whole L-estimator class, including the CTE.

We already defined an L-estimator class by

$$c_1 X_{(1)} + c_2 X_{(2)} + \dots + c_n X_{(n)}, \quad (4.24)$$

where weights c_1, \dots, c_n are implicitly assumed to be mass weights. For example, the empirical VaR and CTE estimators belong to this class. Noting that the bootstrapped and the bias-corrected L-estimators again belong to the same L-estimator class, we examine the IF of an L-estimator at a more general level, where an L-estimator is defined by

$$t(G) = \int_{-\infty}^{\infty} x J(G(x)) dG(x) = \int_0^1 J(u) G^{-1}(u) du, \quad (4.25)$$

where G is the distribution function and $J(u)$ is called the score function defined on $[0, 1]$. Depending on whether the score function is discrete or not the resulting IF becomes quite different.³ When the score function is discrete (4.25) reduces to a linear combination of order statistics in (4.24). The CTE is a good example to explain the difference, and to illustrate what J looks like. First note that we differentiate the CTE and its empirical estimator, though both belong to the L-estimator class defined in (4.25). The CTE at α , by definition, is

$$t(F) = E[X|X > Q_\alpha] = (1 - \alpha)^{-1} \int_{Q_\alpha}^{\infty} x dF(x) = (1 - \alpha)^{-1} \int_\alpha^1 F^{-1}(u) du$$

³If J is bounded and continuous a.e. Lebesgue and a.e. F^{-1} , then $t(F)$ is called Fréchet differentiable. Quantiles, however, are not Fréchet differentiable. See, e.g., Appendix B of Staudte and Sheather (1990) or Chapter 1 of Rieder (1994) for details.

and the corresponding score function is a step function given by

$$J(u) = \begin{cases} 0 & \text{if } u \leq \alpha \\ (1 - \alpha)^{-1} & \text{if } u > \alpha \end{cases}$$

The ‘empirical CTE’ estimator however is an average of high side order statistics and can be expressed by

$$t(\hat{F}) = \int_0^1 J(u) \hat{F}^{-1}(u) du = (n - n\alpha)^{-1} \sum_{n\alpha+1}^n X_{(i)},$$

for integer $n\alpha$, and the score function is a discrete mass function given by

$$J(u) = \begin{cases} 0 & \text{for } u = 1/n, 2/n, \dots, \alpha \\ (1 - \alpha)^{-1} & \text{for } u = \alpha + 1/n, \alpha + 2/n, \dots, 1 \end{cases}$$

The following lemma shows that if the score function consists of discrete masses, its IF necessarily involves the density form.

Lemma 4.6 (e.g., pp.21 of Rieder (1994)) *Consider an L-estimator with discrete weights, $\sum_1^n c_i X_{(i)}$. Write $\alpha_i = i/n$, $i = 1, \dots, n$. Then the IF is given by*

$$\sum_{i=1}^n \frac{c_i}{f(F^{-1}(\alpha_i))} [\alpha_i - H(F^{-1}(\alpha_i) - x)],$$

where $H(F^{-1}(\alpha_i) - x)$ is the unit step function jumping 0 to 1 at $x = F^{-1}(\alpha_i)$. The asymptotic variance of the L-estimator is

$$\begin{aligned} v_L(F) &= E(IF^2) \\ &= \sum_{j=1}^n \sum_{i=1}^n \frac{c_i c_j E_F[(\alpha_i - H(F^{-1}(\alpha_i) - x))(\alpha_j - H(F^{-1}(\alpha_j) - x))]}{f(F^{-1}(\alpha_i)) f(F^{-1}(\alpha_j))} \\ &= \mathbf{c}' \Sigma^{\mathbf{IF}} \mathbf{c}, \end{aligned}$$

where

$$\Sigma^{\mathbf{IF}} = \left\{ \frac{\min(\alpha_i, \alpha_j) - \alpha_i \alpha_j}{f(F^{-1}(\alpha_i)) f(F^{-1}(\alpha_j))} \right\}_{i,j=1}^{i,j=n}. \quad (4.26)$$

The implication of this result is that one should be able to estimate $f(F^{-1}(\alpha_i))$ to obtain the variance, but it might be impractical because estimating the value of density at a large α from the given sample is not easily available, as discussed in the previous section.

Recalling however that we are interested in estimating the variance of bootstrapped CTE, a closer look at the EB formula given in Theorem 3.3 sheds a new light. From this theorem we first note that the EB quantile again belongs to the L-estimator class given in (4.25). Secondly, since each weight $w_{i(r)}$ is a small piece of a Beta density function, the score function $J(u)$ for the EB quantile is smooth (continuous and bounded) for EB quantiles. This observation suggests that the IF of the bootstrapped quantile may not involve the density term, allowing a direct estimation of the variance using the given sample. The following shows that this indeed is the case.

Lemma 4.7 *The IF of the bootstrapped α -th quantile, $E(X_{(r)}|\hat{F}) = \mathbf{w}'_r \mathbf{X}_{\cdot n}$, $\alpha = r/n$, is given by*

$$\frac{1}{B(r, n-r+1)} \left[\int_{-\infty}^x F(y)^{r-1} (1-F(y))^{n-r} dy - \int_{-\infty}^{\infty} F(y)^{r-1} (1-F(y))^{n-r+1} dy \right] \quad (4.27)$$

Proof:

From Theorem 3.3, the EB quantile is defined by

$$t(F) = \int_0^1 J(u) F^{-1}(u) du, \quad (4.28)$$

where $J(u) = u^{r-1}(1-u)^{n-r}/B(r, n-r+1)$. Then the IF of $t(F)$ is

$$\int_0^1 J(u) L_t(x|F) du = \int_0^1 J(u) \left[\frac{u - H(F^{-1}(u) - x)}{f(F^{-1}(u))} \right] du,$$

from Lemma 4.6. Using the variable transformation $y = F^{-1}(u)$, this becomes

$$\begin{aligned} & \int_{-\infty}^{\infty} J(F(y))(F(y) - H(y - x))dy \\ &= \int_{-\infty}^x J(F(y))dy - \int_{-\infty}^{\infty} (1 - F(y))J(F(y))dy. \end{aligned}$$

Replacing $J(u) = u^{r-1}(1 - u)^{n-r}/B(r, n - r + 1)$ completes the proof. It is easy to verify that the expected value is zero using differentiation. We emphasize that the IF is a function of x . Q.E.D.

The importance of this result lies in the fact that it shows the exact influence curve of the bootstrapped quantile which in turn can be extended to any bootstrapped L-estimator; note that the final form does not involve the density f , leading to easy implementation with a sample. In the nonparametric setting, (4.27) is computed by replacing F with \hat{F} , and x with x_j , the j -th observation of the sample:

$$\begin{aligned} & \frac{1}{B(r, n - r + 1)} \left[\sum_{i=1}^{No. obs \leq x_j} (X_{(i)} - X_{(i-1)}) \left(\frac{i}{n}\right)^{r-1} \left(1 - \frac{i}{n}\right)^{n-r} \right. \\ & \quad \left. - \sum_{i=1}^n (X_{(i)} - X_{(i-1)}) \left(\frac{i}{n}\right)^{r-1} \left(1 - \frac{i}{n}\right)^{n-r+1} \right]. \end{aligned}$$

where $X_{(0)}$ is taken as 0 for non-negative loss data. For implementation the following approximation⁴ is recommended:

$$\int_{\frac{i-1}{n}}^{\frac{i}{n}} u^{r-1}(1 - u)^{n-r} du \approx \frac{1}{n} \left(\frac{i}{n}\right)^{r-1} \left(1 - \frac{i}{n}\right)^{n-r} \quad (4.29)$$

and therefore the empirical version of the bootstrapped quantile's IF (4.27) is com-

⁴The right side of (4.29) often produces 0 for large n and r in mathematical software whereas the left side is computed by the incomplete beta function which is more reliable.

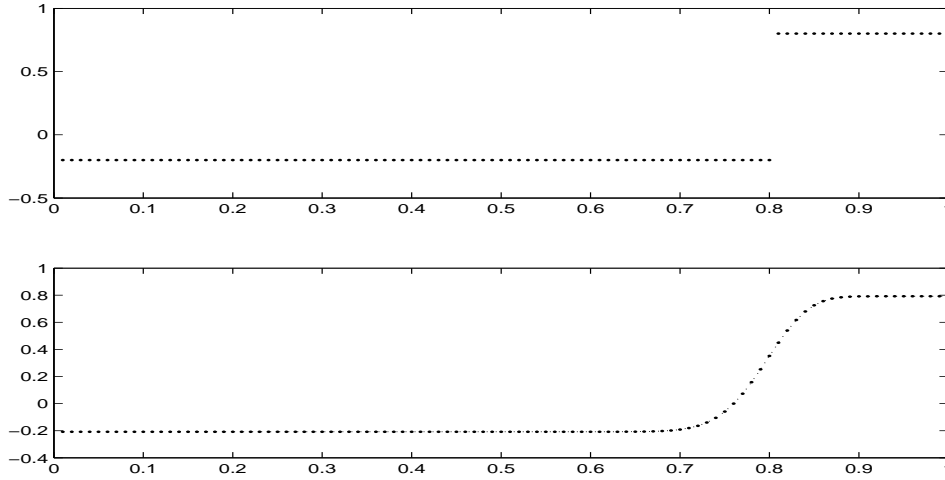


Figure 4.2: IF of quantile: Empirical (top) vs. EB (bottom) for $U(0, 1)$, $\alpha = 80\%$

puted from sample as

$$\frac{n}{B(r, n - r + 1)} \left[\sum_{i=1}^{No. obs \leq x_j} (X_{(i)} - X_{(i-1)}) \int_{\frac{i-1}{n}}^{\frac{i}{n}} u^{r-1} (1-u)^{n-r} du - \sum_{i=1}^n (X_{(i)} - X_{(i-1)}) \int_{\frac{i-1}{n}}^{\frac{i}{n}} u^{r-1} (1-u)^{n-r+1} du \right], \quad (4.30)$$

for each x_j , $j = 1, \dots, n$. Figure 4.2 compares the IF of a quantile before and after bootstrapping. The bootstrap indeed makes the quantile estimator qualitatively robust because the IF is now continuous and bounded. The above result is easily extended to the whole L-estimator class, giving a powerful way to estimate the variance of any bootstrapped L-estimator because any L-estimator is a linear combination of single order statistics. Furthermore the influence function of the bias-corrected sample quantile can be readily computed in a similar fashion. The following corollary is immediate from the above lemma.

Corollary 4.8 *The IF of the bootstrapped L-estimator $\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}}$, with $\mathbf{c} = (c_1, \dots, c_n)'$, is given by*

$$\sum_{k=1}^n \frac{c_k}{B(k, n-r+1)} \left[\int_{-\infty}^x F(y)^{k-1} (1-F(y))^{n-k} dy - \int_{-\infty}^{\infty} F(y)^{k-1} (1-F(y))^{n-k+1} dy \right]$$

Proof: First note that

$$\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} = \sum_{k=1}^n c_k \mathbf{w}'_k \mathbf{X}_{:\mathbf{n}}.$$

Each term $\mathbf{w}'_k \mathbf{X}_{:\mathbf{n}}$ on the right side is the bootstrapped quantile. The result follows immediately from (4.27). Q.E.D.

As an example, the IF of the bootstrapped CTE at α is obtained by setting $\mathbf{c} = (n - n\alpha)^{-1}(0, \dots, 0, 1, \dots, 1)'$ and given by

$$\sum_{k=n\alpha+1}^n \frac{(n - n\alpha)^{-1}}{B(k, n-r+1)} \left[\int_{-\infty}^x F(y)^{k-1} (1-F(y))^{n-k} dy - \int_{-\infty}^{\infty} F(y)^{k-1} (1-F(y))^{n-k+1} dy \right]$$

Again, one can use (4.30) to compute each term in the summation. Figure 4.3 compares the IF of the CTE and the bootstrapped CTE. The sharp edge has been smoothed out in the EB case. However the CTE itself is not robust by definition. We are not able to comment on its systematic behavior change after bootstrapping.

We note that even though the IF of the bootstrapped L-estimator is developed to estimate the variance of the bootstrapped L-estimator, one can use this to estimate the variance of ordinary (non-bootstrapped) L-estimator as well.

Consider a functional $t(F)$. The key idea of the nonparametric delta method in (4.18) states that $L_t(x|F)$ can be estimated by $L_t(x|\hat{F})$, based on the convergence of

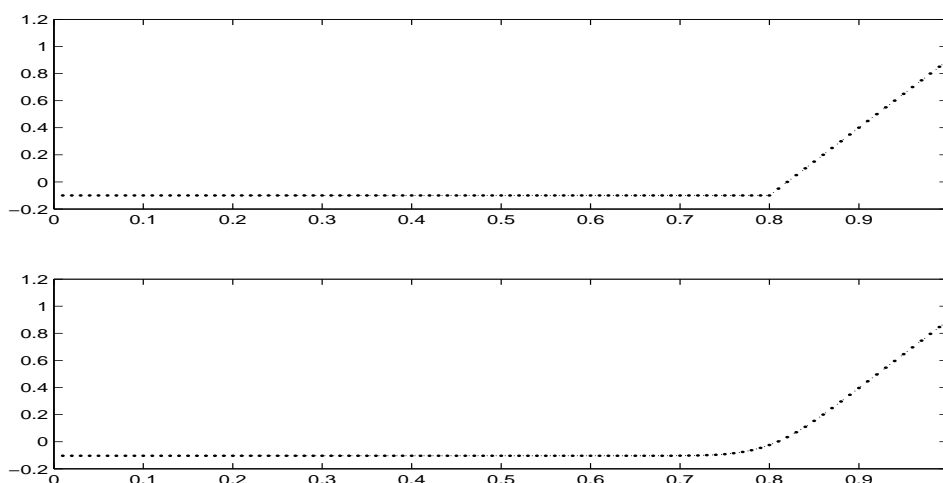


Figure 4.3: IF of CTE: ordinary CTE (top) vs. EB (bottom) for $U(0, 1)$, $\alpha = 80\%$

\hat{F} to F . Since \hat{F}^{EB} also converges to the true F , estimating $L_t(x|F)$ using $L_t(x|\hat{F}^{EB})$ can be likewise justified. This gives a powerful alternative way to estimate the variance of a statistic t where applying the nonparametric delta method is difficult. For example in estimating the variance of a (non-bootstrapped) quantile there is no practical way to compute its empirical IF due to the density form involved; we can use the IF of the EB quantile to estimate the variance, which is straightforward. Similarly in estimating the variance of the CTE one can use the IF of the CTE in (4.21) which can be easily computed; one may instead choose to use the IF of the EB CTE in (4.31) to estimate the same variance.

4.5 Numerical example

For the simulation study we consider the same three parametric models discussed in the previous chapter, namely, the LN put, RSLN2 put, and the Pareto loss. Since our focus is on the nonparametric estimation of the bootstrapped L-estimator, we consider $Var(\widehat{CTE}^{EB})$, with the CTE level set at 95%, for each model. In particular, the influence function (IF) based nonparametric delta method developed in the previous section is compared with the variance estimate using the ordinary bootstrap (OB) using 1,000 resamplings for each sample, under the MSE criterion. We report the results for sample sizes 200 and 1,000, respectively along with standard errors (s.e.), using 4,000 sets of simulations; Table 4.1 shows the numbers in percentage of the true variance of each model.

From the result we make the following comments.

- For samples of size 1,000 or less, the bootstrap method is preferred to the IF method, except for the Pareto case with sample size 1,000; but the differences get smaller as the sample size increases, as we expect.
- The Pareto loss model, the heaviest tail among three examples, shows that, as sample size increases from 200 to 1,000, the actual rMSE values decrease, but if expressed in percentage of the true variance, rMSE gets larger. This indicates that for the CTE variance estimate, the convergence of its variance is slower than that of the estimate as sample increases. The rMSE values under both methods indicate the variance estimate for this model is highly unstable.
- In all cases the MSE of these variance estimators are large even for samples of size 1,000, making both methods less than satisfactory for use in practice. This

LN put Model		Using OB			Using IF		
n	True val.	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
200	26.0479	-5.55%(0.64%)	40.65%(9.62%)	41.02%	3.59%(0.73%)	46.43%(12.56%)	46.56%
1000	5.3614	-1.14%(0.33%)	21.15%(0.54%)	21.18%	0.55%(0.34%)	21.36%(0.55%)	21.37%

RSLN put Model		Using OB			Using IF		
n	True val.	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
200	28.4564	-4.07%(0.64%)	40.48%(10.43%)	40.69%	4.20%(0.72%)	45.84%(13.37%)	46.03%
1000	5.8420	-1.01%(0.33%)	20.94%(0.57%)	20.97%	0.63%(0.34%)	21.30%(0.59%)	21.31%

Pareto loss		Using OB			Using IF		
n	True val.	Bias (s.e.)	Std (s.e.)	rMSE	Bias (s.e.)	Std (s.e.)	rMSE
200	132.8100	-9.33%(2.24%)	141.74%(596.67%)	142.04%	-2.96%(2.31%)	146.07%(633.74%)	146.10%
1000	26.8802	2.63%(3.08%)	194.48%(227.37%)	194.50%	3.50%(2.86%)	180.92%(196.77%)	180.96%

Table 4.1: Variance of the EB CTE 95% for three examples

is because the estimator is in the tail region. However, as we discuss in Chapter 5, using more simulation is often not feasible in practice.

- We repeated the same exercise for CTE 99% and found the same message, only this time the variance estimates are more unstable, leading to much bigger MSEs.

Since both nonparametric methods presented here rely on the same asymptotic argument, we generally cannot consider one over the other without a theoretical justification, and as far as we know there is no such one. Considering that the OB for variance needs more bootstrap resamplings than for mean, we believe that the IF method can sometimes be advantageous because it does not involve any simulation. The numerical examples shown in this section basically show that both the OB and the proposed IF method work reasonably well, and confirm that estimating the CTE variance requires more simulations for fat tailed distributions, although increasing the sample size may not be feasible in practice.

4.6 Concluding remarks

In this chapter we derived the influence function (IF) of the EB quantile estimate and showed it exists in an analytic form where no density function needs to be computed. The result directly extends to the whole L-estimator class. Using the finding we estimated the variance of the bootstrapped CTE and compared it against the ordinary resampling method (OB), through numerical examples.

These examples indicate that, if the model's tail is reasonably heavy, the OB method may be preferred over the IF method for the variance of the bootstrapped

CTE, even though its superior performance for small sample is unknown in principle. The result also indicates that the estimated variance in both methods are quite unstable in terms of its MSE, warranting further research on this topic. A possible improvement could be obtained by bringing in the second term of the von Mises expansion, though the second functional derivative of the CTE is unknown at this time. Alternatively a semi-parametric approach, such as kernel density, can be investigated to estimate the CTE and its variance, with no reference to the true distribution.

It should be noted that even though the finding has been used for variance of the bootstrapped L-estimator, it can also be applied to non-bootstrapped empirical estimator because both \hat{F} and \hat{F}^{EB} converge to F . In particular, since we verified that the bootstrap indeed makes an empirical quantile estimator qualitatively robust in the sense of, e.g., Staudte and Sheather (1990)(see Section 3.2.4), one can use the bootstrapped quantile's IF instead of the empirical quantile's IF, which is discontinuous and involves the density function of the population. If the quantile is around the centre of a distribution, such as median, we conjecture that this method can produce a more reliable result.

Chapter 5

A simulation study

5.1 Introduction

This chapter is devoted to a simulation study by applying the theoretical findings in the previous two chapters to a real industry sized numerical example. Let us briefly explain the usage of the stochastic simulation in life insurance business. By stochastic simulation actuaries mean that the uncertainty or risk associated with a certain product is assessed using scenario-based Monte Carlo simulations, rather than through the factor-based formula described in the first Chapter. One can easily appreciate the advantages of the simulation approach over the factor-based one in principle, and this is the trend of the current risk management practice across different industries. However the stochastic approach requires accurate models for both asset and liability projection. The setup of a well-functioning system can be costly as well, because the uncertainty can be best measured only when all the separate models - asset, liability,

administration, etc - are incorporated¹.

When it comes to cash flow projection to assess the risk and capital requirement for a life insurance portfolio, typically the following four steps are taken:

1. Random economic scenario sets are generated from a specified model. There are many models available to describe the long-term behavior of various economic variables, such as stocks, bonds, inflation, etc. Models could be in parametric form, with parameters estimated from historical data, or they could simply sample the historical data itself. Sometimes regulators provide a set of pre-packaged scenarios for insurers without expertise to do the stochastic simulation; see, e.g., the Report of Life Capital Adequacy Subcommittee (2005) of the AAA.
2. Each generated scenario is input to the insurance product model. The product model consists of a huge number of variables. It would include pricing variables, valuation variables, reinsurance treaties, taxes, and other regulations, for each specified cohort. For the inforce business each policy's information will be fueled through the administration system and for new business some suitable assumptions are made. Except for some variables dependent to economic environment, such as lapse, the product model is often deterministic because insurance liability risks are assumed diversifiable, though modifications are common for sensitivity analysis or regulatory reporting.

3. Running the product model under the given scenario will generate output values,

¹Typically insurers have different IT systems for different products and purposes. If a company acquires a business blocks from other insurers or regulations change, the current system needs to be updated as well.

i.e., future cash flows streams, such as surplus, profits, tax amounts, embedded values, and the return on capital, etc. This is the most time-consuming step.

4. Repeat the above steps as necessary with a new random scenario set. A sample obtained after repetitions is then considered to be i.i.d. and can be used for further statistical inference.

One of the problems, as explained in the previous chapter, assuming that the scenario generating model and liability model are reasonably accurate, is that simulating future cash flows can take a significant amount of time for each scenario. This prevents actuaries from generating large enough samples for accurate estimation, particularly for tail risk measures. The issue of heavy computation and slow simulation run times are well recognized in actuarial community; see the attempts by, e.g., Chueh (2002) and Christiansen (1998) to get around this problem by using “representative scenarios”. The Report of Life Capital Adequacy Subcommittee (2005) of the AAA provides a guideline to use the representative scenarios. However this technique is not proven; it is easy for one representative scenario to become a scenario that is not representative any more with a slight change in any model.

In the previous two chapters we presented small-scaled numerical examples by applying the theoretical findings on the CTE bias in Chapter 3, and on the variance estimation of the bootstrapped CTE in Chapter 4. In this chapter we take an industry-scaled model to repeat the same exercise incorporating both chapters’ findings simultaneously. The example presents a typical challenge for actuaries in practice, thus becomes a good illustration of the work in the practical context of small i.i.d. simulation sample.

Another problem is about the model itself. If the assumed models fail to represent

the true risk of the company or the economic variables, statistical inferences based on the model-generated sample are basically wrong, not because of specific inference methodology, but because the estimates will not produce what they are supposed to. This aspect of risk is discussed later in this chapter in general terms.

5.2 The model

For our simulation study we consider a block of segregated fund of an anonymous Canadian life insurer, implemented using AXIS. Segregated funds, also known as variable annuities in the US, are hybrid contracts of insurance and investment. Under the basic contract, the single premium is invested in mutual funds and monthly administration fees are deducted from the fund. Many segregated funds offer a Guaranteed Minimum Maturity Benefit (GMMB) at the end of the contract term and a Guaranteed Minimum Death Benefit (GMDB) during the contract term, with guarantee level being commonly set at between 75% and 100% of the underlying fund's initial value. There are other guarantees and modifications to this product such as a guaranteed reset, a guaranteed minimum income; see Chapter 1 of Hardy (2003) for further details.

AXIS², the dominating actuarial system for life insurers in Canada, is used for pricing and product development, as well as financial projections, surplus adequacy testing, valuation and asset/liability modeling. The implemented block is basically a collection of model points called 'cells' in AXIS. Each cell represents different co-

²AXIS is the product of Gilliland, Gold and Young (GGY), based in Toronto. It is currently used in over 80 companies worldwide (See www.ggy.com), and was named one of most frequently used cash flow testing softwares in North America; see an SoA survey by Zhang *et al.* (2005).

hort (or segment) of segregated fund policyholders; for instance a certain cell might represent the following characteristics:

- Mail smoker in age band 45-55
- Single premium under 1 million dollars
- 75% GMDB and 100% GMMB
- The premium is invested in 30% bond, 40% TSX index, and 30% S&P 500 with a specified reinvestment strategy throughout the contract term
- Resets are allowed at the end of each 3 years

In fact, within each cell hundreds of assumptions are built in to employ various pricing and business characteristics including the above profile. Since there are so many different combinations possible, a dataset representing one product block ends up involving several hundred cells. Figure 5.1 shows two screen shots of AXIS at a cell level for illustration. In this study we focus on the GMMB guarantee cost and set the guarantee level at 100%³ of the initial fund value.

AXIS has a deterministic structure for the liability side except for some economic-dependent variables such as lapse or reinvestment strategy, but it allows users to choose from several long-term stochastic asset models to project future cash flows distribution. Among these we chose economic scenarios from Wilkie stochastic model; see, for original papers, Wilkie (1986) and Wilkie (1995). The Wilkie model is suitable for long-term economic variables prediction and is widely used in the UK and elsewhere. It uses so called the cascading structure where one variable is integrated with

³For confidentiality, we made this override for all cells.

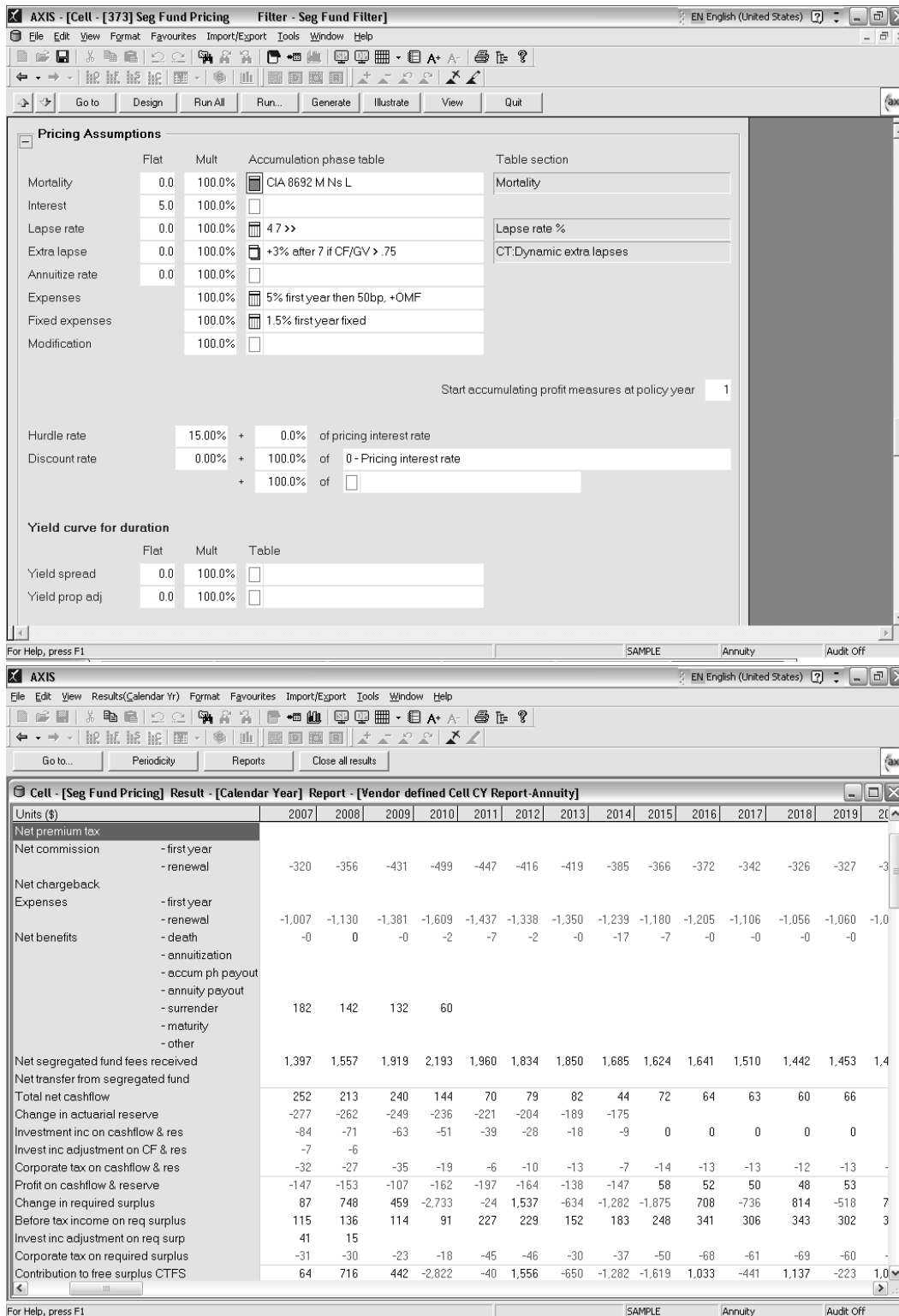


Figure 5.1: Cell assumption screen (top) and cash flow projection (bottom)

other variables plus random fluctuations; see Hardy (2003) and Panjer *et al.* (2001) for an introductory description of the model. We selected the Wilkie model derived from Canadian data, as implemented in the AXIS model, to generate scenarios. The generated scenarios are then used as an input to the liability model to project future cash flows, one scenario at a time, which is the most time-consuming step.

5.3 Simulation backgrounds

For this specific block, it took approximately 25 minutes to project cash flows for one scenario, based on pilot simulations; 1,000 scenarios would require about 416 hours on a personal computer. If several blocks are to be combined the run would take proportionately longer. For this reason we reduced the number of policies significantly to produce the simulation result in a manageable run time. The run time for this reduced model, which consists of 1,055 policies, is approximately 30 seconds for each Wilkie model scenario. We obtained 2,000 scenario runs -which we will call the 'full data' - and computed the empirical CTE at 95% of the GMMB guarantee cost which will be considered to be the benchmark value for our analysis.

Note that all the simulations in this chapter have been performed at the headquarters of the GGY. The computer used has a Pentium 4 2.53GHz processor with 1GB of RAM. While AXIS has a function that allows computing on more than computer simultaneously, which can significantly reduce the simulation times, it is not always possible to have extra computers available in practice. Also it is not hard to speculate that the complexity of models will always keep up with the improvement in computers.

Mean	95.31
Variance	8,699.20
Stand. dev.	93.27
Skewness	4.49
Kurtosis	33.98
Minimum	10.23
Maximum	1,206.90
Empirical CTE 95%	408.04

Table 5.1: Basic statistics based on the full data (Unit: CDN dollar)

5.4 Analysis

Table 5.1 and Figure 5.2 show some basic information from the full data. Each observation represents the cost of the GMMB liability for each scenario run. From its histogram it is evident that it is right skewed, like many loss distributions that actuaries manage.

We started with 200 scenario runs to estimate the CTE 95% using the proposed guideline for three candidates as in Chapter 3: the empirical estimator (\widehat{CTE}), the EB estimator (\widehat{CTE}^{EB}), and the bias-corrected estimator ($\widehat{CTE}^{EB.bc}$). We repeated the same exercise adding additional 200 runs each time, up to 1,000, which would be around the maximum number of simulations used in practice.

For each sample, the bias and the variance (and thus the rMSE) are estimated using the guideline proposed in Section 3.6. A separate set of variance estimates based on the influence function (IF) was also computed, based on the developments in Chapter 4. It is assumed that the true CTE is unknown. The results are presented in Table

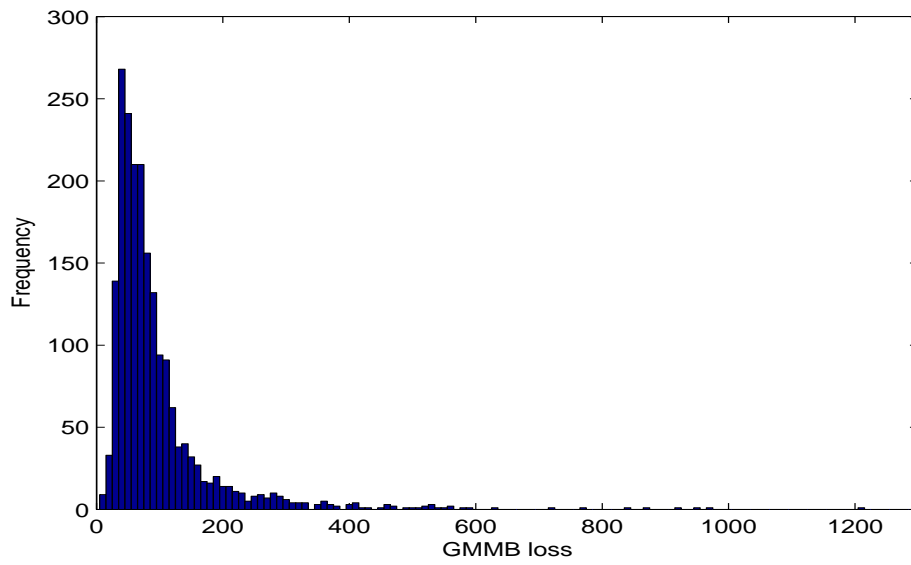


Figure 5.2: Histogram of GMMB loss based on the full data

5.2 along with standard errors⁴.

A few comments on the result follows:

- The optimal estimator among the three candidates would be the \widehat{CTE}^{EB} for sample size 200 and 400. The standard deviation derived from both IF and OB show the same relative rankings.
- As the sample size gets larger the best estimator switches, and the two methods, IF and OB, indicate different estimates.
- Also as the sample size increases, the difference between the two methods starts to disappear; at sample size 1,000 the two methods produce very similar rMSE values, even though the winners are different.
- One can see that the estimated CTE dramatically changes in between sample size 400 and 600 due to one extreme value. Actually, even at size 1,000 the estimated CTE is still well off the benchmark CTE value 408.04, and the estimated standard deviation does not look small enough.

5.5 Alternative model

Assuming that the model is adequately correct the bootstrap seems to give a reasonable method to estimate the tail risk measures with no reference to the true loss distribution, especially when the structure of the model and its *input* is unknown and hard to interpret. From the analytic form of the EB of L-estimator, we can think of

⁴This is computed by normal approximation.

Sample size: 200						
Estimator	Estimated value	Estimated bias	Est. std.		Est. rMSE	
			OB (s.e.)	IF (s.e.)	OB	IF
\widehat{CTE}	428.929	-6.660	80.913 (55.55)	89.405 (61.38)	81.187	89.653
\widehat{CTE}^{EB}	422.269	-13.321	78.127 (53.64)	84.976 (58.34)	<u>79.255</u>	<u>86.014</u>
$\widehat{CTE}^{EB.bc}$	435.589	0	83.782 (57.52)	94.025 (64.56)	83.782	94.025
Sample size: 400						
\widehat{CTE}	394.995	-2.163	55.461 (31.59)	57.085 (32.52)	55.503	57.125
\widehat{CTE}^{EB}	392.832	-4.326	54.672 (31.14)	56.038 (31.92)	<u>54.843</u>	<u>56.205</u>
$\widehat{CTE}^{EB.bc}$	397.158	0	56.265 (32.05)	58.163 (33.13)	56.265	58.163
Sample size: 600						
\widehat{CTE}	471.868	-2.268	59.636 (30.56)	59.756 (30.62)	59.679	<u>59.799</u>
\widehat{CTE}^{EB}	469.600	-4.536	59.287 (30.38)	59.650 (30.57)	<u>59.460</u>	59.822
$\widehat{CTE}^{EB.bc}$	474.136	0	59.992 (30.74)	59.874 (30.68)	59.992	59.874
Sample size: 800						
\widehat{CTE}	462.519	-1.248	47.339 (22.53)	46.675 (22.21)	47.356	46.692
\widehat{CTE}^{EB}	461.271	-2.497	47.248 (22.48)	46.816 (22.28)	<u>47.314</u>	46.883
$\widehat{CTE}^{EB.bc}$	463.767	0	47.436 (22.57)	46.542 (22.15)	47.436	<u>46.542</u>
Sample size: 1000						
\widehat{CTE}	440.938	-1.411	39.438 (17.73)	39.574 (17.79)	<u>39.464</u>	39.599
\widehat{CTE}^{EB}	439.527	-2.821	39.382 (17.70)	39.389 (17.70)	39.483	<u>39.490</u>
$\widehat{CTE}^{EB.bc}$	442.348	0	39.498 (17.75)	39.764 (17.87)	39.498	39.764
\widehat{CTE}^{GPD}	Estimated CTE: 443.89			Estimated s.e. of CTE: 86.33		

Table 5.2: Simulation result of the Seg. fund model

the CTE bootstrapping as a special way of smoothing. In this section, we consider an alternative parametric model, as another way of smoothing, to estimate the CTE based on the first 1,000 scenario runs, which we will call the ‘data’ throughout this section. This exercise may give us a further insight on the data (e.g., tail behavior), and allow us to identify and translate the scenarios linked to the extreme outputs. In the presence of other guarantee options, different driving forces and their interactions will affect the output, making it harder to translate the corresponding scenarios.

First, recognizing its thick tail, let us consider a log transform. In Figure 5.3 the Q-Q plot of the transformed data shows that the tail is substantially heavier than the normal. So, for example, fitting Lognormal to the original data would result in significant underestimation of tail risk measures. In parametric approach one can choose to fit the whole data with a list of known heavy tailed distributions⁵, in which case estimated models using standard parameter estimations such as maximum likelihood or method of moments, however well they fit, generally show a better fit around the center of the distribution than in tails. Alternatively one could fit the tail region only for a given model, e.g., using likelihood of truncated distribution. Instead of taking this path, we try to fit the tail region directly using a tool in Extreme Value Theory (EVT). By using this technique, we implicitly assume that the tail events can be considered separately from the other sample values, thus we can ignore the observations below a certain threshold, from which tail starts. A successful application of EVT however needs some specific assumptions on the underlying distribution’s tail behavior that can be hardly observed or justified in practice; see for example McNeil *et al.* (2005) or Embrechts *et al.* (1997) for technical details. This is a general issue with

⁵This exercise needs a well-defined model selection procedure, such as the SBC or AIC criterion.

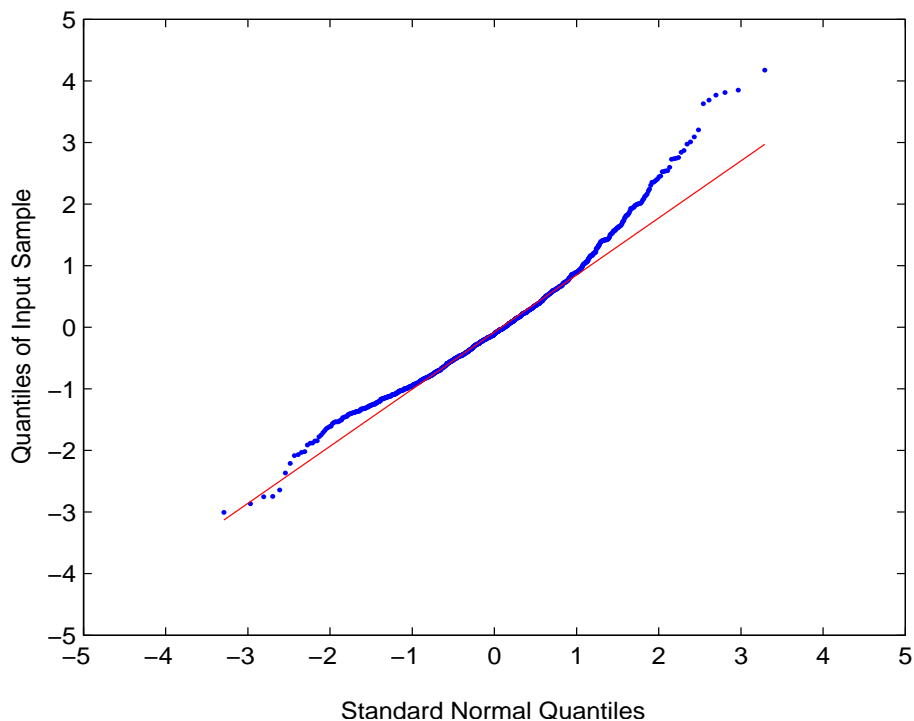


Figure 5.3: Standardized Q-Q plot after log transform

tail estimation and there is still no definite solution to it.

In EVT, Generalized Pareto Distribution (GPD) has been widely utilized to fit exceedances over threshold. Here we adopt this model to fit the tail of the data as an alternative model among other methods to fit tails. The distribution of GPD is defined as

$$F_X(x) = 1 - \left[1 + \frac{\xi}{\beta} x \right]_+^{-1/\xi}, \quad \beta > 0.$$

So the domain is $x \geq 0$ if $\xi > 0$, or $0 \leq x \leq -\beta/\xi$ if $\xi < 0$; when $\xi = 0$ this reduces to the exponential distribution with mean β . It is well known that the mean excess

function with threshold u , defined by $e(u) = E(X - u|X > u)$, of GPD is a linear function of u :

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

Further for any $v \geq u$, we have

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

which is again linear on v . The sample mean excess function is thus used to assess the tail linearity and to identify the threshold. Figure 5.4 gives the sample mean excess function of the data. The picture indicates that except the last few observations the tail region seems to be close to linear⁶ from around 200. We set the threshold equal to the VaR 95% of the data, of which empirical estimate is 258.04, that is $u = 258.04$.

Using the maximum likelihood method, the fitted parameters are $\hat{\xi} = 0.248$ and $\hat{\beta} = 139.74$ with estimated covariance matrix

$$\hat{\Sigma} = \begin{pmatrix} 0.0393 & -4.629 \\ -4.629 & 1134.300 \end{pmatrix},$$

indicating a heavy tail⁷. The fitted exceedances of the data given in Figure 5.5 shows the selected model reasonably fits the data except the last few points. Since the threshold is set at the VaR 95%, The CTE 95% of the data is computed by

$$E(X - Q_{0.95}|X > Q_{0.95}) + Q_{0.95} = 185.85 + 258.04 = 443.89,$$

⁶It is common to omit the final few points from graphical consideration due to their highly unstable nature. In our case one should go back to the corresponding scenarios and investigate the source of extreme outcomes, but this task is impossible for our case.

⁷In GPD $E(X^k)$ is infinite for $k > 1/\xi$ for $\xi > 0$, thus only the first four moments may exist for the data if we take $\hat{\xi}$ as being correct. In practice, there will be a maximum limit, however big it is, for any insurance liability, leading to finite moments of all orders.

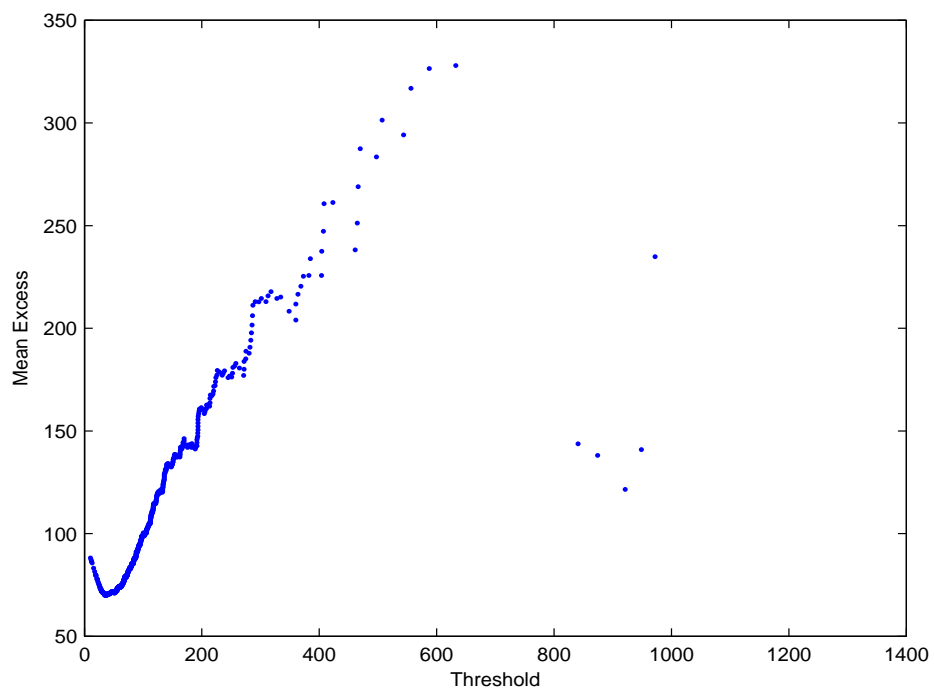


Figure 5.4: Sample mean excess function of the data

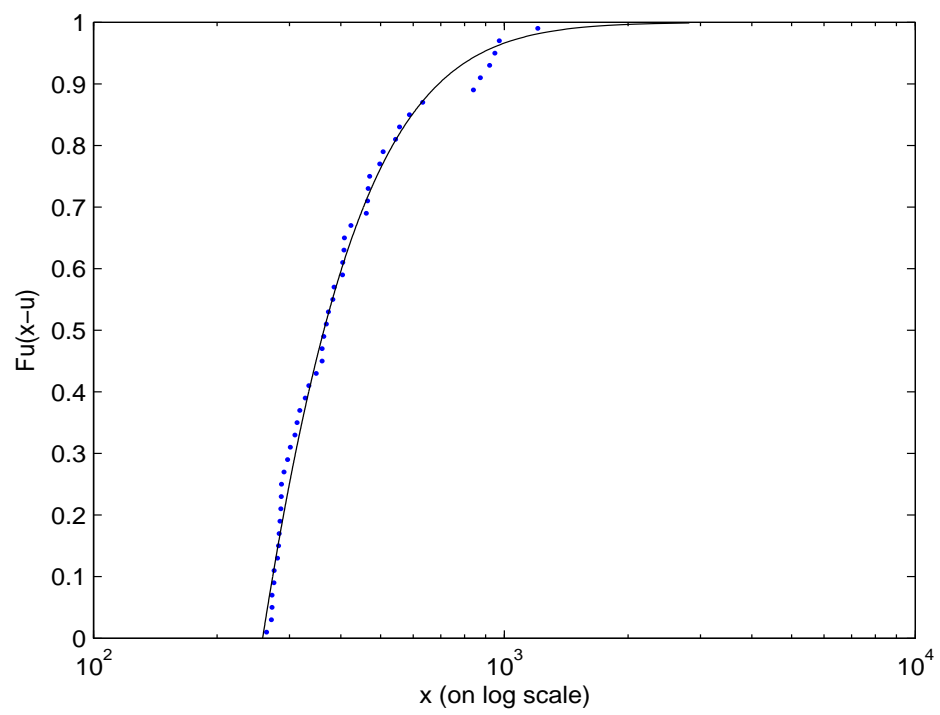


Figure 5.5: Fitted GPD for the data above threshold

where the mean excess term is the mean of the fitted GPD. The estimated standard error of the CTE is 86.33. The estimated CTE value is very close to the empirical (440.94) and bootstrapped value (439.53) with sample size 1,000 as shown in Table 5.2, even though the standard error seems bigger.

Since there are many other possible methods to fit the whole data or tail region, our comment is necessarily general. We saw that the tail of real actuarial data can be significantly heavy and it would be hard to find the best single model that can produce an accurate estimation of the CTE. Generally both parametric and nonparametric models have their strengths and weaknesses, and they can supplement each other. In our example, a detailed data analysis revealed more information on the data such as the tail thickness and a few points that are exceptionally extreme, leading to a possible identification of the driving forces of extremes. This information could give actuaries opportunities to *qualitatively* assess the risk within the product model as well as the model itself; this is something that a blind use of the nonparametric method, such as the bootstrap, could not offer.

There always are source of errors in modeling. Model error, parameter error, and in our context sampling error also plays an important role due to simulation constraint. Recognizing these uncertainties, insurance regulators in many cases use naive empirical approach and add extra cushion to make the estimate conservative, rather than prescribe certain parametric constraints. In the next section we discuss in general terms the uncertainty around the selected model and parameters as a warning to the audience about the potential significance of these risks in practice.

5.6 Model and parameter uncertainty

So far in chapter 3, 4, and 5 the focus is on the outputs simulated from the given liability model. The given model and the economic scenarios fused into the model have implicitly assumed to be correct by using the i.i.d. sample from an unknown true distribution F (therefore the the economic scenarios generator too has been assumed to be so). In reality F is a combination of a complicated liability model combined with the economic scenario model, both of which are man-made machineries. In principle there is no *correct* model in the sense that every model is a proxy or simplified version of reality, however well it fits the data, so there will be errors.

Consider our case study for example. For the economic scenario generator, there are alternative models to forecast future economic factors other than the Wilkie model, and we cannot have an absolute confidence in one model and thus in its outputs. Similarly, for the liability model to which the generated Wilkie scenarios are fused there is a risk that the product and company level variables are incorrectly or inadequately modeled. These uncertainties are called model error. Even in the case where the given model is proved to be the best for a specific purpose there still are many parameters to be estimated in the model which represents another source of error, called parameter error. These two errors are generic in any modeling procedure and could affect the quantity of interest depending on the problem in hand. In our case the uncertainty in the estimated tail risk measure could be a fair concern in this regard since both CTE and VaR involve the tail region where the events are rare by definition; the lack of available observations makes the problem harder since simulation ability is limited, as seen in this chapter. Thus in practice examining the underlying models and parameter sets is as important as estimating the tail risk mea-

sure under the assumption of an i.i.d. sample, because the accuracy of the estimated risk measures eventually depends on the accuracy of the model used to represent the true risk of the company.

To assess the sensitivity of the desired quantity to model and parameter errors, one generally needs to examine different models and parameter sets and see the variation of the result. To examine parameter uncertainty one can do the sensitivity analysis or scenario testing by changing parameter values, within the given model. If the model is highly complicated however the ability to do sensitivity analysis will be limited as well, as shown in our example. The model error however can be difficult to investigate unless alternative models are simple to implement and compare. In insurance practice these uncertainties are recognized and sometimes prescribed by regulators, e.g., the provision for adverse deviation (PfAD) of the individual life and annuity products and the segregated fund calibration requirement (see, e.g., Hardy (2003) Chapter 4). These are designed to produce a conservative result by padding additional cushions against variation of model and parameters. At the business line or the corporate company level, however, comparing different liability models has largely been ignored. This is because the insurance liabilities are often assumed deterministic through the law of large numbers; also building alternative models for complicated liabilities seems infeasible because of its large scale.

Finally we comment on tail behavior under model and parameter uncertainty. Using the Bayesian approach, Cairns (2000) and Hardy (2002) show actuarial examples where model and parameter uncertainty make a significant difference in the final quantity of interest. For the examples in these papers it is also shown that

incorporating these uncertainties tend to fatten the tail of loss distribution. In particular, the latter paper illustrates that the CTE of the segregated fund guarantee costs can be substantially higher for certain guarantees if parameter uncertainty is incorporated. If we think tail thickness as the degree of uncertainty of the quantity of interest, which is a random variable, it would be reasonable to say this observation would hold in general. This indicates that the true tail risk measure of a loss could be substantially larger than what we obtain from the sample generated through the single model, simple or complex, currently in place, leading to a riskier position of the company due to less capital than needed. Related to this, another important issue is to identify the drivers of extreme outputs in the model. We already mentioned that simulation is often the only way to understand the dynamics of insurance office models due to their complexity. There could however be dominant input factors in the model that drive extreme loss outputs depending on product, such as interest rate, stock prices, or lapse. Therefore it would be interesting to construct a statistical model that could mimic the behavior of the real insurance office model using relatively small, but important, input and output variables; a regression could be a feasible candidate for this purpose. For instance, the input factors could be predictor variables and would replace the whole liability model, and outputs, typically cash flow paths or some important financial indices, can be thought to be the response variables. While translation of input and output values into the regression framework needs care and more investigation due to its multivariate and time-dependent structure, the advantages of this modeling, if successful, can be beneficial in risk management; it can reveal the driving force of main risks of company's business and thus give much more information about the tail behavior of outcomes using a much simpler model than the original company model.

5.7 Concluding remarks

If we had known the exact distribution of the loss, we would have been able to select the optimal estimator. In practice, we need to estimate the CTE with only one sample from an unknown distribution, and often there is no definite winner, as seen in our example.

In this chapter we estimated the CTE 95% based on a small sized sample from an industry segregated fund model. In particular the bias and the variance has been estimated for the MSE approximation through both the bootstrap and the nonparametric delta method, for three different CTE estimators. We found that the optimal estimator based on the proposed guideline changes over sample size, though the two different variance estimates become closer as the sample increases, which supports that the two methods are consistent. We also found that even a size of 1,000 may not be big enough for an accurate CTE estimation for this specific model.

Through this example we justified the small i.i.d. sample situation and showed how the proposed guideline can be applied to a real industry data. Considering that in practice the true CTE value is unknown and the ability for large simulations is limited, the guideline offers a theoretically sound and economical way for actuaries to estimate the CTE by selecting the optimal estimator with a high probability. If the company's aggregate CTE is computed by adding the CTEs of each block of business, as the current regulation prescribes, the impact of the accumulation of the CTE bias will also be alleviated via the guideline, because bias already is a part of consideration in its mechanism. The bias, unlike the variance, cannot be reduced through diversification.

However, actuaries should be aware of the risks attached to the selected model itself and its parameters because the true tail risk measure could be significantly larger than what one obtains from the sample through the model in place due to misspecified model and (or) parameters, as discussed in the previous section. If one incorporates model and parameter uncertainty in stochastic modeling at corporate level, there will be no CTE aggregation and these uncertainties would serve as dominant factors in CTE estimation. The proposed guideline is still recommended in this context, but the impact could be marginal compared to the value change in tail risk measures in the presence of model and parameter uncertainty.

Chapter 6

A new capital allocation method

6.1 Introduction

Insurers, like other financial entities, balance two financial considerations; minimizing risk and maximizing return. In insurance business these can be translated into managing solvency and maximizing the embedded value. In the middle of these two conflicting objectives lies capital determination and its management. Excessive capital will protect policyholders at a safer level, but it will inevitably be detrimental to the profitability of shareholders and the value of the company as a financial investment. On the other hand, inadequate capital would increase the profitability for shareholders at the cost of a high risk of bankruptcy of the insurer, which would be a devastating consequence for policy holders. In general policy holders and regulators place constraints on value-maximizing strategies by capping the risk of the insurer, leading to different perspectives than that of the shareholders on how the capital is set and managed.

From a theoretical point of view these two different perspectives can be related to different probability measures; the P measure and the Q measure. Understanding and differentiating these measures can be important because they are used for different purposes and will produce different numbers even from the same business book.

Another important aspect of capital is its allocation to each line of business. Capital allocation can be useful in assessing the risk of each business line and is widely used in practice for many insurers. In particular the management of a company uses allocated capital as the basis of profitability and strategic planning for different business blocks within multi-line company; see Section 1.6 for details. Determining the allocated capital for a line is similar to determining the total capital of company in a sense that both try to do the same thing for a given business entity. They can be however quite different because capital allocation is an *ex post* process, given the total capital. While there are several popular allocation methods available, this topic is fairly new and there is no universal agreement on how to allocate the required capital and researchers still actively try to find the optimal allocation as part of the developments of risk management.

In this chapter we first review the idea of the two different probability measures and their application to valuation and risk management. Then we propose a new allocation method which is inspired by the notion of the solvency exchange option of the company. The proposed method has some interesting properties but it also violates one of the widely accepted allocation axioms. After a close examination we argue that this violation is not as bad as one might think, and further that the axiom itself is not aligned with economic reality.

6.2 Value vs. Risk

6.2.1 Two measures

The P measure, also referred to as the physical measure, states the real world probability. The P measure is the ordinary measure that we call “probability” everyday and becomes the basis of our model of the future behavior of the uncertain outcomes and risks attached to them. The P measure is used in most risk management actions including the projections of the company’s future cash flows, the solvency position testing, and the determination of required capital and its allocation.

The Q measure, also known as the equivalent martingale measure or the risk neutral measure, has been developed to assist in setting the no arbitrage price of financial assets. It is an artificial probability measure based on cash flow replications. It is, generally, not the same as the P measure, so it does not model the uncertain future outcomes for the risk. It is designed to produce current values of financial assets, not to predict their future behavior. The risk neutral measure is thus utilized in valuation of derivatives and also the valuation of a financial liability consistent with no arbitrage principle.

In theory the no-arbitrage condition of a market is actually equivalent to the existence of at least one risk neutral probability measure in the market. We also know that the uniqueness of the risk neutral measure is equivalent to the completeness of the market; see, e.g., Panjer *et al.* (2001). Here the complete market means the market where any cash flow can be replicated by constructing a suitable portfolio. This relationship reversely indicates that it is possible to have a series of different Q

measures if the market is incomplete, which is the case for most insurance business. These are among central concepts in financial economics.

Even though it has been said that the Q measure is used to price, the P measure can also be used in pricing. In fact, P measure pricing, using equilibrium pricing, is the only pricing method if the desired cash flows are not replicable using traded securities. P measure pricing takes risk directly into consideration through the use of utility function, whereas with the Q measure the risk adjustment is needed to reflect the constructed portfolio's risk free characteristic.

In insurance business two main issues shared by its stake holders would be valuation and solvency. The current trend of accounting principles is to move toward a market value (MV), or fair value basis, which means that the valuation process is naturally related to the Q measure. On the other hand solvency is dealt with by using real world probabilities, reflecting the actual distribution of surplus, which leads to the estimated insolvency risk of the company.

6.2.2 Valuation perspective

The valuation of assets is relatively easy and well established because items on the asset side of the balance sheet are mostly tradable securities including stocks, bonds, and mortgage backed securities. For most insurance liabilities (or products) however the secondary market does not exist; exceptions would be reinsurance, transactions of a business block between insurers, securitized insurance products, and viatical settlements¹, all of which do not trade everyday, so MVs are not easily established. Because

¹A viatical settlement is the sale of life insurance policy by the policy holder at discounted face amount for immediate cash settlement.

they are not traded, insurance liability cash flows cannot in general be replicated by other securities in the market, making the insurance market incomplete. Hence determining the unique Q measure for insurance liability is generally not possible.

Assuming the Q measure is available for the given liability model, let us consider a company, which just started its business, and characterized by the following financial variables. For simplicity the model is examined over a single time period, that is time 0 to time 1:

- The insurance loss (or liability) will be realized at time 1 by random variable L and its risk neutral value is set at $V(L)$ at time 0.
- We assume that the MV of the whole asset at time 0, $V(A)$, is the sum of premium collected and shareholder's capital and is invested in a series of different financial securities at a random rate of return. At time 1 the asset amount will be realized by random variable A .

Table 6.1 shows the ordinary form of the insurer's balance sheet at time 0 and 1. Throughout this chapter we differentiate between the capital and the surplus, even though they are used interchangeably in many practical situations. The surplus is the profit from the insurer's operation and is defined by the premium plus investment income less the loss payout. The capital is the amount of money that is provided by the shareholders. Shareholders consider the surplus as the profit from their capital investment and expect it to be distributed (or at least part of it) as dividends.

Recalling that the values at the end of time period are random we can think of two possible outcomes at time 1; the insurer stays solvent, $A \geq L$, or it becomes insolvent, $A < L$. If the insurer is solvent at the end of the period, the liability L is

	Time 0	Time 1
Asset	$V(A)$	A
Liability	$V(L)$	L
Capital and surplus	$V(A) - V(L)$	$A - L$

Table 6.1: Ordinary balance sheet of an insurer

	Shareholders get	Policy holders get
When solvent ($A \geq L$)	$A - L$	L
When insolvent ($A < L$)	0	A

Table 6.2: Entitlements of shareholders and policy holders at time 1

expected to be fully paid and the excess amount of asset over liability, $A - L > 0$, will belong to shareholders. On insolvency the liability is not fully paid and this affects both policyholders and shareholders. The insurer would liquidate the asset A and distribute it to the policy holders; the shareholders get nothing but they are not liable to pay any further loss either. Hence policy holders get something less than they expected - they expected to get L but only get A , which is less than L - and shareholders lose everything they invested in the insurer. Table 6.2 describes this situation.

This limited liability to policy holders indicates that the actual liability payoff -which we call the economic liability from now on - should be less than the gross-up liability L because it is possible for the insurer to fail to pay all L . The economic liability thus can be expressed in a conditional form with conditioning on solvency status:

	Time 0	Time 1
Asset	$V(A)$	A
Liability	$V(L) - V(D)$	$L - D = \min(A, L)$
Capital and surplus	$V(A) - V(L) + V(D)$	$A - L + D = \max(A - L, 0)$

Table 6.3: Economic balance sheet of an insurer

$$L - \max(L - A, 0) = L - (L - A)^+ = \begin{cases} L, & \text{if } A \geq L \\ A, & \text{if } A < L \end{cases} \quad (6.1)$$

Since $(L - A)^+$ is nonnegative the economic liability is less than the gross-up liability L . The fair price at time 0 of this economic liability requires the use of option pricing techniques, using the \mathbb{Q} measure, as seen from its resemblance to a call option payoff, written on L with strike at A . The quantity

$$D = (L - A)^+ \quad (6.2)$$

in particular is referred to the solvency exchange option because it swaps the asset and the liability depending on the solvency status of the insurer, as shown in (6.1). The price of this option is, assuming a constant risk-free rate r ,

$$V(D) = e^{-r} E^{\mathbb{Q}}[(L - A)^+], \quad (6.3)$$

which can be uniquely determined under the complete market assumption.

Table 6.3 shows the economic balance sheet of the insurer including the solvency exchange option. Note that the capital and surplus account increased due to the addition of $V(D)$, reflecting the limited liability of the shareholders.

6.2.3 Solvency perspective

Focusing on the CTE, among other tail risk measures, to ensure the solvency of insurers, there are two possible methods in setting the economic capital. The first one is based on liability side of risk with capital given by

$$E^P[L|L > Q_\alpha(L)] - \Pi(L), \quad (6.4)$$

where $\Pi(L)$ is a value of liability prescribed in regulations² or determined internally using, say, the mean of the loss. The obtained capital will be discounted to produce its current value using, say, the hurdle rate set by the management. There has been much literature devoted in finding the first term of (6.4) for many parametric models; see, e.g., Landsman and Valdez (2003), Landsman and Valdez (2005), Cai and Li (2005), and the developments in Chapter 2 of this thesis, but this type of CTE can be limited in its usage because the asset side plays no role in the capital determination here.³ The second method would involve

$$E^P[L - A|L - A > Q_\alpha(L - A)] \quad (6.5)$$

This form is especially suitable where asset is correlated to the liability, because $L - A$, the net loss, represents the true risk of the company. For example the segregated fund fees collected periodically, the asset of the balance sheet, will be a certain portion of the fund amount, which in turn is a crucial input for guarantee liability payoff at maturity. If one relies on (6.5) in setting the capital the optimal amount of asset would satisfy the equation

$$E^P[L - A|L - A > Q_\alpha(L - A)] = 0 \quad (6.6)$$

²In the segregated fund business in Canada, for example, $\Pi(L)$ is set at approximately the 75-th percentile of the distribution of L by regulation.

³All the CTE formulas for nonnegative random variables implicitly use this method.

which is typically solved through changing the capital amount - which will change the whole asset amount - heuristically. Note that the dynamics of liability and the remaining asset portion do not change. When the whole asset is already defined, the risk manager will compute the value of (6.5) to see its sign and the magnitude to assess the capital adequacy; a small negative number signals adequate but not excessive capital. However the CTE risk measure does not reflect the limited liability of the shareholders.

6.3 Allocation review

In Section 1.5 we introduced the motivation of capital allocation, and briefly discussed some proposed allocation axioms and simple examples. In this section, we revisit the axioms and examples to obtain a further insight needed for later developments. As before, we consider a multi-line P&C insurer with aggregate loss⁴ given by

$$X = \sum_{i=1}^n X_i,$$

where X_i represents each line's loss. To introduce the fair allocation axioms more formally we cautiously develop a notation, aligned with that used in Valdez and Chernih (2003), that can help understand the concept of allocation more easily. As mentioned earlier we can see a similarity and difference between the capital allocation and the total capital determination. Suppose that the capital is determined by a risk measure $\rho(\cdot)$. Then the total capital amount of the P&C insurer is given by $\rho(\sum_1^n X_i)$. Now

⁴In Chapter 1 the aggregate loss was denoted by S to stand for the "sum", but we use X throughout this chapter to minimize the use of different capital letters.

an *ex post*⁵ capital allocation follows. We denote the set of all lines (or the whole portfolio) by

$$\Omega = \{X_1, \dots, X_n\} \quad (6.7)$$

and define the allocated capital for line i by

$$\rho(X_i|\Omega), \quad (6.8)$$

where the use of Ω in the condition indicates that the allocated capital for line i has been allocated after the total capital has been computed based on Ω , or the whole portfolio. Similarly if we define a subportfolio $\mathcal{H} \subset \Omega$, $\rho(X_i|\mathcal{H})$ represents the allocated capital where the total capital has been computed based on lines in \mathcal{H} . Thus in general $\rho(X_i|\Omega)$ and $\rho(X_i|\mathcal{H})$ have different meanings and different values. Putting conditions inside the risk measure notation thus enables us to tell subtle differences in allocation process.

Even though we use the same letter ρ for both total capital and the allocated one because of conceptual similarity, they are not meant to be the same mathematical function and often there is no analytic resemblance between them. One equation however should always hold by definition, that is, for any subportfolio $\mathcal{H} \subset \Omega$,

$$\rho\left(\sum_{\mathcal{H}} X_i|\mathcal{H}\right) = \rho\left(\sum_{\mathcal{H}} X_i\right),$$

where

$$\sum_{\mathcal{H}} X_i \equiv \sum_{X_i \subset \mathcal{H}} X_i$$

Now we are ready for the axioms. Axioms below are adapted from Valdez and Chernih (2003), and are mathematically elaborated from the literature including Denault

⁵We assume that capital allocation always occurs after the total capital is determined.

(2001) and Hesselager and Anderson (2002). We have slightly modified the presentation in Valdez and Chernih (2003) for an easier exposition.

Definition 6.1 (Fair allocation axioms) *Suppose that a company has n business lines, and each line's loss is represented by X_i , $i = 1, \dots, n$. $\rho(\cdot)$ is the risk measure for capital and $\rho(\cdot)$ is the capital allocation rule as explained above. An allocation $\rho(\cdot)$ is said to be fair if the following four properties hold.*

Full allocation For $\Omega = \{X_1, \dots, X_n\}$,

$$\rho\left(\sum_{\Omega} X_i\right) = \sum_{\Omega} \rho(X_i|\Omega) \quad (6.9)$$

No undercut For any $\mathcal{H} \subset \Omega$,

$$\sum_{\mathcal{H}} \rho(X_i|\Omega) \leq \rho\left(\sum_{\mathcal{H}} X_i|\mathcal{H}\right) = \rho\left(\sum_{\mathcal{H}} X_i\right) \quad (6.10)$$

Symmetry *Select arbitrary two lines X_i, X_j ($i \neq j$) and create \mathcal{H} such that $\{X_i, X_j\} \subset \mathcal{H} \subset \Omega$ (but $\mathcal{H} \neq \Omega$). If $\rho(X_i|\mathcal{H}) = \rho(X_j|\mathcal{H})$ holds for every such set \mathcal{H} , we must have*

$$\rho(X_i|\Omega) = \rho(X_j|\Omega) \quad (6.11)$$

Consistency For any $\mathcal{H} \subset \Omega$,

$$\sum_{\mathcal{H}} \rho(X_i|\Omega) = \rho\left(\sum_{\mathcal{H}} X_i|\left\{\sum_{\mathcal{H}} X_i\right\} \cup \mathcal{H}^c\right), \quad (6.12)$$

where $\mathcal{H}^c = \Omega - \mathcal{H}$.

Let us briefly explain each axiom. The full allocation means that the sum of the allocated capital equals the capital of the sum of the risks. The no undercut is analogous to the subadditivity of a risk measure. It means that the capital allocated to a

group of lines, given the full lines Ω , is less than the capital allocated if the group is offered on a stand-alone basis. The symmetry and the consistency are a little more subtle to understand. The symmetry means that if two different lines have the same allocated capital amount given any subset lines $\mathcal{H} \subset \Omega$ ($\mathcal{H} \neq \Omega$), those two lines should have the same allocated capital given the full lines Ω . As Denault (2001) pointed out in the original paper, the symmetry property ensures that a portfolios allocation depends only on its contribution to risk within the firm, and nothing else. Finally the consistency states that the allocated capital amount is independent of the hierarchical structure of the company. For example, suppose that the company has a few divisions, and each division holds some individual lines under its supervision. The collection of all lines across all divisions makes Ω , of course. The consistency compares capital allocation at the division level with that at the individual line level. This axiom means that the allocated capital amount for one division, the right side of (6.12), should be the same as the sum of the allocated capital amounts of the each line in the division, the left side of (6.12).

Along with the developments of general allocation principles, there have been much effort to find specific allocation methods by academics and practitioners. Some methods are aligned with the proposed axioms and others fail to meet them all. The relative allocation

$$\rho(X_i|\Omega) = \frac{\rho(X_i)}{\rho(X_1) + \rho(X_2) + \dots + \rho(X_n)} \quad (6.13)$$

is a simple example. Valdez and Chernih (2003) shows that this allocation method satisfies the full allocation and the symmetry axioms, but fails the other two axioms.

They considered another example

$$\rho(X_i|\Omega) = \frac{Cov(X_i, X)}{Var(X)} \times \rho(X) \quad (6.14)$$

and proves that this method meets all fair allocation axioms. We remark however that the proof is partly incorrect and it actually fails the no undercut axiom⁶. The allocated capital in this example is directly linked to its correlation to the whole portfolio of the company. The third example is the famous CTE allocation, found in Overbeck (2000) and Panjer (2002):

$$\rho(X_i|\Omega) = E[X_i|X > Q_\alpha(X)] \quad (6.15)$$

Panjer (2002) reported that this method satisfies the first three axioms, but consistency was not in his list. However we note that it also satisfies the consistency axiom. The CTE allocation under the full portfolio Ω , given on the left side of (6.12), is based on the aggregate loss $\sum_1^n X_i = \sum_\Omega X_i$, and similarly the allocation under portfolio $\{\sum_{\mathcal{H}} X_i\} \cup \mathcal{H}^c$, the left side of (6.12), is based on $\sum_{\mathcal{H}} X_i + \sum_{\mathcal{H}^c} X_i$. Since we always have

$$\sum_{\Omega} X_i = \sum_{\mathcal{H}} X_i + \sum_{\mathcal{H}^c} X_i,$$

for any $\mathcal{H} \subset \Omega$, both sides measure the same aggregate loss. Finally, the first summation on the right side of (6.12) can be put in front due to additivity of the CTE, making the CTE allocation consistent. This argument is also relevant to the consistency of the allocation method to be introduced later.

For other allocation methods and discussion, readers are referred to, for example, Dhaene *et al.* (2003), Wang (2002), Tasche (1999) and Myers and Read Jr. (2001).

⁶In the proof of Theorem 1 (pp.522) the second inequality $\sum \sigma(X_i) \leq \sigma(\sum X_i)$ should be reversed. We also present a numerical example to support this finding in Section 6.6.

Goovaerts *et al.* (2005) argue that the allocation problem can largely take two different forms. The first approach is called the risk measure approach that uses the properties of the given risk measure. A simple example is the CTE allocation given in (6.15). By denoting $X = \sum_1^n X_i$, the allocation satisfies

$$E[X|X > Q_\alpha(X)] = \sum_{i=1}^k E[X_i|X > Q_\alpha(X)] \quad (6.16)$$

The second approach is based on the insolvency event. In this approach the residual risk defined by $(X - u)^+$, where u is the total capital amount, is to be somehow split onto each line. A unique aspect of this approach is that the allocation rule depends on the amount of the total capital $u = \rho(X)$, but is independent of analytic form of $\rho(\cdot)$. To obtain the allocated capital in this approach one solves the equation

$$\min E\left[\sum_{i=1}^n (X_i - u_i)^+\right], \quad (6.17)$$

with respect to (u_1, \dots, u_n) , the vector of the allocated capital, under constraint $\sum u_i = u = \rho(X)$. The second method is the preferred one in their paper and the solution (u_1^*, \dots, u_n^*) is given by

$$u_i^* = F_{X_i}^{-1}(1 - s),$$

where s is determined as $F_{X_i^c, X_2^c, \dots, X_n^c}(u) = 1 - s$, in which $F_{X_i^c, X_2^c, \dots, X_n^c}$ is the c.d.f. of the comonotonic random vector $(X_i^c, X_2^c, \dots, X_n^c)$ with the same marginal distribution functions as (X_i, X_2, \dots, X_n) .

In the next section we derive a new allocation method that is independent of analytic form of $\rho(\cdot)$ like the second approach but has some similarity to the first one as well.

	Shareholders get	Amount liable to policyholder
When solvent ($A \geq L$)	$A_i - L_i$	L_i
When insolvent ($A < L$)	0	$\frac{L_i}{L}A$

Table 6.4: Entitlements of shareholders and policy holders at time 1 for line i

6.4 A new allocation

The idea of including the insolvency exchange option in the balance sheet to determine the economic liability value is not new. In fair valuation framework this is translated to the issue of reflecting company's own credit risk, and this has been a controversial issue among regulators; including the option would effectively reduce liability and increase the capital in the financial statement; see Section 1 of Girard (2002) for a brief discussion and further references.

In the financial economics literature, there has been several papers focusing on the allocation of option D , defined in (6.2). Merton and Perold (1993) and Myers and Read Jr. (2001) fall into this category. Later Sherris (2006) derived a way to allocate D in an additive manner based on the equal priority of each line on insolvency. To briefly review the allocation idea of Sherris, we start with Table 6.4 which is similar to Table 6.2 but this time focused on i -th line. (No specification on how to obtain A_i , the asset allocated to line i , is given yet). Assume again that each policy holder ranks equally for the liquidating asset on insolvency, which is a standard procedure in practice. The economic liability of line i at time 1 then can be defined by

$$L_i - L_i \max\left(1 - \frac{A}{L}, 0\right) = L_i - L_i \left(1 - \frac{A}{L}\right)^+ = \begin{cases} L_i, & \text{if } A \geq L \\ \frac{L_i}{L}A, & \text{if } A < L, \end{cases} \quad (6.18)$$

as shown in Table 6.2. Sherris denotes $L_i(1 - \frac{A}{L})^+$, the option payoff to the i -th line in the event of insurer default, by D_i and uses this to allocate the solvency exchange option D . That is,

$$D = (L - A)^+ = \sum_{i=1}^k D_i, \quad (6.19)$$

where the equality is trivial, and the corresponding option prices are

$$V(D) = \sum_{i=1}^k V(D_i), \quad (6.20)$$

where $V(D)$ is given in (6.3) and

$$V(D_i) = e^{-r} E^Q[L_i(1 - \frac{A}{L})^+] \quad (6.21)$$

Observe that D_i is a function of L_i and A , but there is no need to specify A_i . This raises the important point of his paper that the allocation of the asset plays no role in either valuing or allocating the solvency exchange option; asset allocation can be arbitrary in the following sense, as quoted from Sherris (2006):

“There is no unique way to do this since the allocation of assets to line of business is an internal insurer allocation that will have no impact on the payoffs or risks of the insurer since assets are available to meet the losses of all lines of business.”

Recognizing however the need of asset allocation in management’s decision making, he considers two alternatives; making each line either have the same solvency ratio or have the same expected return on capital. The allocation of assets is equivalent to the allocation of the capital in this framework because the capital and surplus at time 1 is $A - L + D$ and so far both L and D have been allocated, but A has not. If

	Time 0	Time 1
Asset	$u + P$	$ue^{r_u} + Pe^{r_P}$
Liability	$V(L) - V(D)$	$L - D$
Capital and surplus	$u + P - V(L) + V(D)$	$ue^{r_u} + Pe^{r_P} - L + D$

Table 6.5: Economic balance sheet of an insurer with premium inflow

A_i is determined, the capital and surplus for i -th line would be $A_i - L_i + D_i$.

We now propose a new asset allocation method. Unlike other methods, this one explicitly accommodates the capital reduction due to the shareholders' limited liability. For this new method first we keep the same one period model and further split the asset into premium and capital. More specifically,

- A premium amount of P_i is collected at time 0 for each line in exchange for uncertain loss at time 1, with the total premium $P = \sum_{i=1}^n P_i$. The insurer invests P in securities that produce a random return rate of r_P over the period.
- The shareholders inject an initial capital amount of u at time 0 to safeguard the company from extreme liability outcomes where loss exceeds premium. It is assumed that the capital is invested at a random rate r_u .

We assume r_P and r_u to be random for a more general framework; the final result would be the same even if the rates are constant. Table 6.5 is the revision of Table 6.3 using these new variables. Note that in the table all variables except u can be allocated to each line directly with no difficulty.

The construction of the new allocation method starts with considering the follow-

ing trivial equality where $A = ue^{r_u} + Pe^{r_P}$.

$$(L - ue^{r_u} - Pe^{r_P})^+ = (L - A)^+ = \sum_1^k D_i, \quad (6.22)$$

where

$$D_i = L_i \left(1 - \frac{A}{L}\right)^+ = \begin{cases} 0, & \text{if } A \geq L \\ L_i - \frac{L_i}{L}A, & \text{if } A < L, \end{cases}$$

is the payment shortfall to the i -th line policy holders in the event of insurer's default, due to Sherris (2006). Next we take the expectation of equation (6.22) under the P measure. We use the P measure because capital allocation, like the risk measure, is designed to assess the true risk. Here the new method considers actual occurrence of insolvency and the physical behavior of the net loss under insolvency scenarios, rather than its fair price. Thus

$$E^P[(L - ue^{r_u} - Pe^{r_P})^+] = \sum_1^k E^P[D_i]$$

To solve this equation with respect to u , the right side is modified to

$$\sum_1^k E[D_i] = \sum_1^k E \left[L_i \left(1 - \frac{A}{L}\right)^+ \right] = Pr(L > A) \sum_1^k E \left[L_i \left(1 - \frac{A}{L}\right) | L > A \right]$$

and similarly for the left side,

$$E[(L - ue^{r_u} - Pe^{r_P})^+] = Pr(L > A) E[L_i - u_i e^{r_u} - P_i e^{r_P} | L > A],$$

noting that the two events $L - ue^{r_u} - Pe^{r_P} > 0$ and $L > A$ are equivalent in that both represents insolvency. Hence equation (6.22) reduces to

$$E[L - ue^{r_u} - Pe^{r_P} | L > A] = \sum_1^k E \left[L_i - \frac{L_i}{L}A | L > A \right],$$

for $i = 1, \dots, k$. Before proceeding further we note that on the left side the allocations of the total premium P and liability L to each line are trivial but the same task for capital u is not because capital is available to all lines in case of adverse outcomes. We choose here, among other methods, to allocate u in such a way that each term of both sides matches for each line. That is, for each i ,

$$E[L_i - u_i e^{r u} - P_i e^{r P} | L > A] = E \left[L_i - \frac{L_i}{L} A | L > A \right], \quad (6.23)$$

where u_i represents the allocated capital to the i -th line. Matching both sides for each i means that, in the event of insurer default, the allocated capital should make each line's average net loss, which is the left side, equal to the average payment shortfall to the line's policy holders, which is the right side. This equation is actually exploiting the fact that insolvency is the only situation where asset allocation is clearly defined through the equal priority for all lines, as shown in Table 6.4.

Finally we rearrange equation (6.23) to obtain the allocated capital

$$u_i = \frac{E \left[\frac{L_i}{L} A - P_i e^{r P} | L > A \right]}{E[e^{r u} | L > A]} \quad (6.24)$$

If the capital is invested in a risk free bond at r , which is a reasonable assumption where the liability is positively linked to r_P , the allocated capital is

$$u_i = e^{-r} E \left[\frac{L_i}{L} A - P_i e^{r P} | L > A \right] \quad (6.25)$$

In the next section we examine some interesting properties of this allocation method, along with its advantages.

6.5 Properties of the new allocation

Generally, for any insurance business, we expect the capital to be an increasing function of the liability loss and a decreasing function of premium collected, because the real risk lies in the loss exceeding the premium, or the net loss. The allocation method proposed here is aligned with this principle. From (6.24), when the line's loss share L_i/L increases, with L , A , and P_i unchanged, the allocated capital increases; the remaining lines will get less capital. Similarly the allocated capital decreases if P_i increases with L , A , and L_i kept fixed. This allocation is also consistent with the idea of different level of implied leverage by line; each line will have different expected rates of return depending on premium adequacy of that line. The impact of higher allocation is that the line will be required to earn more profit, to maintain the same rate of return as other lines, because the line's business carries more risk, either through an inadequate premium or a riskier liability.

Another feature of the proposed allocation method is that it is independent of how the total capital is computed and only requires the amount of the total capital u . The allocation is then based on insolvency scenarios, $L > A$, and insolvency scenarios alone; solvent scenarios make no contributions in allocating the capital. In fact the allocation (6.24) involves a type of conditional tail expectation, with a little different form than the usual CTE. The similarity to the CTE however turns out to provide several desirable technical advantages just as the CTE allocation does.

6.5.1 Fairness of the allocation

First of all, the allocation ‘adds up’, meaning that it satisfies the full allocation axiom because

$$\begin{aligned}
\sum_{i=1}^k u_i &= \frac{1}{E[e^{r_u}|L > A]} \sum_{i=1}^k E\left[\frac{L_i}{L}A - P_i e^{r_P}|L > A\right] \\
&= \frac{1}{E[e^{r_u}|L > A]} E\left[\frac{\sum_{i=1}^k L_i}{L}A - \sum_{i=1}^k P_i e^{r_P}|L > A\right] \\
&= \frac{1}{E[e^{r_u}|L > A]} E[A - P e^{r_P}|L > A] \\
&= \frac{1}{E[e^{r_u}|L > A]} E[ue^{r_u}|L > A] \\
&= \frac{u}{E[e^{r_u}|L > A]} E[e^{r_u}|L > A] \\
&= u
\end{aligned}$$

To prove symmetry, suppose that lines i and j have been arbitrarily picked and the assumption of the axiom is met for any $\mathcal{H} \supset \{X_i, X_j\}$. Then we should have

$$\frac{E\left[\frac{L_i}{L_{\mathcal{H}}}A_{\mathcal{H}} - P_i e^{r_P}|L_{\mathcal{H}} > A_{\mathcal{H}}\right]}{E[e^{r_u}|L_{\mathcal{H}} > A_{\mathcal{H}}]} = \frac{E\left[\frac{L_j}{L_{\mathcal{H}}}A_{\mathcal{H}} - P_j e^{r_P}|L_{\mathcal{H}} > A_{\mathcal{H}}\right]}{E[e^{r_u}|L_{\mathcal{H}} > A_{\mathcal{H}}]},$$

where $A_{\mathcal{H}}$ represents the aggregate asset of lines in set \mathcal{H} , and similarly for $L_{\mathcal{H}}$. This equation holds only if $L_i = L_j$ (in distribution) and $P_i = P_j$ whenever $L_{\mathcal{H}} > A_{\mathcal{H}}$. Because \mathcal{H} can be any strict subset of Ω , lines i and j must be identical in loss and premium for any situation. Therefore the allocated capital for both lines should be the same under the full portfolio Ω as well, because the proposed allocated capital produces the same amount as long as the loss and premium are identical. The proposed allocation method is also consistent and the argument is identical to that of

the CTE allocation discussed earlier.

However this method fails the no undercut axiom as shown by an example later in this chapter. We resume our discussion on this in Section 6.7 for a further investigation.

6.5.2 Further decomposition

Here we provide three alternative ways to interpret the allocated capital by further decomposing it into smaller components. These decompositions give the management and line managers more information about the structure of the allocated capital and could also promote a better communication among the various stakeholders. For simplicity we set $r_u = r$, the risk free rate, and substitute A with $ue^r + Pe^{rP}$.

Decomposition I

The first decomposition breaks the allocated capital into two pieces: the capital based on the gross-up loss and the premium adjustment. Under this decomposition the allocated capital (6.24) becomes

$$\begin{aligned} u_i &= e^{-r} E \left[\frac{L_i}{L} (ue^r + Pe^{rP}) - P_i e^{rP} | L > A \right] \\ &= e^{-r} E \left[\frac{L_i}{L} ue^r + e^{rP} \left(\frac{L_i}{L} P - P_i \right) | L > A \right] \\ &= u E \left[\frac{L_i}{L} | L > A \right] + e^{-r} E \left[e^{rP} \left(\frac{L_i}{L} P - P_i \right) | L > A \right] \end{aligned}$$

Here the first term is the capital allocation due to each line's gross up loss and is always positive for all lines. The second term is the adjustment due to each line's

premium collection and can be negative or positive depending on the relative premium adequacy of the line; a line with adequate premium will have a negative adjustment. The sum of the first term across the lines is equal to the total capital and the sum of the second term is zero.

Decomposition II

The second decomposition of the allocated capital is

$$\begin{aligned}
 u_i &= e^{-r} E \left[\frac{L_i}{L} (ue^r + Pe^{rP}) - P_i e^{rP} | L > A \right] \\
 &= e^{-r} E \left[\frac{L_i}{L} (ue^r + Pe^{rP} - L + L) - P_i e^{rP} | L > A \right] \\
 &= e^{-r} E \left[\frac{L_i}{L} (ue^r + Pe^{rP} - L) + (L_i - P_i e^{rP}) | L > A \right] \\
 &= e^{-r} E \left[\frac{L_i}{L} (A - L) + (L_i - P_i e^{rP}) | L > A \right] \\
 &= e^{-r} E \left[\frac{L_i}{L} (A - L) | L > A \right] + e^{-r} E \left[L_i - P_i e^{rP} | L > A \right]
 \end{aligned}$$

The first term, being always negative due to its condition $L > A$, represents the policyholder loss on insolvency, as noted from Table 6.4 because

$$e^{-r} E \left[\frac{L_i}{L} (A - L) | L > A \right] = e^{-r} E \left[\frac{L_i}{L} A - L_i | L > A \right]$$

Thus this term always reduces the allocated capital amount and reflects the limited liability of the shareholders. The second term represents the capital due to premium inadequacy of the line, again on insolvency. The premium, even though reasonable at time 0, can be inadequate at the end of period for two reasons: poor investment performance on the asset side and worse than expected loss on the liability side.

Employing this idea, we can further decompose the second term into three pieces.

$$e^{-r}E[L_i - P_i e^{rP} | L > A] = e^{-r} \left(\{E[L_i | L > A] - E[L_i]\} \right. \\ \left. + \{E[P_i e^{rP}] - E[P_i e^{rP} | L > A]\} \right. \\ \left. - \{E[P_i e^{rP}] - E[L_i]\} \right)$$

Putting these components together the allocated capital becomes

$$u_i = e^{-r} \left(\{E[L_i | L > A] - E[L_i]\} + \{E[P_i e^{rP}] - E[P_i e^{rP} | L > A]\} \right. \\ \left. - \{E[P_i e^{rP}] - E[L_i]\} - E\left[\frac{L_i}{L}(L - A) | L > A\right] \right) \quad (6.26)$$

and each of the four terms has a unique meaning as elaborated below.

1. $E[L_i | L > A] - E[L_i]$ represents the need for capital due to insolvency loss over expected loss, with insolvency loss computed by the average loss under insolvency scenarios
2. $E[P_i e^{rP}] - E[P_i e^{rP} | L > A]$ ⁷ represents the need for capital due to poor investment performance over expected investment income, with poor investment performance computed by the average cumulative premium under insolvency scenarios
3. $-\{E[P_i e^{rP}] - E[L_i]\}$ represents a reduction of the capital due to profit loading for the line on average; and
4. $-E\left[\frac{L_i}{L}(L - A) | L > A\right]$ represents a reduction of the capital due to the limited liability of shareholders.

⁷This reduces to $P_i(E[r_P] - E[r_P | L > A])$ when the premium payment is known, but it is possible that premium is random, if it is charged as percentage of the underlying fund, for example.

The last term, as mentioned before, represents a capital reduction due to the limited liability of shareholders in the event of insolvency, sum of which across all lines is denoted by as $V(D)$ in the time-0 capital and surplus account of Table 6.3 and 6.5. This is a unique feature of this allocation method that other methods do not have, and in this sense it is more aligned with economic reality. Inclusion of this component in the capital allocation explains how much each line's policy holders lose when the insurer goes bankrupt.

Decomposition III

In the third decomposition, we try to attribute risks to different stakeholders involved in the company operation, namely, the liability manager (LM), the investment manager (IM), and the corporate risk manager (RM). Here the LM and the IM are assumed to have no control on the dependency structure of the whole company, and the capital is assigned as such. The RM is the one who is responsible for all of the dependency issues on both liability and investment sides. This is analogous to separating a multivariate density into the marginals and the copula. The main advantage of this decomposition is that the management is able to see each manager's risk accountability. The allocated capital for line i has the following decomposition under this approach:

$$u_i = e^{-r} \left[LM_i + IM_i + RM_i - E\left[\frac{L_i}{L}(L - A) | L > A\right] \right] \quad (6.27)$$

The first three terms represent components under each manager's responsibility. The last term again represents a capital reduction due to the limited liability of shareholders and it is rather an output of the company's business than an input that managers should control and optimize.

LM_i : $E[L_i|L_i > \rho(L_i)] - P_i e^r$ represents the capital amount needed for line i 's liability manager on a stand alone basis. No uncertainty is assumed in the investment side. ρ can be any tail risk measure but the consistent risk measure with the total capital setting is recommended for conceptual agreement. For instance, if the total solvency capital is computed by the CTE at level α , the same applies to ρ . The liability manager would be encouraged to minimize this quantity. If necessary LM_i can be split into two pieces:

$$\{E[L_i|L_i > \rho(L_i)] - E[L_i]\} + \{E[L_i] - P_i e^r\},$$

where the second term is the profit loading.

IM_i : $P_i e^r - E[P_i e^{rP} | P e^{rP} < -\rho(-P e^{rP})]$ represents the capital amount for the investment manager on a stand alone basis. The nested negative signs in the condition of the second term reflects the fact that the asset side risk, unlike the liability side, is concentrated on cases where the earned rate is below than expected. Note that the condition is on the whole investment side because only one investment manager is assumed here; the manager is responsible for the investment of all lines, not just line i . The manager would be encouraged to minimize the sum of this quantity across all lines, or $\sum IM_i$. If different investment managers are hired for different lines, one can easily adapt the method by replacing the condition.

RM_i : $\{E[L_i|L > A] - E[L_i|L_i > \rho(L_i)]\} + \{E[P_i e^{rP} | P e^{rP} < -\rho(-P e^{rP})] - E[P_i e^{rP} | L > A]\}$ represents a reduction in the capital due to liability and investment diversification of the company. The sum of each term across all lines should be negative even though some can be positive indicating risk concentration. A

larger negative value of this quantity means a more efficient risk reduction for the line.

The function of the risk manager, in this framework, is to interact with the other managers by monitoring each RM_i and their sum $\sum RM_i$; the former provides an understanding of each lines dependency among lines and the latter gives the total amount of capital reduction due to diversification. Overall, the corporate risk manager should minimize $\sum RM_i$, maintaining a healthy tension between each RM_i .

6.6 Numerical example

In this section we provide a numerical example to apply the proposed allocation method. The example uses simplified life insurance products for illustration. The goal is to compare different allocation methods to see if they properly accommodate well-known profitability indices. Also we will examine the implication of including the solvency option in these measures.

Consider company A with two business lines. The liability of both lines is the guaranteed minimum maturity benefit of a segregated fund with 1 year contract terms. In particular, line 1's liability is defined by $(G_1 - F_t)^+$, ignoring mortality, where F_t is the fund value at maturity and guarantee level G_1 is set at 90% of the initial fund value, or $G_1 = 0.9F_0$. Line 2 has the identical underlying fund but different guarantee level at $G_2 = 1.0F_0$.

All contracts are one year and have the identical underlying fund. The fund value at any time is determined by

$$F_t = F_0 \frac{S_t(1 - m)^t}{S_0} \quad (6.28)$$

where m is the annual management charge and S_t is the underlying asset at time t . In our example the monthly management charge is set at 1% of the assets and the assets follow a geometric Brownian Motion with monthly parameters $\mu = 0.00947$, $\sigma = 0.04167$, derived from the TSE 300 data during 1956-2001. The company thus effectively consists of two put options written on the same underlying fund with different strike prices. We also consider a separate company B which is a mono-line insurer specializing in equity indexed annuity (EIA) products. For simplicity, we bring in a simplified version of the equity linked annuity, of which liability is defined by $(F_t - G_3)^+$ where $G_3 = 1.0F_0$; all other dynamics are kept the same as company A including management charge. So company B effectively has a call option.

Since we deal with the allocation issue, no dynamic risk management such as hedging is assumed. On the other side of the balance sheet, we define the premium by the management charge; the collected monthly premiums are assumed to be invested at risk free $5\%/12 = 0.0042\%$ monthly effective rate. By assuming the same management charge for all three lines we have the same premium dynamics. For the details of how these products work in practice, see Chapter 6 and 13 of Hardy (2003).

Since we adopt the geometric Brownian Motion dynamics, the Q measure is defined uniquely, which allows a fair valuation of the company. The complete market assumption however does not fit fully this model since the premium charged is not the market value of premium. Table 6.6 shows basic valuation results for both companies. Now suppose that these two companies merge to establish a multiline insurer called ANB. We can create a similar valuation table for ANB, which is presented in Table 6.7. There are two notable impacts of the merger. First the required capital decreases significantly, reflecting that the aggregate risk has been reduced due to diversification.

Company	Company A			Company B	
Line	Seg Fund 1	Seg Fund 2	Total	EIA	Total
Nominal liability $V(L_i)$	3.28	8.48	11.76	2.93	2.93
Premium $V(P_i)$	11.52	11.52	23.04	11.52	11.52
Nominal surplus $V(P_i) - V(L_i)$	8.24	3.04	11.28	8.59	8.59
Option price $V(D_i)$	0.191	0.274	0.465	0.033	0.033
Capital u @CTE 95%			19.12		18.47
Capital & surplus (incl. option)			30.87		27.09

Table 6.6: Valuations of company A and B

Company	Company ANB			
Line	Seg Fund 1	Seg Fund 2	EIA	Total
Nominal liability $V(L_i)$	3.28	8.48	2.93	14.69
Premium $V(P_i)$	11.52	11.52	11.52	34.56
Nominal surplus $V(P_i) - V(L_i)$	8.24	3.04	8.59	19.87
Option price $V(D_i)$	0.185	0.265	0.001	0.451
Capital u @CTE 95%				9.35
Capital & surplus (incl. option)				29.67

Table 6.7: Valuations of the merged company ANB

Alloc. method	Seg fund 1	Seg fund 2	EIA	Total capital
Covariance	2.93	4.69	1.72	9.35
CTE	3.62	12.26	-6.54	9.35
Proposed	4.79	12.45	-7.90	9.35

Table 6.8: Capital allocations to each line of company ANB

With a minimal difference in the capital and surplus account, before and after the merger, this would make the ROC of the merged company much higher than that of the original two companies. The second impact is the reduction in the solvency exchange option value for each line. This means that there is a benefit to the policy holders due to risk reduction of the aggregated risk. These impacts however cannot be guaranteed to occur for any merger in general.

Now we allocate the total capital u of ANB into its three lines using the covariance, the CTE, and the proposed allocation methods. For the first two methods the allocation is applied to the end of period and discounted at the risk free rate. Unlike the covariance allocation, the CTE and the proposed methods show that the EIA line is providing a natural hedge for the other lines by negative allocation. In fact the EIA is countermonotonic to the segregated fund lines. Negative capital can be problematic for ROC calculation in practice but properly indicates its risk position nonetheless. Table 6.9 shows detailed components of the allocated capital using three different decomposition methods developed in the previous section. In this example, the premiums collected are random based on the fund value at each month. In decomposition III, for instance, we would have the following observations by looking at each component:

- In an increasing order, line 1 has the least liability risk, following by line 2 and 3, on a stand alone basis. The premium loading also shows that line 3 has the smallest margin.
- The risk accountable to the investment manager is the same across all lines because the premiums have been identically collected and invested.
- The corporate risk manager enjoys the diversification benefit on both the liability and the investment sides. The liability side reduction indicates that line 3 contributes the greatest value toward diversification.
- The capital reduction component due to the limited liability of the shareholders indicates that the line 2 gets the most benefit. However, one can also say that line 2 is to be blamed most, then line 1 and line 3, when the company is insolvent. So this component gives an idea of each line's contribution to insolvency.

One can derive a similar decomposition for the CTE allocation, but the capital reduction component due to limited liability of the shareholders will not be available, and thus information on each line's risk attribution under insolvency scenarios will be hardly known. Another difference is that the proposed method fails on the no undercut axiom and we deal with this problem in the next section.

6.7 No undercut axiom

In this section we show an example where the proposed allocation does not satisfy the no undercut axiom. However violating this property may not be as bad as one might think. Let us go back to the example given in the previous section and suppose that company A wants to add another line of segregated fund product instead of merging

Type	Attribution	Corresponding formula	Seg fund 1	Seg fund 2	EIA
I	Loss-based	$uE[\frac{L_i}{L} L > A]$	3.56	5.37	0.42
	Prem. adjust.	$e^{-r}E[e^{rP}(\frac{L_i}{L}P - P_i) L > A]$	1.23	7.08	-8.31
II	Excess liab. loss	$e^{-r}\{E[L_i L > A] - E[L_i]\}$	16.39	21.81	-2.96
	Excess inv. loss	$e^{-r}\{E[P_i e^{rP}] - E[P_i e^{rP} L > A]\}$	1.58	1.58	1.58
	Prem. loading	$-e^{-r}\{E[P_i e^{rP}] - E[L_i]\}$	-10.18	-6.60	-6.26
	Limited liab.	$-e^{-r}E[\frac{L_i}{L}(L - A) L > A]$	-3.00	-4.34	-0.25
III	LM(loss)	$e^{-r}\{E[L_i L_i > \rho(L_i)] - E[L_i]\}$	17.05	22.89	33.59
	LM(Prem. load.)	$-e^{-r}\{E[P_i e^r] - E[L_i]\}$	-10.18	-6.60	-6.26
	IM	$e^{-r}\{E[P_i e^r] - E[P_i e^{rP} Pe^{rP} < -\rho(-Pe^{rP})]\}$	2.03	2.03	2.03
	RM(liab. divers.)	$e^{-r}\{E[L_i L > A] - E[L_i L_i > \rho(L_i)]\}$	-0.66	-1.08	-36.55
	RM(inv. divers.)	$e^{-r}\{E[P_i e^{rP} Pe^{rP} < -\rho(-Pe^{rP})] - E[P_i e^{rP} L > A]\}$	-0.44	-0.44	-0.44
	Limited liab.	$-e^{-r}E[\frac{L_i}{L}(L - A) L > A]$	-3.00	-4.34	-0.25
All	Total allocated capital for each line (same for all three types)		4.79	12.45	-7.90

Table 6.9: Different decompositions of the allocated capital (ρ is set at CTE 95%)

Alloc. method	Seg fund 1	Seg fund 2	Total capital
Covariance	6.87	12.25	19.12
CTE	4.85	14.27	19.12
Proposed	5.54	13.58	19.12

Alloc. method	Seg fund 1	Seg fund 2	Seg fund 3	Total capital
Covariance	7.86	14.76	20.19	42.81
CTE	4.85	14.27	23.69	42.81
Proposed	6.03	14.27	22.51	42.81

Table 6.10: Capital allocations of company A, before and after adding Seg fund 3.

with company B. The added line shares the same dynamic as the existing two lines, but has a different guarantee level set at $G_3 = 1.1F_0$; the new line is comonotonic to the existing lines. The values are compared in Table 6.10; note that the capital amount for Seg fund 3 is 23.69 on a stand alone basis at CTE 95%. The covariance and the proposed methods apparently violate the no undercut axiom by showing increased allocated capital for each line after adding the third segregated fund line.

It would be appropriate to discuss the rationale behind the no undercut axiom at this point. The motivation of the axiom is that if one line has more capital assigned by the corporate management than it would have on a stand alone basis, the line manager would rather leave the company and go solo, because there is no point staying in a team if the required capital increases by doing that. The axiom therefore implicitly carries the notion that the assigned capital is owned by line manager.

We believe that this rationale is weak. In particular, an attempt to link the allocated capital to the line manager's ownership is not aligned with the economic reality. The decision of whether a line is staying in the company or not is made by the shareholders, not by the line manager because the shareholders are the owner of the capital and the company. For the shareholders the decision is solely based on the total capital value change after adding (dropping) a line; as long as the total capital measure is subadditive no further restriction on allocation is needed in terms of internal reallocation of the capital from the shareholders point of view. For the line managers the allocated capital is a phantom fund, and no line manager can actually claim the right on the allocated capital; they are hired by the shareholders and work for them.

Let us take a closer look at what has happened. First there is no capital reduction for a company as a whole through diversification for this merger since the newly added line is comonotonic - or perfectly dependent - to the existing lines. Actually the required capital, set at CTE 95%, of the merged company 42.81 equals the sum of two capitals before the merger, which are 19.12 and 23.69 respectively; this indicates that the risk concentration⁸, rather than diversification, has happened here. Upon this event, for the CTE method the allocated capitals of line 1 and 2 are not affected by adding the 3rd line, and line 3 holds the same capital after joining the company. On the other hand the covariance and the proposed method reallocates the total capital in such a way that the existing lines 1 and 2 are loaded more and the new third line benefits from capital reduction, indicating that line 3 distributes some of its risk to

⁸Note that due to the coherency of the CTE measure the total capital does not exceed the sum of each capital requirement.

existing lines due to its comonotonicity. We know that all lines benefit from capital reduction in most mergers as long as the merger involves risk reduction, or positive diversification. Here the question is what should happen if there is no capital reduction on merger. If the answer is that each existing line's capital should not change in this case, then it effectively means that, from the no undercut perspective, adding a comonotonic line is equivalent to adding nothing to the portfolio, and line managers should not be affected at all. If the answer is that each line's capital should increase to reflect the increase of the company's risk concentration and each line managers should participate in the total risk increase, it means that the no undercut axiom can be dropped.

Even though we do not have a definite answer to this, this question seems to lead us to the fundamental role of capital allocation. What do we expect from allocation? Where can we use it? Despite several motivations noted in Section 1.6, the capital allocation methodology has mainly been developed based on a list of axioms rather than on motivation, and little has been researched on how the given capital allocation can be used in light of its motivations. The main reason for this gap is the fact that it is not easy to translate the economic motivations into equivalent mathematical conditions; it may be that satisfying all the motivations would be too constraining to give any reasonable allocation method. There is evidence that many available allocation methods are rather weakly connected to the original motivation. For example, we find negative amount of capital is allowed in the CTE allocation, even though this is not useful in assessing the profitability of the line or line manager. Even if allocated capital is positive, using it directly for manager's compensation cannot be justified because interdependency of the allocation necessarily awards or penalizes

line managers based on the line's position within the company, rather than their own performance.

6.8 Concluding remarks

In this chapter we reviewed the valuation of a company in the market context using the Q measure and introduced a new capital allocation method that is aligned with the notion of limited liability of the shareholders under the P measure. The proposed method can further decompose the allocated capital, so that each stakeholder can have a clearer understanding of their contribution. Since the proposed method is based on the P measure, it can be applied to incomplete markets, where most insurance businesses belong. We also challenged one of the widely accepted allocation axioms, namely the no undercut axiom, and argued that this axiom may not be too convincing in some situations, in particular, when the risks are comonotonic. As Sherris (2006) says, if “allocation is just internal matter”, the no undercut axiom is merely a property that an allocation method may or may not satisfy. Many practitioners use the capital allocation to get various insights from it, and it does give valuable information on each line's risk position among the company. But it is also not clear what we can obtain from a general capital allocation algorithm.

Chapter 7

Discussion and future work

7.1 Summary and discussion

Throughout this thesis different topics in the actuarial risk management have been explored. In Chapter 2 the conditional tail expectation (CTE) and its higher order moments were derived for the exponential family (EF) distribution class. Later in the chapter the same quantity was provided for other distributions which are transformed from the EF class. Since the EF class and its transformed distributions cover many known distributions, we are able to analytically compute the conditional tail moments for a wide range of distributions, without relying on simulation. Since the conditional tail moments take unique forms for each distribution, it will be interesting to use this technique to fit a given loss data. Therefore in the next section we present an example for this idea.

Turning to the nonparametric approach, Chapter 3 and 4 make one package in estimating tail risk measures; in particular, the bias and the variance, respectively.

Focusing mainly on the CTE we found that the exact bootstrap (EB) technique provides a reasonable bias estimate for small i.i.d. samples based on both theoretical findings and numerical examples. In Chapter 4 the influence function of the bootstrapped quantile is derived. Even though the focus is on the CTE the theoretical result is extended to the whole L-estimator class, making the result applicable to many other problems.

On the theoretical side, the findings in these two chapters help us understand the implications of double (or further iterated) bootstrapping of the L-estimator class. For example the double bootstrapping is well-known for its heavy computation, but in the L-estimator class it simply is expressed by

$$E[E(\mathbf{c}'\mathbf{X}_{:n}|\hat{F}^*)|\hat{F}] = E[\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n}^*|\hat{F}] = \mathbf{c}'(\mathbf{w}^2)'\mathbf{X}_{:n}$$

and can be used for further bias correction, because the bias estimate B given in (3.13) is itself biased. So the adjusted bias is shown to be

$$B^{adj} = B - (\mathbf{c}'(\mathbf{w}^2)'\mathbf{X}_{:n} - 2\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} + \mathbf{c}'\mathbf{X}_{:n}) \quad (7.1)$$

$$= 3\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} - 2\mathbf{c}'\mathbf{X}_{:n} - \mathbf{c}'(\mathbf{w}^2)'\mathbf{X}_{:n} \quad (7.2)$$

The variance of the (adjusted) bias-corrected estimate is expected to be higher than the other estimators and a further research will be needed. Another interesting aspect is to examine the bootstrap of the L-estimator from a Markovian perspective. There already is a rich literature connecting the doubly stochastic matrix with the Markov chain. We conjecture that the bootstrap of the order statistics could be translated into the state change in the Markov chain, with each state representing the order of an observation.

The findings in these chapter can also be applied to practical problems. For ex-

ample, it can be used in the variance of the median, and in estimating the quantile variance without estimating the density value. It will be interesting to compare the proposed method to the ordinary bootstrap in estimating variance of a quantity that lies in the middle of the distribution, rather than the tail side. In Chapter 5, to justify our motivation of assuming the small i.i.d. sample and to demonstrate how to estimate the CTE in a practical context, we brought in a industry-scaled actuarial data and applied the algorithm developed in Chapter 3 and 4.

A relatively new topic in the actuarial literature, capital allocation problem is investigated in Chapter 6. Inspired by the idea of the allocation of the solvency exchange option under the Q-measure by Sherris (2006), a new capital allocation method is proposed under the P-measure. This method satisfies most fair allocation axioms and provides further decomposition, which would be useful for financial stakeholders to understand the structure of the allocated capital. At the end of Chapter 6 the no undercut axiom was questioned. There are two possible future extensions for the proposed method. First we can extend it to a multi-period setting. By doing this we should consider the release of current liability over time. Secondly we can develop a new metric based on the allocated capital decompositions. If properly done, this would provide a user friendly information sheet that describes a company's risk position, not just capital structure. For example, actuaries can translate a specified copula into the tabulated values, which could be a useful addition to the executive report for the management.

7.2 Future ideas

In this section we briefly describe some specific ideas for future research emerging from the topics discussed in the thesis. Some of them have already been elaborated to produce preliminary results.

7.2.1 CTE Comparison: parametric vs. nonparametric

The examples considered in Chapter 3 and 4 do not have analytic forms for the CTE. It would be interesting to see apply these techniques to the EF class where the CTE is analytically available from Chapter 2. For a given sample, we can estimate the MLEs (or moment based estimators) for most EF distributions and therefore the CTE becomes a function of estimated parameters. In context of the nonparametric CTE estimate, such as empirical or EB, there will be sampling error involved in the parametric approach even when the model is correct. Generally we would expect that the parametric approach to be better than the nonparametric one, but we do not know its relative efficiency for the tail risk measures. Also if the model is misspecified in the parametric approach, it is possible that the nonparametric CTE could outperform the parametric one.

7.2.2 The Shifted Power Transform

Kernel density (KD) estimation is another useful nonparametric method to estimate the CTE or other tail risk measures. The KD is defined by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (7.3)$$

where X_i is i -th observation of the sample, and KD distribution function is

$$\hat{F}_h(t) = \int_{-\infty}^t \hat{f}_h(x) dx. \quad (7.4)$$

Basically this technique tries to mimic the unknown density function with a given data. The kernel K is usually assumed to be the normal (or gaussian) for its easy convolution. Problems with using standard symmetric kernels for the nonparametric density estimation of these data are two-fold; it violates the domain boundary by placing some weight in negative values. Also one global bandwidth cannot represent ideal smoothing in both left and right tails of skewed data. Wand *et al.* (1991) proposed to apply the ordinary kernel after transforming the data which is skewed or (and) positively distributed, which is the case for many actuarial data. In particular, they proposed the shifted power (SP) transformation with two parameters, given by

$$\tilde{g}_\lambda(x) = \tilde{g}_{\lambda_1, \lambda_2}(x) = \begin{cases} \text{sign}(\lambda_2)(x + \lambda_1)^{\lambda_2}, & \lambda_2 \neq 0 \\ \ln(x + \lambda_1), & \lambda_2 = 0 \end{cases}$$

Then Y , the transformed data, is defined as

$$Y = (\sigma_X / \sigma_{\tilde{Y}}) \tilde{Y} = g_{\lambda_1, \lambda_2}(X) \quad (7.5)$$

This transform basically turns the given skewed data into a normal-like shaped one (in unimodal case). Then the kernel estimate of Y becomes

$$\hat{f}_{Y,h}(t; \lambda_1, \lambda_2) = n^{-1} \sum_{i=1}^n K_h(t - Y_i) \quad (7.6)$$

where K_h is the normal kernel in our case. Bolance *et al.* (2003) showed this transformation performs better than non-transformed kernel method, using some actuarial data. We believe that the same would be true in estimating tail risk measures as

well, such as the VaR and the CTE. Moreover it can estimate beyond the maximum data, where the bootstrap can do nothing. It will be interesting to compare these two approaches for some actuarial models.

7.2.3 Bootstrap after transformation

Using the above SP transform, we obtain a normal-like data, as long as it is unimodal. An alternative to the KD technique, after the transform, is to apply the bootstrap to obtain the confidence interval. In the normal case the $CTE_\alpha(X)$ is given by

$$\mu + \sigma h_Z(z_\alpha),$$

where h_Z is the hazard function of the standard normal and z_α is its α quantile. One of the standard bootstrap methods is to use the pivotal quantity to get the confidence interval. Since the CTE of the normal has two pivotal quantities μ and σ , its standardized one, $h_Z(Q_\alpha)$, is independent of the parameters. Thus the confidence interval can be obtained in a simpler and easier manner; see Davison and Hinkley (1997) or Casella and Berger (2003) for example. The computed interval then can be back-transformed to give the required interval in terms of the original data. This technique is related to the confidence interval of the CTE than its point estimate.

7.2.4 A semi-parametric method for CTE estimation

In estimating tail risk measures it is possible to combine the bootstrap with other parametric models. Lee (1994) proposed the use of an estimator of the form

$$\epsilon t(F_n) + (1 - \epsilon)t(F_\xi)$$

in the univariate setting, where F_n is the empirical distribution function (or the bootstrap sample), F_ξ is the estimated parametric d.f., and ϵ is the weight constant in

$(0, 1)$, derived from data. The author argues that this would asymptotically produce the optimal MSE under some conditions.

Based on this argument we can estimate F_ξ by the transformed kernel introduced in the previous section, making the combined estimator totally nonparametric. This could be an alternative way of estimating the VaR or the CTE. It is also possible to extend this to multivariate setting as well.

7.2.5 Fitting the GPD for Tail Region

This idea is inspired from the findings of Chapter 2. In Extreme Value Theory the Generalized Pareto Distribution (GPD) is often used to fit the tail of a sample, under some distributional assumptions. That is, the k largest values of the sample can be assumed to follow the GPD. The starting point of the tail, called the threshold, or the choice of k , is not easy to determine and it has been a problem of interest to actuaries. From Chapter 2, the distribution of GPD is defined as

$$F_Y(y) = 1 - \left(1 + \frac{\xi}{\beta}y\right)^{-1/\xi},$$

where the domain is $y \geq 0$ if $\xi \geq 0$, or $0 \leq w \leq -1/\xi$ if $\xi < 0$. In that chapter we derived its CTE:

$$CTE_\alpha(Y) = \frac{\beta}{\xi} \left(\frac{(1-\alpha)^{-\xi}}{1-\xi} - 1 \right), \quad \xi < 1.$$

The choice of k affects the value of estimated parameters and there is no universally agreed method to find the optimal k (thus ξ and β as well) so far. Typically one should determine k first, and then estimate MLE's of ξ and β for a fixed k .

We can tackle this problem differently using the relation between the CTE and the conditional tail variance (CTVar). Again from Chapter 2 the CTVar is given by

$$CTVar_{\alpha}(Y) = \frac{\beta^2(1-\alpha)^{-2\xi}}{(1-2\xi)(1-\xi)^2}, \quad \xi < 1/2.$$

Equating the CTVar with the CTE then gives, after some algebra,

$$CTE_{\alpha}(Y) = \frac{\sqrt{(1-2\xi)CTVar_{\alpha}(Y)} + \beta}{\xi},$$

for any α , $0 < \alpha < 1$. This indicates that the (\sqrt{CTVar}, CTE) forms a straight line with slope $\sqrt{1-2\xi}/\xi$ and intercept at β/ξ , for the largest k order statistics. We can investigate the following issues:

1. Graphical determination of k .
2. Estimating ξ and β using the least square estimation, rather than the MLE, due to its similarity to the linear regression problem
3. Advantages using the relation of CTE and CTVar over using just CTE, in parameter estimation.

7.2.6 Allocation via Bootstrap

This idea is related to the capital allocation. We consider the bivariate case for the purpose of this discussion but higher dimensional cases can be handled in the same fashion with no additional requirements. Suppose that the company has only two business lines or risk components, Y and Z , with $S = Y + Z$. The CTE allocation states

$$E(S|S > F_S^{-1}(\alpha)) = E(Y|S > F_S^{-1}(\alpha)) + E(Z|S > F_S^{-1}(\alpha)), \quad (7.7)$$

where $S = Y + Z$. Provided that there is an i.i.d. random sample with size n , from this company's past records, $\mathbf{X}_1 = (Y_1, Z_1), \dots, \mathbf{X}_n = (Y_n, Z_n)$, the sample estimate of the allocation is computed as follows, using a formal matrix notation for later developments.

We first create $S_i = Y_i + Z_i$, $i = 1, \dots, n$, and sort them in an ascending order so that $S_{(1)} \leq S_{(2)} \leq \dots \leq S_{(n)}$ and let $\mathbf{S}_{:n} = (S_{(1)}, \dots, S_{(n)})'$, $n \times 1$ vector containing the ordered aggregate losses. Next we order the given sample (Y_i, Z_i) , $1 \leq i \leq n$, based on the order of S_i . This ordered sample makes a $n \times 2$ matrix whose i -th row is denoted by $(Y_{(i)}^S, Z_{(i)}^S)$, with $Y_{(i)}^S + Z_{(i)}^S = S_{(i)}$. The superscript emphasizes the criterion of the ordering, which is the aggregate loss S . We can call this matrix $\mathbf{X}_{:n}^S$ in the spirit of this naming convention. If we denote the first column of $\mathbf{X}_{:n}^S$ by $\mathbf{Y}_{:n}^S = (Y_{(1)}^S, \dots, Y_{(n)}^S)'$ and the second by $\mathbf{Z}_{:n}^S = (Z_{(1)}^S, \dots, Z_{(n)}^S)'$, we have $\mathbf{X}_{:n}^S = [\mathbf{Y}_{:n}^S \quad \mathbf{Z}_{:n}^S]$.

Using this notation the sample version of (7.7) becomes

$$\frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^n S_{(i)} = \frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^n Y_{(i)}^S + \frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^n Z_{(i)}^S \quad (7.8)$$

The sum of two terms of the right side is equal to the aggregate capital since $S_{(i)} = Y_{(i)}^S + Z_{(i)}^S$ for each i , satisfying the full allocation. This equation can be expressed in matrix form as

$$\mathbf{c}'\mathbf{S}_{:n} = \mathbf{c}'\mathbf{Y}_{:n}^S + \mathbf{c}'\mathbf{Z}_{:n}^S, \quad (7.9)$$

where $\mathbf{c} = \frac{1}{n(1-\alpha)}(0, \dots, 0, 1, \dots, 1)'$ with first $n\alpha$ elements being zeros. Note that both (7.8) and (7.9) are equivalent to the sample version of (7.7)

Turning to the bootstrap technique, we can create a bootstrap sample (or a resample) of size n from the original sample $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ with replacement. The created resample, $\{\mathbf{X}_1^*, \dots, \mathbf{X}_n^*\}$, is then used to compute the total and allocated capitals,

just as in (7.9):

$$\mathbf{c}'\mathbf{S}_{:\mathbf{n}}^* = \mathbf{c}'\mathbf{Y}_{:\mathbf{n}}^{\mathbf{S}^*} + \mathbf{c}'\mathbf{Z}_{:\mathbf{n}}^{\mathbf{S}^*}.$$

Repeating this bootstrap simulation R times we ultimately get the bootstrap estimate of the total and allocated capitals by averaging the bootstrapped results over the R repetitions. In fact the exact bootstrap result can be applied to this allocation problem.

Lemma 7.1 *The exact bootstrap estimate of $Y_{(r)}^{\mathbf{S}}$, $1 \leq r \leq n$, is given by*

$$E(Y_{(r)}^{\mathbf{S}}|\hat{F}) = \sum_{j=1}^n w_{j(r)} Y_{(j)}^{\mathbf{S}},$$

where

$$w_{j(r)} = \frac{\int_{(j-1)/n}^{j/n} t^{r-1}(1-t)^{n-r} dt}{B(r, n-r+1)} \quad \text{and} \quad B(a, b) = \int_0^1 t^{a-1}(1-x)^{b-1} dt.$$

Corollary 7.2 *Using the notation given above the bootstrapped CTE capital allocation for two risk components is given by*

$$\mathbf{c}'\mathbf{w}'\mathbf{S}_{:\mathbf{n}} = \mathbf{c}'\mathbf{w}'\mathbf{Y}_{:\mathbf{n}}^{\mathbf{S}} + \mathbf{c}'\mathbf{w}'\mathbf{Z}_{:\mathbf{n}}^{\mathbf{S}}, \quad (7.10)$$

where each term of the right side represents each line's allocated capital.

Equipped with this tool one can easily estimate the bias and variance of the allocated capital for each business line. For example consider line Y . Since there exists $\tilde{\mathbf{c}} = (\tilde{c}_1, \dots, \tilde{c}_n)'$ such that $\mathbf{c}'\mathbf{Y}_{:\mathbf{n}}^{\mathbf{S}} = \tilde{\mathbf{c}}'\mathbf{Y}_{:\mathbf{n}}$ by rearranging the elements, the bias estimate is

$$\tilde{\mathbf{c}}'(\mathbf{w}' - \mathbf{I})\mathbf{Y}_{:\mathbf{n}}$$

and the variance estimate is

$$\tilde{\mathbf{c}}'(\mathbf{w}' - \mathbf{I})\Sigma(\mathbf{Y}_{:\mathbf{n}})(\mathbf{w} - \mathbf{I})\tilde{\mathbf{c}},$$

where $\Sigma(\mathbf{Y}_{\cdot n})$ is the bootstrap estimate of the covariance matrix with each element $Cov(Y_{(i)}, Y_{(j)})$.

Bibliography

Academy Joint RBC Task Force (2002). Comparison of the NAIC life, P&C and health RBC formulas. Report, American Academy of Actuaries.

http://www.actuary.org/pdf/finreport/jrbc_12feb02.pdf (accessed February 8, 2007).

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measure of risk. *Mathematical Finance*, pages 203–228.

Artzner, P., Delbaen, F., Eber, J. M., Heath, D., and Ku, H. (November 15, 2006). Coherent multiperiod risk adjusted values and Bellman’s principle. *Annals of Operations Research (online version)*.

<http://www.springerlink.com/content/1572-9338/> (accessed in February 8, 2007).

Atkinson, D. B. and Dallas, J. W. (2000). *Life insurance products and finance*. The Society of Actuaries, Schumburg, IL.

Bao, Y. and Ullah, A. (2004). Bias of a Value-at-Risk estimator. *Finance Research Letters*, **1**, 241–249.

- Berkowitz, J. and O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks? *Journal of Finance*, **57**, 1093–1111.
- Bolance, C., Guillen, M., and Nielson, J. O. (2003). Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, **32**(1), 19–36.
- Brown, L. (1986). *Fundamentals of statistical exponential families*, volume 9 of *Lecture note - Monograph series*. IMS.
- Bühlmann, H. (1980). An economic premium principle. *ASTIN Bulletin*, **11**, 52–60.
- Butsic, R. P. (1993). Report on covariance method for property-casualty risk-based capital. Technical report, Actuarial Advisory Committee to the NAIC P&C Risk-Based Capital Working Group.
<http://www.casact.net/pubs/forum/93sforum/93sf173.pdf> (accessed February 8, 2007).
- Butsic, R. P. (1994). Solvency measurement for property-liability risk-based capital applications. *Journal of Risk and Insurance*, **61**(4).
- Cai, J. and Li, H. (2005). Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, **42**(3), 810–825.
- Cairns, A. J. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, **27**, 313–330.
- Casella, G. and Berger, R. L. (2003). *Statistical inference*. Duxbury, Pacific Grove, CA, 2nd edition.
- Christiansen, S. L. (1998). Representative interest rate scenarios. *North American Actuarial Journal*, **2**(3).

- Chueh, Y. C. (2002). Efficient stochastic modeling for large and consolidated insurance business. *North American Actuarial Journal*, **6**(3).
- CIA Task Force (2002). Report of the CIA task force on segregated fund investment guarantees. Report, Canadian Institute of Actuaries, Ottawa, ON.
- David, H. A. (1981). *Order statistics*. John Wiley & Sons, New York, 2nd edition.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press, New York.
- Denault, M. (2001). Coherent risk allocation of risk capital. Working paper, Ecole des H.E.C. Montreal.
- Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2002a). The concept of comonotonicity in actuarial science and finance: Applications. *Insurance: Mathematics and Economics*, **31**(2), 133–161.
- Dhaene, J., Denuit, M., Goovaerts, M., Kaas, R., and Vyncke, D. (2002b). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31**(1), 3–33.
- Dhaene, J., Goovaerts, M. J., and Kaas, R. (2003). Economic capital allocation derived from risk measure. *North American Actuarial Journal*, **7**(2), 44–59.
- Dhaene, J., Vanduffel, S., Tang, Q., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2006). Risk measures and comonotonicity: A review. *Stochastic Models*, **22**(4), 573–606.
- Dowd, K. (2001). Estimating VaR with order statistics. *Journal of Derivatives*, **8**(3).

- Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall, New York.
- Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modelling extremal events*. Springer, New York.
- Embrechts, P., Lindskog, F., and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In S. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. Elsevier.
- Feldblum, S. (1996). NAIC P&C insurance company risk-based capital requirements. *Proceedings of the Casualty Actuarial Society*, **LXXXIII**(159).
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, (2), 1–25.
- Furman, E. and Landsman, Z. (2005). Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics*, **37**(3), 635–649.
- Girard, L. N. (2002). An approach to fair valuation of insurance liabilities using the firm’s cost of capital. *North American Actuarial Journal*, **6**(2), 18–46.
- Glasserman, P., Heidelberger, P., and Shahabuddin, P. (2000). Variance reduction techniques for estimating Value-at-Risk. *Management Science*, **46**(10), 1349–1364.
- Goovaerts, M. J., den Borre, E. V., and Laeven, R. J. A. (2005). Managing economic and virtual economic capital within financial conglomerates. *North American Actuarial Journal*, **9**(3), 77–89.

- Gourieroux, C., Laurent, J., and Scaillet, O. (2000). Sensitivity analysis of value at risk. *Journal of Empirical Finance*, **7**, 225–245.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. Springer-Verlag, New York.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Function*. John Wiley & Sons, New York.
- Hardy, M. R. (2002). Bayesian risk management for equity-linked insurance. *Scandinavian Actuarial Journal*, **3**, 185–211.
- Hardy, M. R. (2003). *Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance*. John Wiley & Sons, New York.
- Hardy, M. R. and Wirch, J. L. (2004). The iterated CTE— a dynamic risk measure. *North American Actuarial Journal*, **8**(4), 62–75.
- Harrell, F. E. and Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika*, **69**(3), 635–640.
- Hesselager, O. and Anderson, U. (2002). Risk sharing and capital allocation. Working paper, Tryg Insurance, Ballerup, Denmark.
- Hutson, A. D. and Ernst, M. D. (2000). The exact bootstrap mean and variance of an L-estimator. *Journal of Royal Statistical Society: Series B*, **62**, 89–94.
- Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, **50**(4), 361–365.

- Inui, K., Kijima, M., and Kitano, A. (2005). VaR is subject to a significant positive bias. *Statistics & Probability Letters*, **72**, 299–311.
- Jeske, D. R. and Sampath, A. (2003). A real example that illustrates interesting properties of bootstrap bias correction. *The American Statistician*, **57**(1), 62–65.
- Joe, H. (1997). *Multivariate models and dependent concepts*. Number 73 in Monographs on Statistics and applied probability. Chapman & Hall, a CRC company.
- Johnson, N. L., Kotz, S., and Kemp, A. W. (1992). *Univariate Discrete Distribution*. John Wiley & Sons, New York, 2nd edition.
- Jones, B. L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, **7**(4).
- Jorion, P. (1997). *Value at Risk: the new benchmark for controlling market risk*. Irwin, Chicago.
- Kaas, R., Van Heerwaarden, A. E., and Goovaerts, M. J. (1994). *Ordering of Actuarial Risks*. Caire Education Series, Amsterdam.
- Kalkbrenner, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, **15**(3), 425–437.
- Klugman, S., Panjer, H., and Willmot, G. (1998). *Loss Models*. John Wiley, New York.
- KPMG (2002). Study into the methodologies to assess the overall financial position of an insurance undertaking from the perspective of prudential supervision. Report, KPMG.

- <http://www.us.kpmg.com/microsite/FSLibraryDotCom/insurance.aspx> (accessed February 8, 2007).
- Landsman, Z. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, **7**(4).
- Landsman, Z. and Valdez, E. A. (2005). Tail conditional expectations for exponential dispersion models. *ASTIN Bulletin*, **35**(1), 189–209.
- Lee, S. M. S. (1994). Optimal choice between parametric and non-parametric bootstrap estimates. *Math. Proc. Cambridge Philos. Soc.*, **115**, 335–363.
- Lehmann, E. L. (1983). *Theory of point estimation*. John Wiley & Sons, New York.
- Life Capital Adequacy Subcommittee (2005). C3 phase II risk-based capital for variable annuities: Pre-packaged scenarios. Report, American Academy of Actuaries. http://www.actuary.org/pdf/life/c3supp_march05.pdf (accessed February 8, 2007).
- Life Capital Adequacy Subcommittee to the NAIC Capital Adequacy Task Force (2005). Recommended approach for setting regulatory risk-based capital requirements for variable annuities and similar products. Report, American Academy of Actuaries, Boston, MA.
- Manistre, B. J. and Hancock, G. H. (2005). Variance of the CTE estimator. *North American Actuarial Journal*, **9**(2), 129–156.
- Mausser, H. (2001). Calculating quantile-based risk analytics with L-estimators. *Algo Research Quarterly*, **4**(4), 33–47.

- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press, New Jersey.
- Merton, R. C. and Perold, A. F. (1993). Theory of risk capital in financial firms. *Journal of Applied Corporate Finance*, **6**, 16–32.
- Myers, S. C. and Read Jr., J. A. (2001). Capital allocation for insurance companies. *Journal of Risk and Insurance*, (68), 545–580.
- Nelsen, R. B. (1999). *An introduction to copulas*. Number 139 in Lecture Notes in Statistics. Springer.
- Office of the Superintendent of Financial Institutions (2003). *Minimum Continuing Capital and Surplus Requirements (MCCSR) for Life Insurance Companies*. Canada.
- Overbeck, L. (2000). Allocation of economic capital in loan portfolios. In J. Franke, W. Härdle, and G. Stahl, editors, *Measuring risk in complex stochastic systems*, pages 15–30.
- Panjer, H. (2002). Measurement of risk, solvency requirements, and allocation of capital within financial conglomerates. IIPR technical report 01-14, University of Waterloo.
- Panjer, H. H., Boyle, P. P., Cox, S. H., Dufresne, D., Gerber, H. U., Mueller, H. H., Pedersen, H. W., Pliska, S. R., Sherris, M., Shiu, E. S., and Tan, K. S. (2001). *Financial economics*. The Society of Actuaries, Schaumburg, IL.
- Rieder, H. (1994). *Robust Asymptotic Statistics*. Springer-Verlag, New York.

- Rychlik, T. (1998). Bounds for expectations of L-estimates. In N. Balakrishnan and C. R. Rao, editors, *Handbook of Statistics*, volume 16, Amsterdam. Elsevier.
- Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, **14**(1), 115–129.
- Shao, J. and Tu, D. (1995). *The jackknife and bootstrap*. Springer Series in Statistics. Springer, New York.
- Sherris, M. (2006). Solvency, capital allocation and fair rate of return in insurance. *Journal of Risk and Insurance*, **73**(1), 71–96.
- Staudte, R. G. and Sheather, S. J. (1990). *Robust estimation and testing*. John Wiley & Sons, New York.
- Tasche, D. (1999). Risk contributions and performance measurement. Working paper, Zentrum Mathematik (SCA), TU Munchen, Germany.
- Valdez, E. A. (2004). Tail conditional variance for elliptically contoured distributions. Technical report, University of New South Wales.
- Valdez, E. A. and Chernih, A. (2003). Wang’s capital allocation formula for elliptically contoured distribution. *Insurance: Mathematics and Economics*, **33**, 517–532.
- Wand, M. P., Marron, J. S., and Ruppert, D. (1991). Transformations in density estimation (with discussion). *Journal of the American Statistical Association*, **86**, 343–361.
- Wang, S. S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, **26**(1), 71–92.

- Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, **67**(1), 15–36.
- Wang, S. S. (2002). A set of new methods and tools for enterprise risk capital management and portfolio optimization. *Working paper. SCOR Reinsurance company.*
- Wang, S. S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, (21), 173–183.
- Wilkie, A. D. (1986). A stochastic investment model for actuarial use. *Transactions of the Faculty of Actuaries*, **39**, 341–381.
- Wilkie, A. D. (1995). More on a stochastic asset model for actuarial use. *British Actuarial Journal*, **1**(V), 777–964.
- Wirch, J. L. and Hardy, M. R. (1999). A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, **25**, 337–347.
- Wirch, J. L. and Hardy, M. R. (2000). Distortion risk measures: Coherence and stochastic dominance. *Working paper.*
- Zhang, Y., Lu, P., and Jones, B. (2005). Report on the cash flow testing software web survey. Technical report, The Society of Actuaries.
http://www.soa.org/ccm/cms-service/stream/asset?asset_id=9170075&g11n
(accessed February 8, 2007).