ABSTRACT

LU, NA. Statistical Issues in Coherent Risk Management. (Under the direction of of Professor Peter Bloomfield.)

Measuring risk is a crucial aspect of the portfolio optimization problem in finance, and of capital adequacy assessment in risk management. Expected Shortfall (ES_α) has been proposed as a coherent risk measure, by contrast with Value-at-Risk (VaR_α) and the standard-deviation-type of measures. Based on a coherent risk measure, for instance ES_α, we can discuss a coherent capital allocation for the purpose of internal risk management and performance measure, if ES_α is used for economic capital held by financial firms as a cushion to absorb the unexpected losses. Properly allocating risk capital down to the business level is important for the purpose of risk management and portfolio performance measurement. Even if there is a doubt about the reason for allocating ES_α instead of VaR_α, the statistical properties of the statistic, marginal ES_α, from the proposed coherent allocation rule, are still of interest, because it is exactly the sensitivity of the target portfolio's ES_α.

The idea of a coherent capital allocation rule by using a cost sharing rule, the Aumann-Shapley value in game theory, proposed by Denault (2001), happens to result in the same formula as proposed by Tasche (2000), who independently develops the "suitable" allocation rule based on the discussion of risk-adjusted returns. The fact, that two aspects of the concerns are satisfied by the same allocation formula, brings

two fields together in an integrated way, so that a systematic risk management in a banking system seems very promising.

Fundamental statistical issues arise in several places in a coherent risk management system. Primary interests will be, and are always, in modeling the profit/loss (P/L) distribution. Statistical modeling is receiving more and more attention currently, as well as economic modeling. For our purpose, we place more emphasis on the estimation and inference of ES_α and allocation statistics (marginal contribution of ES_α) under different situations. We also modify the back-testing rules based on ES_α. We propose a collection of weighted test statistics aiming at detecting the under-estimated ES_α. Asymptotic properties of the test statistics are offered. The power of the tests in the context of an exponential family and the local alternatives is provided and the optimal weighting scheme is discussed.

STATISTICAL ISSUES IN COHERENT RISK MANAGEMENT

by

Na Lu

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics, minor in Economics

in the

GRADUATE SCHOOL
at
NC STATE UNIVERSITY
2004

Professor Peter Bloomfield
Chair of Advisory Committee

Professor John Seater
Minor Representative

Professor Marc Genton

Professor Albert Kyle

Professor Sastry Pantula

UMI Number: 3154326

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

STATISTICAL ISSUES IN COHERENT RISK MANAGEMENT

To all the people who love me,

whom I love.

Biography

Na Lu was born in Chaoyang, China, to parents Zuozheng Lu and Zhihua Gu on April 4, 1973. She earned her Bachelor of Science degree in Mathematics in July of 1995 from Liaoning University, China, and a Master of Science degree in Applied Mathematics in July of 1998 from Nankai University, China. In August of 1998, she was enrolled in the Department of Statistics at North Carolina State University in Raleigh, North Carolina, USA, and received her Master of Statistics degree in May of 2000. She continued her study at NCSU to pursue a Doctor Degree in Statistics with a minor in Economics.

Acknowledgements

This dissertation would not have been possible without Dr. Peter Bloomfield who not only served as my advisor, doctoral committee chair, but also inspired and challenged me through my doctoral study. He and other committee members, Dr. Genton, Dr. Kyle, Dr. Pantula and Dr. Seater, patiently guided me through the dissertation process. I thank them all.

I want to thank my friends, Jing Wang, Kathryn and Dave Williams, Dr. Kai Wang, Dr. Xiaoyun Yang, and Dr. Jimmy Doi for their hearty supports in these years. Their attitudes toward life deeply influence me and enlighten my life.

Special thanks to my parents, brother and husband, their endless supports and understanding are always the impulse of my journey to the academic world.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction

1.1 Motivation

This research was originally motivated by the recent advances in quantitative financial risk management theory, especially by emerging concepts of risk measures, Value-at-Risk, Shortfall Risk, Coherent Risk Measure, and so on. These concepts in the field of quantitative risk management are defined in terms of quantiles, conditional moments and partial moments in general, with respect to the uncertain outcomes concerned. The uncertain outcomes can be profit/loss of certain assets, return of particular portfolios, or traditional pure losses caused by down-side risks.

Measuring risk is a crucial aspect of the portfolio optimization problem in finance, and of capital adequacy assessment in risk management. Expected Shortfall (ES_α) has been proposed as a coherent risk measure compared to Value-at-Risk (VaR_α) and the standard-deviation-type of measures. ES_α is known to satisfy four properties of coherency as a risk measure, especially the property of sub-additivity, while VaR_α or variance type of measures do not, except in some special situations – the underlying distributions of portfolio returns are jointly normal, or utility functions of individuals are quadratic, for instance.

Based on a coherent risk measure, for instance ES_α, we can discuss a coherent capital allocation for the purpose of internal risk management and performance measure, if ES_α is used as economic capital held by financial firms as a cushion to absorb the unexpected losses. Properly allocating risk capital down to the business level is important for the purpose of risk management and portfolio performance measurement. Even if there is a doubt about the reason for allocating ES_α instead of VaR_α, the statistical properties of the statistic, marginal ES_α from the proposed coherent allocation rule, are still of interest, because it is exactly the sensitivity of the target portfolio's ES_α.

The idea of a coherent capital allocation rule by using a cost sharing rule, the Aumann-Shapley value in game theory, proposed by Denault (2001), happens to result in the same formula as proposed by Tasche (2000), who independently develops the "suitable" allocation rule based on the discussion of risk-adjusted returns. The former is based on axiomatic game theory. By imposing reasonable conditions on an allocation rule to be "fair" as a cost sharing rule, the Aumann-Shapley value is the unique allocation. The latter is based on a popular management concept of risk adjusted performance measurement - Return on Risk-adjusted Capital(RORAC), which is an optimization objective function of portfolio allocation, measured as ratio of expected cash flow and economic capital. The fact, that two aspects of the concerns are satisfied by the same allocation formula, brings two fields together in an integrated way, so that a systematic management in a banking system seems very promising.

Fundamental statistical issues arise in several places in a coherent risk management system. Primary interests will be, and are always, in modeling the profit/loss (P/L) distribution. Statistical modeling is attracting more attention currently, as well as economic modeling. For our purpose, we place more emphasis on the estimation and inference of ES_α and allocation statistics (marginal contribution of ES_α) under different situations. We also modify the back-testing rules based on ES_α.

1.2 Outline of the Thesis

This thesis has three parts. Part I - Introduction- has one chapter, Chapter 1. Chapter 1 consists of motivation of this research and the outline of the thesis. Part II - Economic Theory - has three chapters, Chapter 2, 3, 4. Chapter 2 is an introduction to risk measure and decision problems. We particularly focus on risk measure and financial/risk management problems in portfolio selection/risk capital assessment. Chapter 3 is an introduction to coherent risk management - coherent risk measure and coherent risk capital allocation. Chapter 4 is about coherent allocation and suitable allocation. Part III- Statistical Theory - has two chapters, Chapter 5 and

6. In Chapter 5, we discuss the different modeling methods for ES_α and Sensitivity of ES_α (SES_α). In Chapter 6, we propose a class of test statistics for the purpose of back-testing ES_α, and give their properties: consistency and power considerations. We draw conclusions in Chapter 7, which constitutes Part IV.

1.3 Contributions

This dissertation addresses both economic and statistical issues in a coherent risk management system. Statistical issues are the focus of this thesis. Because of the inherent complexity of the behavior of the uncertain outcomes - financial P/L, this thesis investigates the statistical estimation methods for the coherent risk measure (ES_α) and coherent allocation of risk capital (SES_α), and modifies back-testing procedure based on ES_α. The dissertation makes two main contributions to the field of coherent risk management.

First, we propose a consistent empirical estimator for SES_α, and give analytic formula for VaR_α, ES_α and SES_α of the multivariate t-distribution and the general elliptical family. Monte Carlo simulations are performed to evidence the slow converging speed of the proposed empirical estimator in a multivariate t-distribution framework. When some prior knowledge of the true distribution family is known, an accelerate Monte Carlo simulation can be used to speed up the converging process. Based on the analytic formula for SES_α, we discover the proportional invariance property of the allocation statistics for the centered multivariate t-distribution and the centered elliptical family.

Second, we modifies the back-testing procedure based on ES_α. We propose a collection of consistent test statistics based on ES_α with different deterministic weighting functions imposed on the lower tail of the distribution. We inverse the characteristics functions with the Fast Fourier Transformation (FFT) combined with linear interpolation methods to tabulate the critical values for the two special cases - test statistics with equal weighting and reciprocal weighting functions. The exact method and the approximating methods are utilized and compared at a three-digit precision in this context. We further prove the asymptotic normality of the test statistics in a generalized framework, where the weighting function can be either uniformly bounded deterministic function, or a uniformly integrable and converging random weighting process. The optimal weighting scheme, in terms of optimal local power, is discussed for a special exponential family.

Part II

Economic Theory

Chapter 2

Risk Measure and Decision Problems

People always have to choose actions from a given set of alternatives with uncertain consequences. Consider, for instance, an investor who has to choose his portfolio among different investment opportunities, or an individual who has to decide whether or not to buy a super-ball ticket. We will call the uncertain consequence "risk", although conventionally people are using the term "risk" for different purposes and it has no clear definition. We describe the risks by random variables with distributions that are known to the decision maker. This is sometimes called the situation of decision under risk in contrast to decision under uncertainty where the decision maker doesn't know the distributions of the random variables.

Von Neumann and Morgenstern (1947) [57] described some axioms for a rational decision maker making decisions under risk, which imply the so-called expected utility hypothesis. This hypothesis says that, for a rational decision maker, there exists a utility function $u(\cdot) : \mathbb{R} \to \mathbb{R}$ such that he/she prefers Y to X if and only if $Eu(X) \leq Eu(Y)$. Given exactly a utility function, and distributions of X and Y, it is not difficult to reach a conclusion by making the pairwise comparison. In practice, since it is impossible to know exactly the decision maker's utility function, we are interested in finding rules for the distributions of random variables that enable prediction of the decision maker's choice when we partially know the properties of his utility function: monotonicity, and concavity, for instance. Or similarly, we are interested in finding out whether a group of decision makers with different utility functions will reach the same decision or not. Stochastic Dominance (SD) rules were developed aiming to answer such questions.

Stochastic Dominance The central idea of the SD rules is to simplify the decision problem by sorting out the dominated alternatives. SD allows pairwise comparison of cumulative distribution functions. Let $F(x)$ and $G(x)$ be cumulative distribution functions of X and Y, respectively. X dominates Y by First (FSD), Second (SSD), or Third Order Stochastic Dominance (TSD) if and only if

$$
\begin{aligned}
&F(x) \leq G(x), \forall x \in \mathbb{R}, && \text{(FSD)}\\
&\textstyle\int_{-\infty}^{v} [G(x) - F(x)]dx \geq 0, \forall v \in \mathbb{R}, && \text{(SSD)}\\
&\textstyle\int_{-\infty}^{t} \int_{-\infty}^{v} [G(x) - F(x)]dxdv \geq 0, \forall t \in \mathbb{R} \text{ and } \int_{-\infty}^{\infty} F(x)dx \leq \int_{-\infty}^{\infty} G(x)dx, && \text{(TSD)}
\end{aligned}
\tag{2.1}
$$

It has been shown that the stochastic ordering from the definition of FSD, SSD, and TSD is consistent with the ordering generated by maximizing expected utilities of the specific classes,

$$\begin{aligned}
&U_1 \equiv \{u(x)|u^{'}(x) > 0, \forall x \in \mathbb{R}\} && \Longleftrightarrow \text{ FSD} \\
&U_2 \equiv \{u(x)|u^{'}(x) > 0, \text{ and } u^{''}(x) < 0, \forall x \in \mathbb{R}\} && \Longleftrightarrow \text{ SSD} \\
&U_3 \equiv \{u(x)|u^{'}(x) > 0, u^{''}(x) < 0 \text{ and } u^{'''}(x) > 0, \forall x \in \mathbb{R}\} && \Longleftrightarrow \text{ TSD}
\end{aligned} \tag{2.2}$$

SD rules are playing a prominent role in the literature on decision under risk (Quirk and Saposnik [43], 1962, Fishburn[26], 1964, etc.). The applications of SD rules are widespread in portfolio management, and risk management [1] among others. Since SD rules involve pairwise comparison of the set of alternative probability distributions, it is useful for problems with prespecified and finite number of alternatives, such as capital budgeting type problems, but is computationally infeasible for portfolio-selection-type of problems involving an infinite convex set of choices among alternative probability distributions. In addition, the theory requires the decision maker to take the complete distribution of outcomes into consideration; the decision task can be simplified if one could concentrate on a few attributes that contain the complete information about the distribution under consideration. This idea lies at the heart of the so-called two parameter selection rules, or mean-risk (or return-risk) approaches.

[1] Through the whole discussion of this dissertation, we will loosely use "risk management" for "financial risk management", specifically for market risk management and credit risk management, in which probabilistic approach plays an important role.

The two parameters are the return measure and the risk measure.

Coherence There has been no question of using mean as a return measure, but people have been arguing what measures are appropriate for risk for centuries, not only for portfolio selection problems but also for risk management problems. Artzner, Delbaen and Heath (1997, 1999) [5, 6] proposed the necessary properties for a risk measure to be coherent. By using this criterion, we can discuss the coherency of the proposed risk measures under different conditions.

In section 1.1, for portfolio selection problem, we will introduce Markowitz-Tobin (1959, 1958) [40] mean-minimum variance (MV) selection rule, and Bawa (1975, 1976, 1977, 1978) [11, 12, 13] mean-lower partial moment (MLPM) functional rules (including Roy's Safety-First (SF) rules), which use variance and lower partial moments as risk measures, respectively. In section 1.2, we consider using a risk measure to calculate economic capital held by financial firms as a cushion to absorb the unexpected losses, which is a typical discussion in risk management for financial firms and insurance companies.

2.1 Risk Measure and Portfolio Selection

The traditional way of reducing the dimensionality of the SD rules is to add restrictions to the probability distributions of security and portfolio returns, assuming normal or stable distributions. In the case of normal distribution, the SD rule reduces to the well-known Markowitz-Tobin mean-minimum variance (MV) selection rule for risk averse individuals: the mean-variance-frontier type of discussion. MV is an optimal rule for risk averse individuals with increasing and concave utility functions when the distribution of portfolio returns is normal. Samuelson (1958) [51] and Ross (1976) [48] showed that MV is a reasonable approximation to the optimal selection rules when "riskiness" of returns is limited or the number of securities is large enough, i.e. $n \to \infty$.

In this situation, variance or standard deviation is a natural risk measure. However, the normal-distribution assumption can not be upheld even for market risk as long as the portfolio includes derivatives, let alone the empirical evidence of fat-tailed distribution of returns. The normal distribution assumption seems to be an inadequate approximation in the case of a credit portfolio.

An alternative way is to put restrictions on decision maker's utility functional space, assuming quadratic utility functions,

$$u(x) = ax - bx^2.$$

where $a, b \in \mathbb{R}_+$.

In the case of quadratic utility function, the SD rule reduces to MV again, but quadratic utility functional space is too limited to represent decision makers' preferences.

MLPM rules seem to be a more general framework than the above two approaches.

Without making undue distributional restrictions, it has been shown that the SD rules are MLPM rules where the Lower Partial Moment (LPM) functionals, computed at every point in the domain of the underlying random variable, may be viewed as the risk measure.

This class of risk measures is consistent with the definition of increasing risk; see Rothschild and Stiglitz (1970), Machina and Pratt (1997) [49, 38] (for arbitrary probability distributions). LPMs are measures of downside or shortfall risk in the sense that only negative deviations from a target outcome are taken into consideration. In the case of continuous distributions with outcomes $x \in (-\infty, \infty)$, each LPM can be computed as follows,

$$LPM_n^t(X) = \int_{-\infty}^{t} (t-x)^n dF(x),$$

where X is a random variable, $F(x)$ is its cumulative distribution function (cdf), and $n \in \mathbb{Z}$.

Some of the most frequently used risk measures are special cases of LPMs. For instance, semi-variance corresponds to the LPM_2^{μ}, Value at Risk (VaR_α) is a sort of LPM_0^t, and Expected Shortfall (ES)[2] is the LPM_1^t.

The LPM approach is of special importance for applications in portfolio theory, and to risk management as well, as we will see in the discussion of the later chapters.

[2]Following Definition 1 and Definition 3 in the later chapter, it is not difficult to verify that for a continuous variable X,

$$\mathrm{VaR}_\alpha(X) = LPM_0^t(X)$$

and

$$\mathrm{ES}_\alpha(X) = \frac{1}{\alpha} LPM_1^t - t$$

where $t = \mathrm{VaR}_\alpha(X)$.

Plausible ordering properties of LPMs, applicable to arbitrary distributions and consistent with utility theory are:

For X to be preferred to Y, it is necessary and sufficient that (see Bawa, 1978, [13]),

- $\forall u(x) \in U_1 : LPM_0^t(X) \leq LPM_0^t(Y), \forall t \in \mathbb{R}$, with at least one strict inequality for t,
- $\forall u(x) \in U_2 : LPM_1^t(X) \leq LPM_1^t(Y), \forall t \in \mathbb{R}$, with at least one strict inequality for t,
- $\forall u(x) \in U_3 : LPM_2^t(X) \leq LPM_2^t(Y), \forall t \in \mathbb{R}$, with at least one strict inequality for t and $\int_{-\infty}^{+\infty} F(x)dx \leq \int_{-\infty}^{+\infty} G(x)dx$.

The Safety-First (SF) rule, introduced by Roy (1952) [50], stipulates choice of an alternative that provides a target mean return while minimizing the probability of the return falling below some threshold of a disaster. The SF rule has generally been regarded as outside the expected utility paradigm, but it can be naturally generalized to n^{th} order SF rules using higher order LPMs of the probability distributions; see Bawa (1978) [13].

It is not surprising that VaR_α is in the spirit of Roy's SF rules as a risk measures, while Expected Shortfall (ES_α) is a sort of Bawa's Generalized SF/MLPM, as we will discuss in the later chapters. MV is equivalent to both mean-VaR_α and mean-ES_α, when normality is assumed.

2.2 Risk Measure and Risk/Economic Capital

As we mentioned before, the uncertain consequence is called "risk" and mathematically we use random variables to describe risks. Particularly, for the portfolio selection problems and/or the risk management problems, risks are the uncertainty of the returns of the portfolios[3].

So far, individual investors, financial firms or insurance companies are all treated as decision makers facing choice (or decision) under risk. Next, in order to study the risk management problems, we will treat them separately, because they have different characteristics, in terms of the risk management problems. We will focus on the corporate aspects of the risk management problems after we discuss their differences.

Unlike financial firms or insurance companies, individual investors may or may not buy an insurance (if there exists one, or simply keep certain amount of risk-free asset, or cash) to protect themselves from big losses, if they are not using borrowed money for their portfolio investment or if they are not investing in derivatives. They may be typically suggested to do so anyhow, but it is totally up to them. In this case, risk measure, in the two-parameter selection rules, is nothing but a screening device, to screen out the portfolios with targeted returns, but higher risks, and the risk measure need not to be linked with the concept of economic capital, which we will introduce in the following discussions.

[3]Or, the uncertainty of the net worths of the portfolios, as in Artzner (1999) [6].

Financial firms or insurance companies could well, and would often be regulated to, hold an amount of risk-less capital, as an insurance/collateral against the uncertainty of the return of their portfolios; this is the so-called risk/economic capital. Consider a financial firm: it can be a highly leveraged investor, i.e. an investor financing his assets largely with borrowed money. His risk is creditor's risk, and his loss is creditors' loss. If his investments do not generate enough returns, the investor has to fall back on his equity to meet his obligations. He goes bankrupt as soon as the equity is exhausted, and the creditors suffer losses. In addition, the influence of such events might be even worse to the stability of the whole economic system; we can remind readers easily by the following events:

- In December 1994, Orange County stunned the markets by announcing that its investment pool had suffered a loss of \$1.6 billion (interest rate swaps). This was the largest loss ever recorded by a local government investment pool, and led to the bankruptcy of the county shortly thereafter.

- In 1995, Barings, London's oldest merchant bank, lost £830 million on a speculative position in Nikkei 225 stock index futures and went bankrupt.

- In 1995, Daiwa Bank lost \$1.1 billion - one seventh of the bank's capital because of concealed trades.

- In 2002, Allied Irish Bank lost \$700 million because of concealed trades.

These are just a few examples from a long list in the history. If they had held enough risk capital based on their risk exposures, they might have survived and some of their creditors might not have suffered losses.

In the case of a highly leveraged investor or financial firms, some risk measures (like LPMs) from two-parameter selection rules, can be not only a screening device, but also part of the calculation of risk/economic capital among other reasons.

VaR_α has become a popular risk measures for risk capital adequacy calculation, since the Basel Committee on Banking Supervision (1996) [10] permitted banks to make use of it in their internal models, for the capital required by market risks. However, other risk measures, like ES_α, are potentially suitable for this purpose, too. Moreover, it has been shown that ES_α is coherent but VaR_α is not, as we will discuss in the next part, although ES_α is not yet widely used.

After introducing different risk measures that have been proposed from different perspectives, we will give a close look to the risk measures proposed for risk capital evaluation: VaR_α and ES_α. ES_α is often cited as a coherent risk measure, while VaR_α is not.

Properly allocating risk capital down to the business level is of importance for the purpose of performance measurement and risk management as well, which is often referred as risk capital allocation in the literatures. We will introduce the concept of coherent allocation of risk capital in this part, and use the term coherent risk management to refer to both concepts of coherency in the whole discussion.

While coherent risk measure is based on the axiomatic discussion of the properties that a reasonable risk measure should bear, coherent allocation of risk capital stems from the well-known game theoretic discussion of Aumann-Shapley values and the "suitable" allocation rule based on the discussion of risk-adjusted returns on capital (RORAC). Two independent discussions of the risk capital allocation mechanisms result in the same allocation formula when ES_α is utilized for risk capital evaluation, or more generally a coherent risk measure is used. We further discuss the related findings from this "coincidence".

In Chapter 3, we will introduce the concept of Coherent Risk Management consisting of both Coherent Risk Measure and Coherent Allocation of Risk Capital. In Chapter 4, we will discuss the two approaches, Aumann-Shapley value and "suitable" allocation rule, for the coherent allocation of risk capital.

Chapter 3

Coherent Risk Management

The concept of coherent measures of risk has been introduced recently in a series of papers, including [5, 6, 7], by P. Artzner, F. Delbaen, S. Eber and D. Heath, for the purpose of risk capital calculation. Coherent measures of risk, discussed in the next section, are defined through a set of axioms on a linear space of random variables – representing the feasible choices of a portfolio. VaR_α as a risk measure is heavily criticized for not being coherent, especially for not being sub-additive (see Artzner et al., 1999, [6]). ES_α has been proposed as a natural remedy for the deficiencies of VaR_α.

The concept of coherent allocation of risk capital has been introduced by Denault (2001) in [22], for the allocation of risk capital. A coherent allocation of risk capital is defined through a set of properties to be fulfilled by a fair risk capital allocation principle, when the calculation of economic capital is based on a coherent risk measure.

We borrow the popular expression, "coherent", and talk about coherent risk management, embracing both concepts, here. In section 3.1, we will introduce the concept of coherent measures of risk, and different versions of ES_α. In section 3.2, we will introduce the concept of coherent allocation of risk/economic capital and an allocation rule based on ES_α.

3.1 Coherent Risk Measure and Expected Shortfall

Mathematically, a risk measure is a real valued function

$$r : A \to \mathbb{R} \cup \{\infty\},$$

where $A \subset \mathbb{R}_+^n$, a vector space consisted of the weights of assets in the portfolio, given a base of the linear space representing the feasible assets of the decision maker. We will not impose any special property on the function $r(\cdot)$ to be a risk measure for the time being, but we will always assume it is a differentiable function. Tasche (2000) [55] and Scaillet (2002) [52] gave some differentiability conditions of risk measures. The conditions are relatively mild, in comparison with temerarious assumptions common in the area of risk management. For some specific risk measures like VaR_α and ES_α, explicit first order derivatives were provided in these two papers.

Mapping the riskiness of a set of portfolios by a single function is not an easy task.

Some knowledge about the portfolio is required, such as worse-case scenarios based on human insights or statistical models of the portfolio cash flows based on historical data. We will focus on two risk measures, VaR_α and ES_α, which need statistical models for the cash flows.

Value-at-Risk, or VaR_α for short, a widely used risk measure, answers the question: what is the minimum loss incurred in the α worst cases of the market changes that affect the value of the portfolio? So, values of α is desirable to be close to 0. VaR_α has become one of the most important and generally accepted measures of risk, which has achieved the high status of being written into industry regulations (see, for instance, [42] by M. Pritsker).

To be consistent with the notions in [6] and [3], we define some real valued random variable X on the probability space $(\Omega, \mathcal{A}, P)$, the random P/L of some asset or portfolio.

We formally define VaR_α in the following way.

Definition 1 *Value at Risk.* Given $\alpha \in [0,1]$, the value-at-risk VaR_α of a random variable X with cumulative distribution function $P(\cdot)$ at level α is,

$$q_\alpha^+(X) = \sup\{x \in \mathbb{R} \mid P[X < x] \leq \alpha\}, \;\; \alpha \in [0,1)$$

$$\mathrm{VaR}_\alpha(X) = -q_\alpha^+(X).$$

When the portfolio has a joint normal/elliptical distribution for its assets P/L, X

and Y, we know that VaR_α satisfies:

$$\mathrm{VaR}_\alpha(X+Y) \leq \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y),$$

which is the so-called sub-additivity property. This property reflects the fact that a diversified portfolio should be less risky than a non-diversified one to any investors. Any sensible risk measure should satisfy this basic property if risk measures are supposed to be "good" indicators of riskiness of portfolios, in the sense that more risky, higher risk measures. But when the portfolio does not have a joint normal/elliptical distribution of its assets P/L, VaR_α might not have such a necessary property as risk measures, which is very easy to be illustrated with a counter example, such as the one in 3.3 of [6], page 216.

If any joint P/L can be approximated by a joint normal/elliptical distribution, life will be much easier. VaR_α can be a very good candidate for risk measures, because VaR_α has very good interpretations and is very easy to be understood among other reasons. But more and more empirical results have shown that joint normal/elliptical distributions are not a good approximation for the real world. Then, it becomes an interesting question that whether there exists a "good" (coherent) risk measure, which does not depend on distributional properties of random variables but still keep more nice properties than VaR_α. Artzner et al. (1999) [6] defined the coherency for risk measures in a very general framework, and examined the popular risk measures that are currently used for different purposes. Their definition of coherent risk measures is as follows,

Definition 2 *Coherent Measure of Risk.* Consider a set $0 \in V$ of real-valued mappings from a measurable space $(\Omega, \mathcal{A})$ to $\mathbb{R}$,

$$\rho : V \to (-\infty, \infty],$$

with $\rho(0) = 0$. We call this mapping a coherent measure of risk, if it is

1. monotonic:

$$X, Y \in V,\ X \leq Y \ \Rightarrow\ \rho(X) \geq \rho(Y),$$

2. sub-additive:

$$X, Y, X + Y \in V \ \Rightarrow\ \rho(X + Y) \leq \rho(X) + \rho(Y),$$

3. positively homogeneous:

$$X \in V,\ h > 0,\ hX \in V \ \Rightarrow\ \rho(hX) = h\rho(X),$$

4. translation invariant:

$$X \in V,\ a \in \mathbb{R},\ X + a \in V \ \Rightarrow\ \rho(X + a) = \rho(X) - a$$

This coherency is defined in a very general way, since a probability measure is not even defined on the measurable space $(\Omega, \mathcal{A})$. In the decision-under-risk problems, the probability measure P on the measurable space is assumed to be known and can be used to construct risk measures, usually in terms of marginal probability measures. This leads to the following discussion of probabilistic/statistical risk measures: VaR_α and ES_α.

ES in several variants has been proposed as a remedy for the deficiencies of VaR_α, which in general is not a coherent risk measure. ES_α is a natural coherent measure comparing to VaR_α. The formal definition of ES_α is as follows,

Definition 3 *Expected Shortfall.* Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\alpha \in (0, 1)$. Consider a real random variable X on $(\Omega, \mathcal{A}, P)$ with $E(X^-) < \infty$. Then

$$ES_\alpha(X) = \alpha^{-1} \int_0^\alpha VaR_\mu(X)\, d\mu = -\alpha^{-1} \int_0^\alpha q_\mu^+(X)\, d\mu.$$

Remark 1 *Expected Shortfall.*

- Acerbi et al. [2] first named Expected shortfall. Delbaen [22] and Kusuoka [37] considered ES_α before without naming it and Rockafellar and Uryasev [44, 45] called it Conditional Value-at-Risk ($CVaR_\alpha$) and Artzner et al. [5, 6] call it tail conditional expectation (TCE_α)[1]. Bertsimas et al. [14] gave different versions of Expected Shortfall for continuous distributed P/L.

- Acerbi (2002) [1] studies a space of coherent risk measures M_ϕ, based on the expansions of some coherent elementary basis measures, for instance, ES_α,

$$M_\phi(X) = -\int_0^1 q_p^+(X)\phi(p)dp,$$

with $\phi \in \mathcal{L}^1[0, 1]$. $M_\phi(X)$ is also called "spectral measures of risk generated by

[1]Notice that TCE_α and $CVaR_\alpha$ are not coherent risk measures in general. They coincides with ES_α (and hence coherent) only under the certain conditions such as the case of continuous distributed P/L (see e.g. Acerbi and Tasche (2002)[3]).

ES_α". He further gives ϕ a name "risk aversion function", which is mathematically a weighting function with properties

1. $\phi > 0$ in $\mathcal{L}^1[0,1]$,

2. ϕ is non-increasing[2] in $\mathcal{L}^1[0,1]$,

3. $||\phi||_1 = \int_0^1 |\phi(p)|dp = 1$.

From economics point of view, the choice of weighting functions, or "risk spectrum" in $\mathcal{L}^1[0,1]$ or $\mathcal{L}^1[0,\alpha]$ will be directly related to the risk capital calculation. Thus, the normalized condition is critical in order to have a coherent induced measure. We will discuss, in the later chapters, that the weighting functions for constructing test statistics purposes can be based on a much richer functional space and can be optimally determined in terms of the testing power. As we will point out, the intuition for the whole weighting mechanism remains the same: the more the risk capital is kept aside, the more "risk averse" is the risk manager.

- Kusuoka (2001) [37] and Acerbi (2002) [1] argue that ES_α is the "smallest" coherent and "law invariant"[3] risk measure to dominate VaR_α, which was men-

[2] We change Acerbi's discussion on this property from "decreasing" to "non-increasing" after we notice that equal weighting is a special case. For example, the basis measures generating the space are natural equal weighting cases.

[3] A rough interpretation of law invariance might be "can be estimated from statistical observations only", which makes statistical analysis of great importance in this field. Mathematically, it is

$$X, Y \in V, P[X \le t] = P[Y \le t] \text{ for all } t \in \mathbb{R} \Longrightarrow \rho(X) = \rho(Y)$$

tioned previously in Delbaen (1998) [21][4]. Dasche (2002) [54] give a thorough summary of various properties of ES_α.

3.2 Coherent Allocation of Risk/Economic Capital

The second use of the term "coherency" is for risk capital allocation. Recall that risk capital is risk-free, low return assets, which are held as an insurance/collateral against the unexpected losses of the portfolios in the firm. Financial firms will generally incur dead-weight cost by holding an amount of low return risk capital. It is natural to find a fair allocation of the burden among the constituents, especially when the allocation provides incentives for constituents to reduce risks while keeping a certain level of returns, or for the constituents' internal performance comparison purposes (for example in a RORAC approach, see Tasche [55]).

Again, the problem of risk capital allocation is interesting and non-trivial, because the sum of the risk capital of each constituent is usually larger than the risk capital of the firm taken as a whole. The advantage of reducing total costs by pooling firms' activities need to be shared fairly among constituents. This kind of problems will naturally fall into the category of sharing joint costs or surplus in economic theory in large. The current literatures offered two direct resolutions to this specific risk capital

[4]It is a more probabilistic discussion.

allocation problem. They are Denault [22]'s game theoretic (axiomatic) approach and Tasche [55]'s RORAC approach. As we will show that they offer the same solution formula for the risk capital allocation rules. Denault [22] is a thorough discussion of a game theoretic (axiomatic) approach and defines the coherent allocation of risk capital. Tasche [55]'s RORAC approach reaches the similar results by defining "suitableness" of risk contributions (allocation), whose total risk capital can be based on any risk measures.

1. Game theoretic approach by Denault (2001) [22]

   In Denault (2001) [22], the risk capital allocation problem is modeled as a coalitional fuzzy game between constituents or portfolios. The Aumann-Shapley value emerges as a most promising allocation principle for the fuzzy game (or coalitional game with fractional players). Mathematical details will be furnished in the next chapter for the purpose of comparisons.

2. RORAC approach by Tasche (2000)

   In Tasche (2000) [55], through defining "suitable" allocation for performance measurement, the same allocation formula was proposed for risk capital allocation. An allocation is called "suitable" if when the risk-adjusted return of a portfolio is "above/below average", then at least locally increasing/decreasing the share of this portfolio improves the overall return of the firm. We will give the detailed mathematical discussion in the next chapter for the purpose of comparisons.

Chapter 4

Coherent and Suitable Allocation

Again, the allocation of risk capital is interesting and non-trivial for the risk managers of financial institutions or insurance firms. In practice, different allocation schemes can be utilized internally for different purposes. As an example, we will consider allocation rules for both the down-side risk management problem and the portfolio selection problem in a finite economy [1].

Since risk capital is designed as a capital reserve to cover the potential big losses, which is exactly the case in the banking systems, one concern is about "fair" allocation in terms of sharing cost during the course of holding the low return assets. On the other hand, optimal portfolios need to be constructed to achieve the best profitability of the firm. Risk capital allocation schemes are critical in ranking the profitability of the assets in the portfolio, or in signaling the best investment opportunities. It

[1] By saying finite economy, we mean that there is a finite number of participants in the market, opposing to the usual equilibrium discussion based on infinite economy with infinite many participants in the economic literatures.

seems that one allocation scheme can serve both purposes if the risk measure or risk capital is a coherent one, given Denault (2001) [22] and Tasche (2000) [55]. While the former approach is focusing on the "fair" allocation and the latter is about "suitable" allocation for performance measurement, both approaches reach the same conclusion: the Aumann-Shapley value.

The structure of this chapter is as follows: in Section 3.1, we give the mathematical details in formulating the coherent risk capital allocation rules; in Section 3.2, suitable allocation based on RORAC approach is introduced; Section 3.3 is about the related research findings to the two approaches.

4.1 Coherent Allocation of Risk Capital - a Recap

The concept of Coherent Allocation of Risk Capital was introduced by Denault (2001) [22]. Before introducing this concept, we need to give some background about the fuzzy game (coalitional game with fractional players) and its core so that we can model the risk capital allocation problem as a fuzzy game later.

Definition 4 *A Fuzzy Game (A Coalitional Game with Fractional Players)*, (N, Λ, r), consists of

- a finite set N of players, with $|N| = n$;
- a positive vector $\Lambda \in \mathbb{R}^n_+$, each component representing for one of the n players his full involvement;

- a real-valued cost function $r : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad r : \lambda \mapsto r(\lambda)$ such that $r(0) = 0$.

Definition 5 *The Fuzzy Core ("no-undercut" condition)* of a fuzzy game (N, Λ, r) is the set of allocations [2] $k \in \mathbb{R}^n$ for which $\lambda' k \leq r(\lambda)$ for all $\lambda \in \mathbb{R}^n_+$, $0 \leq \lambda \leq \Lambda$ and $\Lambda' k = r(\Lambda)$.

The axioms needed for a risk capital allocation to be coherent was introduced as follows by Denault (2001):

Definition 6 *Coherent Allocation of Risk Capital.*

Let r be a coherent risk measure as defined in Chapter 3. A fuzzy value

$$\phi : (N, \Gamma, r) \longrightarrow k \in \mathbb{R}^n$$

is **coherent** if it satisfies the properties defined below, and if k is an element of the fuzzy core:

- **Aggregation Invariance** Suppose the risk measure r and $\bar{r}$ satisfy $r(\lambda) = \bar{r}(\Gamma\lambda)$ for some $m \times n$ matrix Γ and all λ such that $0 \leq \lambda \leq \Lambda$ then

$$\phi(N, \Lambda, r) = \Gamma' \phi(N, \Gamma\Lambda, \bar{r})$$

- **Continuity** The mapping ϕ is continuous over the normed vector space M^n of continuously differentiable functions $r : \mathbb{R}^n_+ \longrightarrow \mathbb{R}$ that vanish at the origin.

[2] $k = \phi(N, \Lambda, r)$, with ϕ defined in Definition 6.

- **Non-negativity under r non-decreasing (Monotonicity)** If r is non-decreasing, in the sense that $r(\lambda) \leq r(\lambda^*)$ whenever $0 \leq \lambda \leq \lambda^* \leq \Lambda$, then

$$\phi(N, \Lambda, r) \geq 0$$

- **Dummy Player Allocation** If i is a dummy player, in the sense that

$$r(\lambda) - r(\lambda^*) = (\lambda_i - \lambda_i^*)\frac{\rho(X_i)}{\Lambda_i}$$

whenever $0 \leq \lambda \leq \Lambda$ and $\lambda^* = \lambda$ except in the i^{th} component, then

$$k_i = \frac{\rho(X_i)}{\Lambda_i}$$

- **Fuzzy core** Defined as above, in Definition 5.

Definition 7 *Aumann-Shapley Value*

$$\phi_i^{AS}(N, \Lambda, r) = k_i^{AS} = \int_0^1 \frac{\partial r(\gamma\Lambda)}{\partial \lambda_i} d\gamma$$

Theorem 1 Aumann-Shapley Value is the Unique Coherent Fuzzy Value.

If (N, r, Λ) is a fuzzy game, with a coherent cost function, r, which is differentiable at Λ, then the Aumann-Shapley value is a unique coherent fuzzy value, and

$$\phi_i^{AS}(N, \Lambda, r) = k_i^{AS} = \frac{\partial r(\Lambda)}{\partial \lambda_i}$$

Proof: See Denault (2001) [22].

This is a core result from Denault (2001), which convince us with the existence and uniqueness of a coherent allocation rule - Aumann-Shapley value.

Example 1 *"sensitivity" of SES_α*, is the first order derivative (or gradient, in matrix language) of ES_α with respect to the weights of the asset in the portfolio, $a = (a_1, \cdots, a_n)'$. Let us define,

$$\mathrm{SES}^i_\alpha(a) = \frac{\partial \mathrm{ES}_\alpha}{\partial a_i}, \quad i = 1, \cdots, n.$$

When ES_α is employed as a cost function for the fuzzy games, SES_αs are exactly the Aumann-Shapley values in a linear portfolio setup by Theorem 1. Under certain continuity and differentiability conditions (see Tasche [55] and Scaillet [52]), the resulting formula is

$$\mathrm{SES}^i_\alpha(a) = E[-Y_i| - a^{'}Y > \mathrm{VaR}_\alpha(a^{'}Y)], \tag{4.1}$$

and the risk capital allocated to the i_{th} asset is $a_i \mathrm{SES}^i_\alpha(a)$.

4.2 Suitable Allocation for Performance Measurement

Suitable allocation approach by Tasche (2000) [55] is more or less sparked by the so-called two parameter selection rules for the portfolio selection problems, as we mentioned in the previous chapters. Remember that risk measure has a long history of serving as one of the two parameters in the mean-risk (or return-risk) approaches, during the earlier years of the research on decision under risk in the financial market. Representative discussions include Markowitz-Tobin (1952, 1958) [40], Bawa (1976, 1977, 1978) [11, 12, 13] MLPM functional rules (including Roy's SF rules), etc. Risk measures in those setups are nothing but a nonlinear (especially, convex) condition for the constraint optimization problems. Relating this to the utility theoretic discussion, we can percept the investors' level of risk aversion as we discussed in the previous chapters.

By imposing a subjective return function, as Tasche (2000) defined below, one can talk about an optimal portfolio selection rule if the risk measure/risk capital is a convex function of the shares of the assets in the portfolio. A suitable allocation for performance measurement will naturally need to bear the characteristics sensitive to diagnose/flag the better performing asset. Thus, locally improving a certain amount of the better performing asset should definitely improve the overall return rate (RO-RAC). Given the above intuition, Tasche (2000) defines the "suitability" in a rather general way, i.e. the risk measure does not have to be a coherent measure for this "suitability" to hold. It is a local condition, instead of a global one in supporting the unique solution to an optimization problem.

Definition 8 *Return Function*

Let $\emptyset \neq U$ be a set in $\mathbb{R}^n$ and $r : U \longrightarrow \mathbb{R}$ be some function on U. Fix any $m \in \mathbb{R}^n$, then the function $g : \{u \in U | r(u) \neq m'u\} \longrightarrow \mathbb{R}$, defined by

$$g(u) = \frac{m'u}{r(u) - m'u}$$

is called return function for r.

Return function incorporates both expected returns and required risk capitals of the feasible portfolios. It assigns a ranking score to each portfolio $u \in U$, for the performance ranking purpose.

Definition 9 *Suitable Allocation for Performance Measurement*

Let $\emptyset \neq U$ be a set in $\mathbb{R}^n$ and $r : U \longrightarrow \mathbb{R}$ be some function on U. A vector field $a = (a_1, \cdots, a_n) : U \longrightarrow \mathbb{R}^n$ is called suitable for performance measurement with r, if it satisfies the following two conditions:

- For all $m \in \mathbb{R}^n$, $u \in U$ with $r(u) \neq m' u$ and $i \in \{1, \cdots, n\}$, if $m_i r(u) > a_i(u) m' u$, then there exists an $\epsilon > 0$, such that for all $t \in (0, \epsilon)$ we have $g(-te^{(i)} + u) < g(u) < g(te^{(i)} + u)$;

- For all $m \in \mathbb{R}^n$, $u \in U$ with $r(u) \neq m' u$ and $i \in \{1, \cdots, n\}$, if $m_i r(u) < a_i(u) m' u$, then there exists an $\epsilon > 0$, such that for all $t \in (0, \epsilon)$ we have $g(-te^{(i)} + u) > g(u) > g(te^{(i)} + u)$,

where $e^{(i)} \in \mathbb{R}^n$ denotes the i^{th} canonical unit vector in each condition.

Theorem 2 *Let $\emptyset \neq U \subset \mathbb{R}^n$ be an open set and $r : U \longrightarrow \mathbb{R}$ be a function that is partially differentiable in U with continuous derivatives. Let $a = (a_1, \cdots, a_n) : U \longrightarrow \mathbb{R}^n$ be a continuous vector. Then, a is suitable for performance measurement with r if and only if*

$$a_i(u) = \frac{\partial r(u)}{\partial u_i}$$

Proof: See Tasche (2000) [55] Theorem 4.4.

This is a core theorem showing that a suitable allocation formula is exactly a coherent fuzzy value (or Aumann-Shapley value) by comparing both Theorem 2 and Theorem 1, if the risk measure is a coherent one and with some differentiability assumptions.

4.3 Related Findings

A list of findings related to the previous sections is as follows:

- Aumann (1964, 1975) [8, 9] shows that, in a market with so-called continuum of traders, the core of such a market (exchange economy) coincides with the set of its "equilibrium allocations", i.e., allocations induced by a competitive equilibrium with an appropriate price structure. "Presumably, the results could be extended to economies with production", by Aumann (1964), although a serious mathematical proof on it has not been found in the literature.

- To follow up the portfolio selection problem related to the RORAC approach mentioned in the previous section, asset pricing in a coherent risk measure framework would be a natural extension on this topic. A starting point could be asset pricing in an ES_α framework. Since ES_α is an affine transformation of 1_{st} order MLPM, it can be a special case of Harlow and Rao (1989) [31]'s results which is asset pricing in a generalized MLPM framework. Since we are going to focus on the coherent risk management problems in this thesis, we will leave it for future work.

- Should a single risk measure, such as ES_α, be selected by the regulators? Or would it be better if every financial firm is allowed to use its own measure of risk (from the spectral measures of risk generated by ES_α or other coherent risk measures)? We have heard enough criticism on both proposals, and we are not going to answer questions like these in this dissertation.

Part III

Statistical Theory

Recall that the risk measures (VaR_α and ES_α) and the corresponding risk capital allocations that we are focusing on here, are all defined in terms of statistics, such as quantiles, moments or conditional moments of Profit/Loss (P/L). It is natural to study the statistical properties of the random P/L, such as the density functions, quantiles and moments etc. Given those statistical properties, we can carry out the calculation of the risk measures and risk capital allocations easily. For instance, if we know the PDF or CDF, we, for sure, can find the quantiles and the expectations analytically or numerically.

In the real business, none of the quantities are observable and we do not know the PDFs or CDFs, so feasible solutions must be based on inference from the observed P/L data, and on both economic models and statistical models. Statistical models will be our focus in this dissertation.

Estimating the joint P/L distribution, i.e. PDF or CDF, and the corresponding quantiles and moments, can be non-trivial statistical problems. Back-testing[3] is always necessary, in order to assess the model validity or the forecasting accuracy of the statistical models proposed.

In Chapter 5, we will review the current statistical modeling methods for P/L distribution: nonparametric, parametric, semi-parametric and semi-nonparametric

[3]We did not find a formal definition for it in the literature, since it is neither a statistical concept, nor a pure economic one. In practice, this concept has been well accepted in risk management for years. One sample definition from the American Stock Exchange is "The practice of applying a valuation or forecasting model to historical data to help appraise the model's possible usefulness when current and future data are used". We will discuss both back-testing and hypothesis testing in Chapter 6.

methods. We will discuss the corresponding estimators for ES_α and sensitivity of ES_α. We also give analytic results based on parametric distributional assumptions: multivariate t-distribution, two numerical estimators of which were proposed and compared. In Chapter 6, we will review the current back-testing methods based on VaR_α, discuss their statistical properties, and modify the back-testing procedures based on ES_α. A class of test statistics based on ES_α are proposed. Two typical test statistics are tabulated. Asymptotic properties are discussed.

Throughout this part, we make the following assumptions.

Assumption 1 *Loss Function*

- $a \in A$, choice vector belongs to a feasible set of portfolios A, satisfying imposed requirements. For instance, we may consider portfolios only with non-negative positions (short positions are not allowed).

- Y, random vector, random returns or risk factors, defined on a probability space $(\Omega, \mathcal{A}, P)$.

- $Z = f(a, Y)$, loss function, which is continuous in a, and measurable in Y, and that $E\{|f(a, Y)|\} < \infty$ for each $a \in A$. A typical example is the so-called linear portfolio

$$f(a, Y) = a^{'}Y,$$

Another example will be that

$$f(a, Y) = a^{'}g(Y, \beta),$$

where $g(Y, \beta)$ are vector functions, representing the random returns based on the pricing formula if the prices are not observable, with Y as the vector of risk factors and β as the vector of unknown parameters in the pricing formula. This is often quoted as non-linear portfolio in the literature.

Chapter 5

Statistical Modeling

The statistical modeling methods for P/L (or return) of portfolios, discussed in the following chapters, fall into four categories: nonparametric, parametric, semi-parametric and semi-nonparametric methods.

5.1 Nonparametric Methods

Without imposing any distributional assumption on the iid univariate sample of Z, one can estimate quantiles by the empirical distribution $\hat{F}_T$,

$$\hat{F}_T(z) = \frac{\sum_{i=1}^{T} \mathbf{I}_{\{Z_i \leq z\}}}{T}, \ z \in \mathbb{R}$$

where $\mathbf{I}$ is an indicator function.

By the Glivenko-Cantelli theorem, we know that with probability 1,

$$\sup_{z \in \mathbb{R}} |\hat{F}_T(z) - F(z)| \to 0, \quad T \to \infty$$

Consequently, we estimate

$$z_\alpha = F^{\leftarrow}(\alpha) \equiv \inf\{z \in \mathbb{R} : F(z) \geq \alpha\}$$

by

$$\hat{z}_{\alpha,T} = \hat{F}_T^{\leftarrow}(\alpha).$$

Let's define the ordered sample by

$$Z_{1,T} = \max_{0 \leq t \leq T}\{Z_t\} \geq Z_{2,T} \geq \cdots \geq Z_{T,T} = \min_{0 \leq t \leq T}\{Z_t\},$$

which are called order statistics. Then, for $k = 1, \cdots, T$,

$$\hat{z}_{\alpha,T} = \hat{F}_T^{\leftarrow}(\alpha) = Z_{k,T}, \quad 1 - \frac{k}{T} < \alpha \leq 1 - \frac{k-1}{T}$$

By the Central Limit Theorem (CLT), it is shown that

$$\hat{z}_{\alpha,T} \sim AN\left(z_\alpha, \frac{\alpha(1-\alpha)}{nf^2(z_\alpha)}\right) \tag{5.1}$$

where $f(\cdot)$ is a continuous density with $f(z_\alpha) \neq 0$ and $k = k(n)$ so that $n - k = n\alpha + o(n^{\frac{1}{2}})$. Also, see Embrechts (1996) [25] for some details.

If Z represents the random P/L, then the nonparametric estimator for VaR_α is

$$\widehat{\mathrm{VaR}_\alpha}(Z) = -\hat{z}_{\alpha,T}.$$

(5.1) can be used for sample size calculation, if a targeting accuracy is determined for VaR_α.

Similarly, we have the following estimator[1] for $\mathrm{ES}_\alpha(Z)$

$$\widehat{\mathrm{ES}_\alpha}(Z) = -\frac{\sum_{t=1}^T Z_t \mathbf{I}_{\{Z_t<\hat{z}_{\alpha,T}\}}}{\sum_{t=1}^T \mathbf{I}_{\{Z_t<\hat{z}_{\alpha,T}\}}} = -\hat{z}_{\alpha,T} + \frac{\sum_{t=1}^T (\hat{z}_{\alpha,T} - Z_t)^+}{\sum_{t=1}^T \mathbf{I}_{\{Z_t<\hat{z}_{\alpha,T}\}}}. \tag{5.2}$$

If $\widehat{\mathrm{VaR}_\alpha}(Z) = c$ is fixed, or a known quantity, then

$$\widehat{\mathrm{ES}_\alpha}(Z)|c \sim AN\left(\mathrm{ES}_\alpha(Z), \frac{\sigma^2(Z)}{n\alpha}\right). \tag{5.3}$$

If $\widehat{\mathrm{VaR}_\alpha}(Z) = -\hat{z}_{\alpha,T}$ is estimated from the empirical process, then

$$\begin{aligned}
\widehat{\mathrm{ES}_\alpha}(Z)|\hat{z}_{\alpha,T} &\sim AN\left(\mathrm{ES}_\alpha(Z), \frac{\sigma^2(Z)}{n\alpha}\right) \\
\widehat{\mathrm{ES}_\alpha}(Z) &= E_{\hat{z}_{\alpha,T}}\left(-\frac{\sum_{t=1}^T Z_t \mathbf{I}_{\{Z_t<\hat{z}_{\alpha,T}\}}}{\sum_{t=1}^T \mathbf{I}_{\{Z_t<\hat{z}_{\alpha,T}\}}}|\hat{z}_{\alpha,T}\right)
\end{aligned} \tag{5.4}$$

(5.3) and (5.4) can be used for sample size calculation depending on the situations, if a targeting accuracy is determined for ES_α. (5.4) needs to be estimated to be applicable to the real calculation.

Naturally, we have the following estimator for SES_α^i in a linear portfolio setup,

$$\widehat{\mathrm{SES}_\alpha^i}(a) = \frac{\sum_{t=1}^T Y_{i,t} \mathbf{I}_{\{Z_t<\hat{z}_{\alpha,T}\}}}{\sum_{t=1}^T \mathbf{I}_{\{Z_t<\hat{z}_{\alpha,T}\}}}.$$

Remark 2 Again, the above estimators are consistent estimators of the VaR_α, ES_α and SES_α by the Law of Large Numbers and Slutsky's theorem.

The advantage of this method is that it is straightforward and easy to implement, but it needs large sample size T to achieve accuracy, which can be partially illustrated

[1] Acerbi-Tasche (2002) [3] proposes $\lim_{n\to\infty} \frac{\sum_{k=1}^{[n\alpha]} Z_{k,T}}{[n\alpha]}$, to account for the discrete CDFs, which is equivalent to the estimator here for the continuous cases.

by the numerical examples furnished in the next section. Also, it follows from the definition that the empirical CDF, $\hat{F}_T$, is not a continuous function. Continuity is always a desirable property that we need for the calculation of quantiles and sensitivity of ES. The kernel method is natural and offers smooth estimators. Moreover, it allows for a broader class of data including dependent data satisfying strong mixing conditions (see Doukhan (1994) [23]). Scaillet (2002) in [52] proposed and studied the properties of an estimation method for ES_α and its sensitivity, based on a kernel approach. He also did simulation studies to compare parametric methods and semi-parametric methods, with the proposed kernel methods.

Remark 3 *Other Nonparametric Methods.*

As Scaillet (2002) in [52] pointed out, other nonparametric approaches, like spline methods, might be contemplated. We leave it as future work.

5.2 Parametric Methods

Classical distributional assumptions on Y seem to be violated all the time by empirical evidence of financial data. Meanwhile, the linear relationship of the portfolio weights and risk factors seems too simple to explain an arbitrary portfolio. Researchers try to generalize both the distributional assumptions on the P/L and restrictions on the loss functions. VaR_α for linear portfolios under student t-distribution is considered in the working papers by Albensese, Levin etc. Glasserman et. al (2002) in [30] proposes methods to compute the so-called "$\Gamma - \Delta$ VaR_α", for heavy-tailed non-linear portfolios, with focus on multivariate t-distribution. Kamdem (2004) in [34] generalizes the parametric linear portfolio methods to the elliptical distribution families, with special attention to multivariate t-distribution, which is very similar to our results under a totally independent regime.

The corresponding conditional CDF, PDF and VaR_α and/or ES_α are studied more or less along this line of research, but none of them ever considered SES_α. There are a few reasons why SES_α is under-addressed in the literature. First of all, research in risk measures are ongoing and puzzles remain to be solved. Comparing to risk measures, risk capital allocation seems a secondary problem. Secondly, the idea of coherent allocation is largely motivated by the idea of coherent risk measure recently, so that it does not catch enough attention yet. Thirdly, risk capital allocation meets the business needs of internal performance measurement, while risk capital calculation is more or less affected by the banking regulation, which always has a stronger signal directing the researchers, especially the industrial researchers. In this thesis, we try to consider SES_α whenever it is possible.

The elliptical distributional family will be revisited in this section. Special attention is given to multivariate t-distribution, while both Multivariate Gaussian and multivariate t-distribution are discussed as two special cases in this family. We keep the loss function linear, i.e.

$$f(a, Y) = a^{'}Y,$$

where Y can be viewed as price difference between the current value of the portfolio and an uncertain value of the portfolio at the next period, for instance.

5.2.1 Multivariate Gaussian

The classical assumption on Y is that $Y \sim N(\mu, \Sigma)$ (see [40]), where μ and Σ are either known or unknown parameters that need to be estimated. There are many ways of estimating μ and Σ, depending on the data quality. For example, Maximum Likelihood Estimators (MLE) or Moment Methods (MM) could be employed if we have enough iid samples of Y. If we don't have enough samples, we have to impose structures on μ or on Σ, and then estimate the structured models by using MLE, MM or Generalized Moment method (GMM), and so on.

Once we know μ, Σ or their feasible estimators ($\hat{\mu}$, $\hat{\Sigma}$), we can calculate $\mathrm{ES}_\alpha(Z)$ or its estimators. Notice that

$$Z \sim N(\mu_Z, \sigma_Z^2),$$

where $\mu_Z = a^{'}\mu$, and $\sigma_Z^2 = a^{'}\Sigma a$, so we have

$$\mathrm{ES}_\alpha(Z) = E\left[-Z \mid -Z > \mathrm{VaR}_\alpha(Z)\right],$$

where

$$\mathrm{VaR}_\alpha(Z) = -\mu_Z - \sigma_Z \Phi^{-1}(\alpha).$$

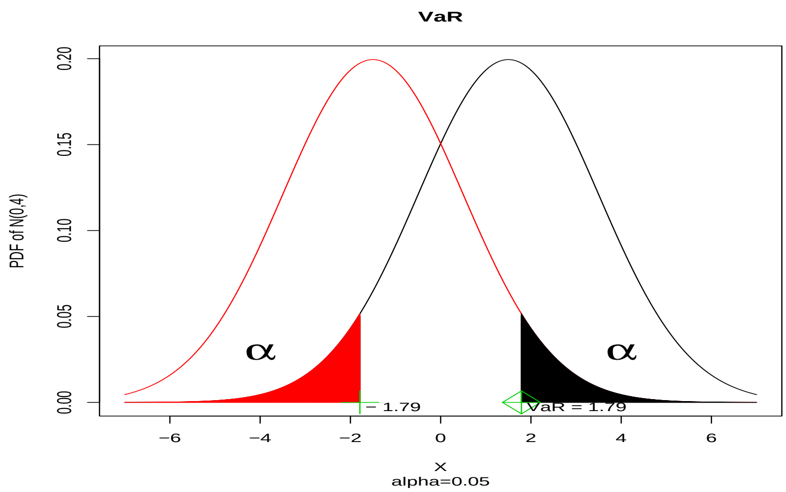

Figure 5.1: **Graphical illustration of VaR_α.**

Let us first find out the conditional CDF and conditional PDF,

$$\begin{aligned} F(t|-Z > \text{VaR}_\alpha(Z)) &= P(-Z \le t|-Z > \text{VaR}_\alpha(Z)) \\ &= \frac{P(\text{VaR}_\alpha(Z) < -Z \le t)}{P(-Z > \text{VaR}_\alpha(Z))} \\ &= \frac{1}{\alpha} P\left(\frac{\text{VaR}_\alpha(Z)+\mu_Z}{\sigma_Z} < \frac{-Z+\mu_Z}{\sigma_Z} \le \frac{t+\mu_Z}{\sigma_Z}\right) \\ &= \frac{1}{\alpha}\left[\Phi\left(\frac{t+\mu_Z}{\sigma_Z}\right) - \Phi\left(\frac{\text{VaR}_\alpha(Z)+\mu_Z}{\sigma_Z}\right)\right] \end{aligned}$$

Thus,

$$f(t|\mathrm{VaR}_\alpha(Z)) = \frac{1}{\alpha\sigma_Z}\phi\left(\frac{t+\mu_Z}{\sigma_Z}\right)$$

where $t > \mathrm{VaR}_\alpha(Z)$, Φ and ϕ are CDF and PDF of standard Gaussian random variables respectively.

Now, let's look at the expected shortfall,

$$\begin{aligned}
\mathrm{ES}_\alpha(Z) &= E[-Z| - Z - \mathrm{VaR}_\alpha(Z) > 0] \\
&= \int_{\mathrm{VaR}_{\alpha(Z)}}^{\infty} tf(t|\mathrm{VaR}_\alpha(Z))dt \\
&= \frac{1}{\alpha\sigma_Z}\int_{\mathrm{VaR}_{\alpha(Z)}}^{\infty} t\phi\left(\frac{t+\mu_Z}{\sigma_Z}\right)dt \\
&= -\mu_Z + \frac{\sigma_Z}{\alpha\sqrt{2\pi}}\exp\left\{-\frac{[\Phi^{-1}(\alpha)]^2}{2}\right\}. \qquad (5.5)
\end{aligned}$$

We denote the sensitivity of ES_α by SES_α or K_i^{AS} (Aumann-Shapley value, see [22] for details). Thus, the corresponding SES_αs are

$$K_i^{AS} = \mathrm{SES}_\alpha^i(a) = -\mu_i + \frac{a'\Sigma_i}{\alpha\sqrt{2\pi\sigma_Z^2}}\exp\left\{-\frac{[\Phi^{-1}(\alpha)]^2}{2}\right\}, \qquad i = 1, \cdots, n, \qquad (5.6)$$

where μ_i is the i^{th} element of μ and Σ_i is the i^{th} column of Σ.

Scaillet (2002) [52] gives some similar results.

5.2.2 Multivariate t-distribution

Multivariate t When the P/L exhibits the symmetric fat-tailed property, multivariate $t(\mu, \Sigma, v)$ family can be very useful to accommodate this with the extra freedom of v, which is a natural generalization of multivariate Gaussian assumption. Assuming $Y \sim t(\mu, \Sigma, v)$, we will give analytic solutions for VaR_α, ES_α and SES_α of Z. Without losing generality, we assume $\mu = 0$.

Recall that $Z = a^{'}Y$ and $\mu = 0$, so we have

$$Z \sim t(0, a^{'}\Sigma a, v).$$

Thus

$$\text{VaR}_\alpha(Z) = \sqrt{a^{'}\Sigma a}\,\text{VaR}_\alpha(t_v) \equiv \lambda_0, \tag{5.7}$$

$$\begin{aligned}
\text{ES}_\alpha(Z) &= E(-Z| - Z \geq \lambda_0) \\
&= \sqrt{a^{'}\Sigma a} E\left(-t_v| - t_v \geq \frac{\lambda_0}{\sqrt{a^{'}\Sigma a}}\right) \\
&= \sqrt{a^{'}\Sigma a} \int_{\frac{\lambda_0}{\sqrt{a^{'}\Sigma a}}}^{\infty} t \frac{1}{\alpha} f_{t_v}(t) dt
\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{a'\Sigma a}\frac{v\Gamma((v+1)/2)}{2\alpha\Gamma(v/2)\sqrt{v\pi}}\int_{\frac{\lambda_0}{\sqrt{a'\Sigma a}}}^{\infty}(1+v^{-1}t^2)^{-\frac{v+1}{2}}d(1+v^{-1}t^2) \\
&= \begin{cases} \sqrt{a'\Sigma a}\,\frac{v\Gamma((v+1)/2)}{2\alpha\Gamma(v/2)\sqrt{v\pi}}\frac{2}{1-v}(1+v^{-1}t^2)^{\frac{1-v}{2}}\Big|_{\frac{\lambda_0}{\sqrt{a'\Sigma a}}}^{\infty} & \text{when } v \geq 2, \\ \infty & \text{when } v = 1 \end{cases} \\
&= \begin{cases} \frac{1}{\alpha(v-1)}\frac{\Gamma((v+1)/2)}{\Gamma(v/2)}\sqrt{\frac{a'\Sigma a v}{\pi}}\left(1+\frac{\lambda_0^2}{va'\Sigma a}\right)^{\frac{1-v}{2}}, & \text{when } v \geq 2, \\ \infty & \text{when } v = 1 \end{cases}
\end{aligned}
\tag{5.8}
$$

since

$$
Pr(-t_v \leq t| - t_v \geq \frac{\lambda_0}{\sqrt{a'\Sigma a}}) = F\left(t| - t_v \geq \frac{\lambda_0}{\sqrt{a'\Sigma a}}\right) = \frac{1}{\alpha}\left[F_{t_v}(t) - F_{t_v}\left(\frac{\lambda_0}{\sqrt{a'\Sigma a}}\right)\right]
$$

and

$$
f\left(t| - t_v \geq \frac{\lambda_0}{\sqrt{a'\Sigma a}}\right) = \frac{1}{\alpha}f_{t_v}(t).
$$

By definition, we know

$$
\frac{\lambda_0^2}{a'\Sigma a} \equiv [\mathrm{VaR}_\alpha(t_v)]^2,
$$

thus,

$$\mathrm{ES}_\alpha(Z) = \begin{cases} \frac{1}{\alpha(v-1)} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)} \sqrt{\frac{a^{'}\Sigma a v}{\pi}} \left(1 + \frac{[\mathrm{VaR}_\alpha(t_v)]^2}{v}\right)^{\frac{1-v}{2}}, & \text{when } v \geq 2; \\ \infty & \text{when } v = 1 \end{cases} \tag{5.9}$$

and

$$K_i^{AS} = \mathrm{SES}_\alpha^i(a) = \begin{cases} \frac{1}{\alpha(v-1)} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)} \sqrt{\frac{v}{\pi a^{'}\Sigma a}} \left(1 + \frac{[\mathrm{VaR}_\alpha(t_v)]^2}{v}\right)^{\frac{1-v}{2}} a^{'}\Sigma_i, & \text{when } v \geq 2; \\ \infty & \text{when } v = 1 \end{cases} \tag{5.10}$$

Remark 4 *Proportional Invariance*

As you may notice, the proportion of K_i^{AS} over $\mathrm{ES}_\alpha(Z)$, which is the proportion of the institution's risk capital allocated to its i_{th} business unit, is invariant for the multivariate t-distribution and multivariate Gaussian distribution families, given the same dispersion matrix Σ and allocation parameter a with mean $\mu = 0$, i.e.

$$\left.\frac{\mathrm{SES}_\alpha^i(a)}{\mathrm{ES}_\alpha}\right| a, \Sigma = \frac{a^{'}\Sigma_i}{a^{'}\Sigma a},$$

This capital allocations strategy has been proposed for the Gaussian families. As we may notice, we can extend its usefulness to other elliptical families. Multivariate t-distribution is certainly one example serving such a purpose. We will show that this invariance holds good for a general elliptical family in the later discussions on elliptical families.

Notice that this invariance works only when mean $\mu = 0$. In the case of non-zero mean elliptical family, there is no such relationship.

Numerical Examples When an analytic capital allocation formula is infeasible for a more general case, a numerical solution by Monte Carlo simulation would be an option. We will illustrate the basic idea by using the following two Monte Carlo simulation methods for the multivariate t-distribution.

Example 2 *Naive Monte Carlo Simulation* We simulate bivariate-t with degree of freedom $v = 1, 2, 3, 5, 20, 100$, correlation $\gamma = -0.7, -0.3, 0, 0.3, 0.7, 0.9, 1$, $\sigma_1 : \sigma_2 = 1, 2, 4$, by Monte Carlo Simulations.

Based on a standard univariate Normal random number generator, we can generate bivariate Normal $(\mathbf{0}, \Sigma)$ in the following steps:

- Covariance matrix Σ can be decomposed into the following form:

$$\Sigma = DRD^T,$$

  where $D = diag(\sigma_1, \sigma_2)$ and set $\sigma_2 = 1$ for convenience; R is the correlation matrix.

- $R^{\frac{1}{2}}$ can be obtained by Cholesky decomposition, given R positive definite as a correlation matrix. Or denoted by L,

$$R = LL^T = R^{\frac{1}{2}} {R^{\frac{1}{2}}}^T.$$

Thus,

$$L = R^{\frac{1}{2}} = \begin{bmatrix} (1-\gamma^2)^{1/2} & \gamma \\ 0 & 1 \end{bmatrix}.$$

- Generate independent random vectors $X_n = (X_{1,n}, X_{2,n})^T$ from two standard Normal distributions independently , with sample size $s = 10^5$ [2], i.e. $n = 1, \cdots, 10^5$.

- Transfer the computer-generated random vectors, X_n, to new vectors called $K_n = (K_{1,n}, K_{2,n})^T$, by

$$K_n = DLX_n, \text{ for } n = 1, \cdots, s.$$

  By construction we know that K_n's are random samples of bivariate Normal $(\mathbf{0}, \Sigma)$.

Next, we generate s random samples of χ^2_v, denoted by $\chi^2_{v,1}, \cdots, \chi^2_{v,10^5}$. Each of the random samples is the summation of v independent χ^2_1's, each of which is essentially the square of a random number from the standard Normal generator.

Thus the random samples of bivariate-t, denoted by $T_n = (T_{1,n}, T_{2,n})^T$, with sample size s, is

$$T_n = \sqrt{\frac{v}{\chi^2_{v,n}}} K_n, \quad n = 1, \cdots, s.$$

[2]For the time being, we fix the sample size. This sample size should roughly give us one digit accuracy, 95% of the chance, for the VaR_α based on (5.1). A sample size calculation can be fulfilled by (5.1), (5.3) or (5.4) based on the targeting precision.

Consider the simplest portfolio Z with two assets and the allocation vector $a = (1,1)^T$.

$$Z = a^T T = T_1 + T_2,$$

with $T \sim t\left(\mathbf{0}, \mathbf{\Sigma}, v\right)$. So

$$Z|\chi_v^2 \sim N(0, \sigma^2_{Z|\chi_v^2}),$$

with $\sigma^2_{Z|\chi_v^2} = a^T \mathbf{\Sigma} a \frac{v}{\chi_v^2}$. We will use the nonparametric formula developed in the previous section, (5.2) and (**??**), for this simulation study.

The naive simulation results for ES_α and SES_α are summarized in the following tables:

Table 5.1: **VaR_α, ES_α, and SES^1_α of Bivariate-t - Naive Simulation vs Analytic Solution.** *Empirical estimation based on sample size* $= 10^5$, $\alpha = 0.05$, $\sigma_1 = 1$ *with parameters corr* $= \gamma$, σ_2, $df = 2$ *specified in the table.*

			Naive Simulation vs Analytic Solution					
df	γ	σ_2	VaR_α	ES_α	SES^1_α	VaR_α	ES_α	SES^1_α
2	-0.9	1	2.82702	7.11876	3.35048	1.30586	2.75681	1.37841
		2	4.83713	11.51646	0.76384	3.45497	7.29383	-4.16791
		4	10.20423	23.46255	-0.58342	9.14100	19.29767	-5.11979
	-0.7	1	3.38093	7.81721	3.69541	2.26181	4.77494	2.38747
		2	5.48056	12.43721	1.88603	4.33104	9.14330	-1.66242
		4	10.59547	23.68708	-0.00877	9.85900	20.81346	-3.28634
	-0.3	1	4.10647	8.50414	4.34927	3.45497	7.29383	3.64692
		2	6.32773	12.88704	2.40955	5.69210	12.01666	1.26491
		4	11.72435	24.81110	1.27985	11.15725	23.55419	-0.32266
	0.0	1	4.58684	8.98488	4.51464	4.12948	8.71780	4.35890
		2	6.90351	14.03528	3.18383	6.52929	13.78405	2.75681
		4	12.41771	26.89350	2.48252	12.03941	25.41653	1.49509
	0.3	1	4.88248	9.57295	4.80503	4.70834	9.93982	4.96991
		2	7.56035	14.64409	4.08626	7.27071	15.34927	3.96110
		4	13.12461	26.96529	3.09080	12.86120	27.15143	3.07903
	0.7	1	5.51755	10.36205	5.11979	5.38419	11.36662	5.68331
		2	8.27827	15.65736	4.38732	8.15508	17.21627	5.29731
		4	13.91033	27.57889	4.06255	13.88145	29.30529	4.92744
	0.9	1	5.80718	11.30743	5.80076	5.69210	12.01666	6.00833
		2	8.65141	15.94276	4.78300	8.56308	18.07761	5.88573
		4	14.34322	27.92321	4.04436	14.36443	30.32491	5.76424
	1.0	1	5.88662	10.57245	5.34716	5.83997	12.32883	6.16441
		2	8.71044	16.29826	5.26095	8.75996	18.49324	6.16441
		4	14.54384	28.64696	4.39694	14.59993	30.82207	6.16441

Table 5.2: **$\mathbf{VaR}_\alpha$, $\mathbf{ES}_\alpha$, and $\mathbf{SES}^1_\alpha$ of Bivariate-t - Naive Simulation vs Analytic Solution.** *Empirical estimation based on sample size* $= 10^5$, $\alpha = 0.05$, $\sigma_1 = 1$ *with parameters corr* $= \gamma$, σ_2, *df* $= 5$ *specified in the table.*

			Naive Simulation vs Analytic Solution					
df	γ	σ_2	$\mathbf{VaR}_\alpha$	$\mathbf{ES}_\alpha$	$\mathbf{SES}^1_\alpha$	$\mathbf{VaR}_\alpha$	$\mathbf{ES}_\alpha$	$\mathbf{SES}^1_\alpha$
5	-0.9	1	1.29345	2.04515	1.037508	0.90116	1.29251	0.646253
		2	2.64554	4.14844	-1.129877	2.38424	3.41965	-1.954084
		4	6.54200	9.84666	-1.632288	6.30810	9.04754	-2.400367
	-0.7	1	1.79362	2.60518	1.316703	1.56085	2.23868	1.119342
		2	3.18709	4.71309	-0.447564	2.98880	4.28675	-0.779410
		4	6.97105	10.28833	-1.102782	6.80359	9.75820	-1.540768
	-0.3	1	2.48902	3.46490	1.759983	2.38424	3.41965	1.709823
		2	4.00908	5.70013	0.580196	3.92805	5.63390	0.593042
		4	7.71940	11.02925	-0.100075	7.69949	11.04317	-0.151276
	0.0	1	2.91922	3.98875	1.977382	2.84971	4.08726	2.043630
		2	4.57484	6.34587	1.186806	4.50579	6.46253	1.292505
		4	8.39640	12.00963	0.528878	8.30826	11.91631	0.700959
	0.3	1	3.27174	4.48406	2.228702	3.24917	4.66019	2.330097
		2	5.05244	7.00123	1.633988	5.01743	7.19636	1.857126
		4	8.90957	12.58639	1.136527	8.87537	12.72970	1.443574
	0.7	1	3.71114	5.06174	2.514974	3.71557	5.32914	2.664567
		2	5.62383	7.74599	2.209218	5.62772	8.07169	2.483597
		4	9.48709	13.26037	1.891794	9.57943	13.73952	2.310184
	0.9	1	3.88742	5.27771	2.626265	3.92805	5.63390	2.816949
		2	5.82114	8.00815	2.443764	5.90928	8.47552	2.759472
		4	9.79349	13.67108	2.265169	9.91273	14.21756	2.702511
	1.0	1	3.97507	5.39705	2.684645	4.03010	5.78026	2.890129
		2	5.97345	8.17806	2.553336	6.04515	8.67039	2.890129
		4	9.97166	13.91291	2.425069	10.07524	14.45065	2.890129

Table 5.3: **$\mathbf{VaR}_\alpha$, $\mathbf{ES}_\alpha$, and $\mathbf{SES}^1_\alpha$ of Bivariate-t - Naive Simulation vs Analytic Solution.** *Empirical estimation based on sample size* $= 10^5$, $\alpha = 0.05$, $\sigma_1 = 1$ *with parameters* $corr = \gamma$, σ_2, $df = 20$ *specified in the table.*

			Naive Simulation vs Analytic Solution					
df	γ	σ_2	$\mathbf{VaR}_\alpha$	$\mathbf{ES}_\alpha$	$\mathbf{SES}^1_\alpha$	$\mathbf{VaR}_\alpha$	$\mathbf{ES}_\alpha$	$\mathbf{SES}^1_\alpha$
20	-0.9	1	0.853394	1.114224	0.558604	0.771317	0.993630	0.496815
		2	2.100403	2.831680	-1.351137	2.040714	2.628899	-1.502228
		4	5.422632	7.118274	-1.675602	5.399222	6.955413	-1.845314
	-0.7	1	1.375782	1.761377	0.898356	1.335961	1.721018	0.860509
		2	2.593318	3.380344	-0.581433	2.558171	3.295499	-0.599182
		4	5.836389	7.642724	-1.106875	5.823319	7.501745	-1.184486
	-0.3	1	2.055897	2.623281	1.319540	2.040714	2.628899	1.314449
		2	3.389919	4.351740	0.428904	3.362095	4.331134	0.455909
		4	6.596714	8.509006	-0.142986	6.590139	8.489582	-0.1162956
	0.0	1	2.439581	3.116013	1.564941	2.439120	3.142135	1.571068
		2	3.868815	4.961395	0.942733	3.856587	4.968152	0.993630
		4	7.129247	9.188534	0.480545	7.111195	9.160820	0.538872
	0.3	1	2.789031	3.544602	1.761537	2.781025	3.582585	1.791293
		2	4.299005	5.517483	1.365590	4.294514	5.532300	1.427690
		4	7.609010	9.797629	1.038416	7.596596	9.786124	1.109767
	0.7	1	3.172202	4.045928	2.021936	3.180223	4.096843	2.048421
		2	4.798645	6.127796	1.839591	4.816876	6.205220	1.909298
		4	8.161366	10.479750	1.668947	8.199217	10.562440	1.775985
	0.9	1	3.344419	4.274013	2.133355	3.362095	4.331134	2.165567
		2	5.050745	6.443723	2.056111	5.057867	6.515670	2.121381
		4	8.455688	10.822150	1.973010	8.484492	10.929930	2.077591
	1.0	1	3.429921	4.364765	2.179136	3.449436	4.443650	2.221825
		2	5.133964	6.561685	2.156548	5.174155	6.665475	2.221825
		4	8.590916	11.009510	2.126935	8.623591	11.109130	2.221825

Table 5.4: **$\mathbf{VaR}_\alpha$, $\mathbf{ES}_\alpha$, and $\mathbf{SES}^1_\alpha$ of Bivariate-t - Naive Simulation vs Analytic Solution.** *Empirical estimation based on sample size* $= 10^5$, $\alpha = 0.05$, $\sigma_1 = 1$ *with parameters corr* $= \gamma$, σ_2, $df = 100$ *specified in the table.*

			Naive Simulation vs Analytic Solution					
df	γ	σ_2	$\mathbf{VaR}_\alpha$	$\mathbf{ES}_\alpha$	$\mathbf{SES}^1_\alpha$	$\mathbf{VaR}_\alpha$	$\mathbf{ES}_\alpha$	$\mathbf{SES}^1_\alpha$
100	-0.9	1	0.757838	0.961644	0.480384	0.742479	0.935835	0.467917
		2	1.977907	2.518946	-1.393556	1.964416	2.475986	-1.414849
		4	5.185186	6.567978	-1.724627	5.197356	6.550843	-1.737979
	-0.7	1	1.295618	1.633379	0.842846	1.286012	1.620913	0.810457
		2	2.474370	3.133913	-0.565497	2.462525	3.103813	-0.564330
		4	5.611616	7.113155	-1.117617	5.605596	7.065397	-1.115589
	-0.3	1	1.972478	2.486291	1.263070	1.964416	2.475986	1.237993
		2	3.246317	4.102897	0.422296	3.236393	4.079209	0.429390
		4	6.354220	8.038726	-0.130398	6.343746	7.995775	-0.109531
	0.0	1	2.350950	2.963797	1.483654	2.347926	2.959369	1.479685
		2	3.725140	4.704259	0.923139	3.712397	4.679174	0.935835
		4	6.837930	8.650717	0.485829	6.845321	8.627970	0.507528
	0.3	1	2.684601	3.378638	1.684791	2.677047	3.374200	1.687100
		2	4.129657	5.206411	1.329709	4.133950	5.210507	1.344647
		4	7.288198	9.224603	1.007975	7.312574	9.216903	1.045216
	0.7	1	3.064494	3.865036	1.928132	3.061321	3.858545	1.929273
		2	4.639251	5.845365	1.783768	4.636782	5.844286	1.798242
		4	7.883717	9.969116	1.646972	7.892664	9.948060	1.672683
	0.9	1	3.229999	4.082028	2.035925	3.236393	4.079209	2.039604
		2	4.858529	6.123875	1.980229	4.868763	6.136679	1.997988
		4	8.148111	10.312080	1.934722	8.167273	10.294180	1.956745
	1.0	1	3.312290	4.184024	2.091375	3.320469	4.185180	2.092590
		2	4.959696	6.274408	2.086203	4.980703	6.277770	2.092590
		4	8.277086	10.479860	2.081550	8.301172	10.462950	2.092590

The above simulation study based on a sample size of 10^5 shows that the naive simulation tend to overestimate VaR_α, ES_α and SES_α when γ is negative, and to underestimate them when γ increases. The precision seems increasing with the degree of freedom by comparing the four tables. In general, we will need a bigger sample size to achieve a more precise result for this naive simulation with the empirical estimation.

Example 3 *Accelerated Monte Carlo Simulation* When computational resource is limited, an accelerated simulation method is in demand. There are various kinds of accelerated methods in the computing literature. We will give one example to illustrate the basic idea.

$$\begin{aligned} \alpha &= E(I_{\{-Z\geq\lambda\}}) \\ &= E_{\chi_v^2}[E(I_{\{-Z\geq\lambda\}}|\chi_v^2)]. \end{aligned} \tag{5.11}$$

Consider

$$\begin{aligned}
E(I_{\{-Z\geq\lambda\}}|\chi_v^2) &= P(Z \leq -\lambda|\chi_v^2 = c) \\
&= P\left(\frac{Z}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/c}} \leq \frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/c}}|\chi_v^2 = c\right) \\
&= \Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/c}}|\chi_v^2 = c\right)
\end{aligned} \tag{5.12}$$

Thus, an estimator of VaR_α, λ_N, can be decided by solving the following equation[3] for λ:

$$F_N(\lambda) \equiv \frac{1}{N}\sum_{n=0}^{N}\Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/\chi_{v,n}^2}}\right) \quad = \quad \alpha \tag{5.13}$$

where $\chi_{v,n}^2$ are random samples of χ_v^2, with known v. We know that

$$F_N(\lambda_N) \equiv \alpha.$$

Since

$$F_{N+1}(\lambda) = \frac{N}{N+1}F_N(\lambda) + \frac{1}{N+1}\Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/\chi_{v,N+1}^2}}\right),$$

[3] We use equal weighting here as an example for simplicity. We can also use other weighting mechanisms, such as Kernel method or χ_v^2 weighting.

and

$$F_{N+1}(\lambda_{N+1}) \equiv \alpha$$

we know that

$$F_{N+1}(\lambda_N) \;\; = \;\; \frac{N}{N+1}\alpha + \frac{1}{N+1}\Phi\left(\frac{-\lambda_N}{\sqrt{\mathbf{a^T \Sigma a} v/\chi^2_{v,N+1}}}\right)$$

Thus,

$$\begin{aligned} F_{N+1}(\lambda_{N+1}) - F_{N+1}(\lambda_N) = F'_{N+1}(\lambda_N)(\lambda_{N+1} - \lambda_N) + o(|\lambda_{N+1} - \lambda_N|) \\ \alpha - \tfrac{N}{N+1}\alpha - \tfrac{1}{N+1}\Phi\left(\tfrac{-\lambda_N}{\sqrt{\mathbf{a^T \Sigma a} v/\chi^2_{v,N+1}}}\right) = F'_{N+1}(\lambda_N)(\lambda_{N+1} - \lambda_N) + o(|\lambda_{N+1} - \lambda_N|) \\ \tfrac{1}{N+1}[\alpha - \Phi\left(\tfrac{-\lambda_N}{\sqrt{\mathbf{a^T \Sigma a} v/\chi^2_{v,N+1}}}\right)] = F'_{N+1}(\lambda_N)(\lambda_{N+1} - \lambda_N) + o(|\lambda_{N+1} - \lambda_N|) \\ \lambda_{N+1} = \lambda_N + \tfrac{1}{N+1}\frac{\alpha - \Phi\left(\tfrac{-\lambda_N}{\sqrt{\mathbf{a^T \Sigma a} v/\chi^2_{v,N+1}}}\right)}{F'_{N+1}(\lambda_N)} + o(|\lambda_{N+1} - \lambda_N|) \end{aligned} \tag{5.14}$$

for $N = 0, 1, \cdots, s$.

Notice that we have to assume $F'_{N+1}(\lambda_N) \neq 0$ to make (5.14) work. This assumption is also critical to the following discussion on the asymptotic properties of F_N and λ_N. However, this condition holds, by the definition of $F_N(\cdot)$, (5.13). Further, in order to smooth our later discussion, we assume there exists $M \in \mathbb{R}^+$, such that $|F'_{N+1}(\lambda_N)| > M$ almost surely, for any N.

We will use the iterative steps to find λ_N,

$$\lambda_{N+1} = \lambda_N + \frac{1}{N+1} \frac{\alpha - \Phi\left(\frac{-\lambda_N}{\sqrt{\mathbf{a}^\mathbf{T}\boldsymbol{\Sigma}\mathbf{a}v/\chi^2_{v,N+1}}}\right)}{F'_{N+1}(\lambda_N)} \tag{5.15}$$

and we will stop at the step where the target accuracy is achieved, i.e.

- λ_{N+1} is close enough to λ_∞, where λ_∞ is the true VaR_α i.e. $F_N(\lambda) \overset{N\to\infty}{\longrightarrow} F_\infty(\lambda)$ for any λ and $F_\infty(\lambda_\infty) \equiv \alpha$. This is almost infeasible, when nothing is known about the true VaR_α before hand.

- λ_{N+1} is close enough to λ_N. In our case, the stopping rule is even easier. Since it is not difficult to show that

$$\frac{\alpha - \Phi(x)}{F'_{N+1}(\lambda_N)} \leq \frac{\max(\alpha, 1-\alpha)}{M},$$

for some $M \in \mathbb{R}^+$ and any $x \in \mathbb{R}$, i.e.

$$|\lambda_{N+1} - \lambda_N| \leq \frac{\max(\alpha, 1-\alpha)}{M} \frac{1}{N+1}.$$

So it is suffice to set any $N \geq \frac{\max(\alpha,1-\alpha)}{\epsilon M} - 1$ for a pre-specified $\epsilon \in \mathbb{R}^+$ by the risk managers.

- $F_\infty(\lambda_N)$ falls into the specified confidence interval for α.

Any of the above criteria could be reasonable to decide the sample size. It is up to the risk manager to choose. As you may notice, λ is in the dollar amount, while $F_\infty(\lambda_N)$ is a scale free quantity, probability of loss exceed λ_N.

In order to make the stopping rules explicit, we will discuss the asymptotic properties of $F_N(\lambda)$ and λ_N in the following paragraphs.

By construction and the combined results from (5.11) and (5.12), we know that $F_N(\lambda)$ is an unbiased estimator of $E_{\chi_v^2}[E(I_{\{-Z\geq\lambda\}}|\chi_v^2)]$ for any $\lambda \in \mathbb{R}$, since

$$\begin{aligned} E[F_N(\lambda)] &= \frac{1}{N}\sum_{n=0}^{N} E\left[\Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/\chi_{v,n}^2}}\right)\right] \\ &= E\left[\Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/\chi_v^2}}\right)\right]. \end{aligned}$$

If the variance of $\Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/\chi_v^2}}\right)$ exists, then, by Central Limit Theorem,

$$\sqrt{N}\left\{F_N(\lambda) - E\left[\Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/\chi_v^2}}\right)\right]\right\} \xrightarrow{\mathbf{D}} N\left(0, \mathrm{Var}\left[\Phi\left(\frac{-\lambda}{\sqrt{\mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}\mathbf{a}v/\chi_v^2}}\right)\right]\right) \tag{5.16}$$

for any $\lambda \in \mathbb{R}$.

Then,

$$\begin{aligned} \alpha \equiv \quad & F_N(\lambda_N) = F_N(\lambda_\infty) + F_N^{'}(\lambda_\infty)(\lambda_N - \lambda_\infty) + o(|\lambda_N - \lambda_\infty|) \\ \Longrightarrow \qquad & \lambda_N - \lambda_\infty = \frac{\alpha - F_N(\lambda_\infty)}{F_N^{'}(\lambda_\infty)} + o(|\lambda_N - \lambda_\infty|) \end{aligned}$$

By the Slutsky theorem, the Central Limit Theorem and (5.16), we know that

$$\sqrt{N}(\lambda_N - \lambda_\infty) \xrightarrow{\mathbf{D}} N\left(0, \frac{\mathrm{Var}\left[\Phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T \Sigma a} v/\chi_v^2}}\right)\right]}{(F_\infty^{'}(\lambda_\infty))^2}\right) \tag{5.17}$$

where $\lambda_\infty = \mathrm{VaR}_\alpha$ and

$$F_\infty^{'}(\lambda_\infty) = E\left[\phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T \Sigma a} v/\chi_v^2}}\right)\frac{1}{\sqrt{\mathbf{a^T \Sigma a} v/\chi_v^2}}\right]$$

Given the confidence level $\beta = 0.05$, the confidence interval for λ_∞ is

$$\left[\lambda_N - z_{\beta/2}\sqrt{\frac{\mathrm{Var}\left[\Phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T \Sigma a} v/\chi_v^2}}\right)\right]}{N(F_\infty^{'}(\lambda_\infty))^2}}, \lambda_N + z_{\beta/2}\sqrt{\frac{\mathrm{Var}\left[\Phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T \Sigma a} v/\chi_v^2}}\right)\right]}{N(F_\infty^{'}(\lambda_\infty))^2}}\right]$$

By the Delta method, we know that

$$\sqrt{N}[F_\infty(\lambda_N) - F_\infty(\lambda_\infty)] \xrightarrow{\mathbf{D}} N\left(0, \mathrm{Var}\left[\Phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T \Sigma a} v/\chi_v^2}}\right)\right]\right)$$

where $F_\infty(\lambda_\infty) = \alpha$, i.e.

$$\sqrt{N}[F_\infty(\lambda_N) - \alpha] \xrightarrow{\mathbf{D}} N\left(0, \mathrm{Var}\left[\Phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T\Sigma a}v/\chi_v^2}}\right)\right]\right).$$

Given the confidence level $\beta = 0.05$, the confidence interval for $F_\infty(\lambda_N)$ is

$$\left[\alpha - z_{\beta/2}\sqrt{\mathrm{Var}\left[\Phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T\Sigma a}v/\chi_v^2}}\right)\right]/N}, \alpha + z_{\beta/2}\sqrt{\mathrm{Var}\left[\Phi\left(\frac{-\lambda_\infty}{\sqrt{\mathbf{a^T\Sigma a}v/\chi_v^2}}\right)\right]/N}\right]$$

Do we know λ_∞ and F_∞? The answer is yes, since for the multivariate t-distribution λ_∞ is known due to the formula (5.7), and thus F_∞ is just a probability that the loss is larger than λ_∞.

Once λ_N is decided, ES_α and SES_α can be estimated either by the empirical estimation or by the parametric formula.

5.2.3 Elliptical Family

Both multivariate Gaussian and multivariate t-distribution are special cases of the elliptical family. To be consistent with Anderson (1993) [4], we define the elliptical family by $Y \sim El_n(\mu, \Sigma; h)$, with density function of the form

$$|\Sigma|^{-\frac{1}{2}} h^n((y-\mu)^{'}\Sigma^{-1}(y-\mu)),$$

where h^n satisfies

$$\int_0^\infty \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\mu^{\frac{n}{2}-1}h^n(\mu)d\mu = 1.$$

By theory, we know that any linear combination of an elliptically distributed random vector is elliptical, i.e.

$$Z \quad = \quad a^{'}Y \sim El_n(a^{'}\mu, a^{'}\Sigma a; h).$$

Thus,

$$\mathrm{VaR}_\alpha(Z) = -a^{'}\mu + q^h_{\alpha,n}\sqrt{a^{'}\Sigma a},$$

with $q^h_{\alpha,n}$ the $\mathrm{VaR}_\alpha(El_n(0, I; h))$; and

$$\mathrm{ES}_\alpha = -a^{'}\mu + K^h_{ES}\sqrt{a^{'}\Sigma a},$$

$$\mathrm{SES}^i_\alpha = -\mu_i + K^h_{ES}\frac{a^{'}\Sigma_i}{\sqrt{a^{'}\Sigma a}},$$

with K^h_{ES} the $\mathrm{ES}_\alpha(El_n(0, I; h))$. Some special cases of the elliptical family have analytic expression for $q^h_{\alpha,n}$ and K^h_{ES}, such as multivariate Gaussian and Multivariate t-distribution, but some not, depending the function h.

As an example, $s = q^h_{\alpha,n}$ can be a unique positive solution to the equation by Kamdem (2004) [34]

$$\alpha = \frac{\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\int_s^{-\infty}\int_{z^2}^{+\infty}(x - z^2)^{\frac{n-3}{2}}h(x)dxdz;$$

and

$$K^h_{ES} = \frac{\pi^{\frac{n-1}{2}}}{\alpha\Gamma\left(\frac{n+1}{2}\right)}\int_{(q^h_{\alpha,n})^2}^{\infty}(x - (q^h_{\alpha,n})^2)^{\frac{n-1}{2}}h(x)dx.$$

5.3 Semi-parametric Methods

Since VaR_α, ES_α and SES_α are all about the tail properties of the P/L distributions, it is natural to model P/L distributions by using the so-called semi-parametric approaches, in the sense that a full parametric model (Generalized Pareto Distribution or GPD, for instance, see [52]) is used to model the tail of the distribution, with the rest unspecified or partially specified.

A flexible model is based on a so-called point process characterization, resulting in Peak Over Threshold (POT) methods, which considers exceedances over a threshold μ. Mathematical theory (see Leadbetter (1991)) supports the condition of a possibly inhomogeneous Poisson process with intensity λ for the number of exceedances combined with independent excesses $W = Y - \mu$ over the threshold with μ given. $W \sim$ GPD (μ, σ, κ), Generalized Pareto Distribution (GPD) with location parameter μ, scale parameter σ and shape parameter κ. See [32] for detailed investigation of estimation methods for the GPD. Scaillet (2002) [52] fitted GPD to the data above threshold by a method of moments (MM), in the case when threshold μ is known.

The smoothing extreme value method fitted by penalized log-likelihood was proposed by Chavez-Demoulin and Embrechts (2001) in [18].

Remark 5 *Relevant Research along This Line.*

McNeil (1997) in [41] summarized the relevant EVT theoretical results and provided an extensive example of their application to Danish data on large fire insurance losses.

5.4 Semi-nonparametric Methods

Semi-nonparametric method (SNP) naturally seems to be another alternative to offer smooth estimators. Gallant and Nychka (1987) in [29] offered a feasible way to estimate P/L densities and their moments, derivatives and integrals. We will propose a SNP approach for the purpose of estimating ES_α and its sensitivity. We leave this for future work.

Chapter 6

Back-testing

Financial institutions have extensively increased their use of statistical/econometrics models to manage their market risk exposure over the past decades, partially because of their increased trading activities, increased emphasis on RAROC and advances in both the theoretical and empirical research. Financial regulators also have begun to focus on the use of such models by regulated institutions.

Beginning in 1998, U.S. commercial banks may determine their regulatory economic/risk capital for financial market risk exposure using VaR_α models, i.e., models of the time-varying distributions of portfolio returns. VaR_α estimates are forecasts of the maximum portfolio loss that could occur over a given holding period with a specified confidence level α[1], while ES_α estimates are forecast of the average portfolio loss given that the loss is no less than the threshold of VaR_α, over the same

[1] The term "confidence level" here is used synonymously with "probability", instead of the widely used term in statistics. It sounds ambiguous, but we hope readers can identify this difference in different contexts.

given holding period. Under the amended capital rules, regulated banks' market risk capital charges can be based on the VaR_α estimates generated by their own VaR_α models. From our previous discussion, we know that ES_α is a coherent risk measure while VaR_α is not, for the purpose of calculating capital charges.

Given the importance of VaR_α estimates to the banks and their regulators, evaluating the model validity or forecast accuracy is always necessary. A variety of model selection methods or testing methods can be found in econometrics and statistics literatures to evaluate models. Different methods can be utilized to access the models from different perspectives. As Lopez (1999) in [35] pointed out, regulatory evaluation differs from institutional evaluations, because:

1. The goal of regulatory evaluation of models is to assure that sufficient risk capital is available to prevent an institution from a big portfolio losses, which may not be shared by an institutional evaluation due to issues of moral hazard.

2. Regulators generally can not evaluate every component of the model and its implementation as well as the originating institution can, although they are allowed to acquire the details of an institution's model.

3. Comparable evaluations across institutions are more desirable to regulators than to institutions.

For financial risk models, "Back-testing" is a process of comparing daily, monthly, quarterly or yearly profits and losses with model-generated risk measures to gauge the quality and accuracy of the institutions' internal risk measurement systems. In this evaluation process, the detection of the under-estimated risk measurement systems is always the main interest, since an over-estimated risk measurement system tends to be a conservative system which will not unstabilized the whole economy. When a reasonable hypothesis about the interested quantities - risk measures, such as VaR_α and ES_α - can be established, statistical hypothesis testing can be a very useful framework to validate the models. We will be focusing on the hypothesis testing methods that are potentially suitable as regulatory approaches to validate ES_α models in this dissertation, although other approaches might be appropriate as well.

6.1 Hypothesis Testing

6.1.1 Hypothesis Testing based on VaR_α

Three hypothesis-testing methods based on VaR_α are available to regulators (see Lopez, 1997, [35] for a review): the binomial method, the interval forecast method (see Christoffersen, 1998, [19]), and the distribution forecast method.

The back-testing discussion will be based on the following assumptions,

Assumption 2 *Back-testing*

- Z_t: portfolio value at time t in dollar terms, and define $z_t = \ln Z_t$,
- $\boldsymbol{\epsilon_{t+k}} = z_{t+k} - z_t$: k-period-ahead return in percentage, a random variable with

continuous true PDF f_{t+k}, and CDF F_{t+k}. $\boldsymbol{\epsilon_{t+k}}$s are log returns of the particular asset, i.e. $\boldsymbol{\epsilon_{t+k}} = \ln \frac{Z_{t+k}}{Z_t}$. In order to distinguish the forecasted log returns from the observed log returns, we will use the unbolded $\{\epsilon_t\}$'s for the observed values.

- Ω_t: information available at time t, which typically includes, but is not limited to, the entire history of the time series of the log returns up to $\boldsymbol{\epsilon_t}$.

- $\boldsymbol{\epsilon_{t+k}}|\Omega_t \sim f_{t+k,t}$, where $f_{t+k,t}$[2] is the conditional PDF of $\boldsymbol{\epsilon_{t+k}}$ given information Ω_t and $F_{t+k,t}$ is the corresponding conditional CDF,

- k: holding period, $k = 1$, or 10, is the typical case,

- α: confidence level, $\alpha = 1\%$ or 5%, is the typical case,

- $\mathrm{VaR}_\alpha^t(\boldsymbol{\epsilon_{t+k}}) = F_{t+k,t}^{\leftarrow}(\alpha)$: VaR_α expressed in percentage,

- $Z_t(1 - \exp\{\mathrm{VaR}_\alpha(\boldsymbol{\epsilon_{t+k}})\})$: VaR_α expressed in dollar,

- $\mathrm{ES}_\alpha^t(\boldsymbol{\epsilon_{t+k}}) = \frac{1}{\alpha}\int_0^\alpha VaR_\mu(\boldsymbol{\epsilon_{t+k}})d\mu$: ES_α expressed in percentage,

- $Z_t(1 - \exp\{\mathrm{ES}_\alpha(\boldsymbol{\epsilon_{t+k}})\})$: ES_α expressed in dollar,

- T: sample size[3], $T = 250$ is the typical case, i.e. one trading year.

The current available approaches are,

[2] It is assumed continuous for the time being in order to avoid the difficulty in defining the ES_α later.

[3] Sample size T is for estimation, and $T + k$ for back-testing.

1. Binomial Method: Based on the null hypothesis that the VaR_α estimates are accurate against the alternative that they are not, in the sense that the unconditional coverage is at least α ,

$$H_0 : E(I_t(\alpha)) \leq \alpha \quad \text{vs} \quad H_a : E(I_t(\alpha)) > \alpha.$$

where

$$\mathbf{I}_t(\alpha) = \begin{cases} 1 & \text{when } -\epsilon_{t+k}|\Omega_t \geq \text{VaR}_\alpha(\boldsymbol{\epsilon_{t+k}}), \\ 0 & \text{when } -\epsilon_{t+k}|\Omega_t < \text{VaR}_\alpha(\boldsymbol{\epsilon_{t+k}}). \end{cases} \qquad t = 1, \cdots, T.$$

Notice that if the VaR_α estimates are accurate in all sense,

$$\mathbf{I}_t(\alpha) \overset{iid}{\sim} \text{Bernoulli}(\alpha), \quad t = 1, \cdots, T.$$

thus a Wald test, Score test or a likelihood Ratio test (exact or asymptotic) can be appropriate.

2. Interval Forecast Method: Interval forecast methods are proposed to check both the conditional coverage and unconditional coverage (or dependence Heteroscedasticity), aiming to diagnose the clustered outliers. See Christoffersen (1998) [19] for details.

3. Distribution Forecast Method: The so-called Kuiper statistics (see Kupiec (1995) [36] for details) or Kolmogorov-Smirnoff Test based on the transformation defined by Rosenblatt (1952) in [46].

6.1.2 Hypothesis Testing Based on ES_α

Next, we will propose a test procedure based on the ES_α transformed through the null CDF, opposing to the methods based on VaR_α. Our motivation is that the accuracy of the forecasted magnitude of large losses is of particular interest to both the financial institutions and regulators. They don't want to reject a model that forecast the conditional/unconditional ES_α well, but fail to match the small day-to-day moves that characterize the interior of the forecast. Rosenblatt (1952) in [46] plays an important role in constructing the test statistics as in Kupiec (1995) [36].

Proposed Test Statistics and Two Examples

The hypothesis testing idea is as follows,

- $F_{t+k,t},\quad t = 1, \cdots, T$ are the forecasted conditional CDFs of $\boldsymbol{\epsilon_{t+k}}$, based on the historical data, such as historical log returns. $\{\epsilon_{t+k}\}_{t=1}^{T}$ are observed log returns.
- Null hypothesis is that $F_{t+k,t}$s' are the accurate (or true) conditional CDFs for ϵ_{t+k}, then by the argument of Rosenblatt (1952) in [46],

$$\gamma_{t+k} \;\equiv\; F_{t+k,t}(\epsilon_{t+k}) \;\stackrel{iid}{\sim}\; \mathrm{Uniform}(0,1), \quad \forall t = 1, \cdots, T,$$

- Test statistic that we propose,

$$X^{\alpha}_{T,W} = \frac{1}{\alpha}\int_0^{\alpha} \hat{F}^W_T(z)dz$$

$$= \frac{1}{T\alpha}\sum_{t=1}^{T}(W(\alpha) - W(\gamma_{t+k}))^+, \quad (6.1)$$

  where

$$\hat{F}^W_T(z) = w(z)\frac{1}{T}\sum_{t=1}^{T}\mathbf{I}_{\{W(\gamma_{t+k})\leq W(z)\}},$$

  a *empirical cumulative distribution function (ECDF)* of $\boldsymbol{\gamma}$ with a scaling/weighting function $w(\cdot) = W'(\cdot) \geq 0$. $X^+ = \max(X, 0)$. Notice that $W(\cdot)$ is required to be a monotone non-decreasing function with $|W(x)| < \infty$, $x \in [0, \alpha]$. $(W(\alpha) - W(\gamma_{t+k}))^+$'s are independent and identical distributed (iid) under the null hypothesis.

- By playing with different weighting functions, we can put different weights on the tails. We are going give two simple weighting functions later, and discuss the properties of the corresponding test statistics.

- Rules of the test: When the alternative hypothesis is

$$H_a : F_{t+k,t}s \text{ are under-estimated in tails,}$$

  we reject the hypothesis if $X^{\alpha}_{T,W} > c_{\beta}$, where c_{β} is some critical value with an confidence level $\beta \in (0, 1)$.

When the alternative hypothesis is

$$H_a : F_{t+k,t}s \text{ are over-estimated in tails,}$$

we reject the hypothesis if $X^{\alpha}_{T,W} < c_{\beta}$, where c_{β} is some critical value with an confidence level $\beta \in (0, 1)$.

Equal Weighting By equal weighting, we mean that $w(x) = 1$ and $W(x)$ is an identical transformation, $x \in (0, 1]$, thus

$$\hat{F}^E_T(z) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}_{\{\gamma_{\mathbf{t+k}} \leq \mathbf{z}\}},$$

and

$$X^{\alpha}_{T,E} \quad = \quad \frac{1}{T\alpha}\sum_{t=1}^{T}(\alpha - \gamma_{t+k})^{+}. \tag{6.2}$$

Recall that $m_t = (\alpha - \gamma_{t+k})^{+}$s are iid with the PDF of mixture of a mass, $1 - \alpha$, at 0 and a Uniform(0, 1) on $(0, \alpha]$, so that we can omit the suffix t and denote the CDF as

$$F(m) = \Pr(M \leq m) = \begin{cases} 0 & \text{when } m < 0, \\ 1 - \alpha & \text{when } m = 0, \\ 1 - \alpha + m & \text{when } 0 < m \leq \alpha, \\ 1 & \text{when } m > \alpha. \end{cases}$$

If the sample size T is reasonably large, we can estimate the critical values for the test by an approximation. One example will be the following asymptotic normal approximation.

We notice that

$$E(M) = \frac{\alpha^2}{2} \leq \frac{1}{2}, \quad \alpha \in (0, 1];$$

and

$$\mathrm{VaR}_\alpha(M) = E(M^2) - [E(M)]^2 = \frac{\alpha^3}{3} - \frac{\alpha^4}{4} \leq \frac{1}{12}, \quad \alpha \in (0, 1].$$

By Central Limit Theory, we have

$$\sqrt{T}\left(X^{\alpha}_{T,E} - \frac{\alpha}{2}\right) \xrightarrow{\mathbf{d}} N\left(0, \frac{\alpha}{3} - \frac{\alpha^2}{4}\right). \tag{6.3}$$

For a large sample size T, the result of (6.3) can be used to construct the critical values and confidence intervals for the test statistics (6.2):

$$c_\beta = \frac{\alpha}{2} + \frac{1}{\sqrt{T}} z_\beta \sqrt{\frac{\alpha}{3} - \frac{\alpha^2}{4}},$$

when β is the specified the confidence level, T and α are given, and z_β is the critical value of the standard normal distribution.

$T = 250$ is a typical case; the above value may not be very close to the true one (A Monte Carlo study will illustrate this point later). Another approximation is considered here based on a Poisson distribution in hope of doing a better job.

Notice that we can rewrite the test statistic (6.2) as

$$X^{\alpha}_{T,E} = \frac{1}{T}\sum_{t=1}^{N} U_t. \qquad (6.4)$$

where $N \sim$ Binomial(T, α) and $U_t \overset{iid}{\sim}$ Uniform$(0, 1)$. N and U_t's are independent.

The characteristic function (c.f.) of $X^{\alpha}_{T,E}$ is

$$\begin{aligned}
\phi(t) &= E(e^{itX^{\alpha}_{T,E}}) \\
&= \int_0^1 e^{itx} d[F(x)] \\
&= E(e^{it\frac{1}{T}\sum_{t=1}^{N} U_t}) \\
&= E[E(e^{it\frac{1}{T}\sum_{t=1}^{N} U_t})|N] \\
&= E[(\frac{T}{it})^N (e^{\frac{it}{T}} - 1)^N |N]
\end{aligned} \qquad (6.5)$$

where $F(x)$ is the CDF of $X^{\alpha}_{T,E}$.

Poisson $(T\alpha)$ can be a good approximation to N ~Binomial (T, α), with PDF

$p_n = Pr(N = n) = e^{-T\alpha}\frac{(T\alpha)^n}{n!}$. Thus, the characteristic function (c.f.) of the Poisson approximation, denoted by $\phi_P(t)$, is

$$\begin{aligned}
\phi(t) \approx \phi_P(t) &= \sum_{n=0}^{\infty}(\frac{T}{it})^n(e^{\frac{it}{T}} - 1)^n e^{-T\alpha}\frac{(T\alpha)^n}{n!} \\
&= \sum_{n=0}^{\infty}\left[\frac{T^2\alpha}{it}(e^{\frac{it}{T}} - 1)\right]^n \frac{e^{-T\alpha}}{n!} \\
&= e^{-T\alpha + T\alpha\frac{T}{it}\left(e^{\frac{it}{T}} - 1\right)} \\
&= e^{-T\alpha\left(1 - \frac{T}{it}\left(e^{\frac{it}{T}} - 1\right)\right)}
\end{aligned} \tag{6.6}$$

When the maximum sample size is T, the use of the truncated Poisson $(T\alpha)$ at T might be reasonable, that is

$$\phi(t) \approx s(T)\sum_{n=0}^{T}\left[\frac{T^2\alpha}{it}(e^{\frac{it}{T}} - 1)\right]^n \frac{e^{-T\alpha}}{n!} \tag{6.7}$$

where $s(T) = \frac{e^{T\alpha}}{\sum_{n=0}^{T}\frac{(T\alpha)^n}{n!}}$.

The exact (Binomial) c.f. is

$$\phi(t) \;=\; \sum_{n=0}^{T} (\frac{T}{it})^n (e^{\frac{it}{T}} - 1)^n p_n \tag{6.8}$$

where $p_n = \Pr(N = n) = \binom{T}{n} \alpha^n (1-\alpha)^{T-n}$.

Thus

$$\begin{aligned} \phi(t) \;=\; E(e^{itX^{\alpha}_{T,E}}) &= (1-\alpha)^T \sum_{n=0}^{T} \binom{T}{n} \left(\frac{\alpha}{1-\alpha}\right)^n \left(\frac{T}{it}\right)^n (e^{\frac{it}{T}} - 1)^n \\ &= \left[1 - \alpha + \alpha\frac{T}{it}\left(e^{\frac{it}{T}} - 1\right)\right]^T, \end{aligned} \tag{6.9}$$

For any given sample size T, we know that there is a mass p at 0, which contributes discontinuity to the whole distribution functions of the proposed statistics. Let $F_c(x)$ be the continuous part of the distribution function defined on $(0,1]$ and $\phi_c(t)$ be the corresponding characteristic function; then we have

$$\begin{aligned} F(x) &= p + [1-p]F_c(x)I_{[x>0]} \\ \Longrightarrow \quad \phi(t) &= p + [1-p]\phi_c(t) \\ \Longrightarrow \quad \phi_c(t) &= \frac{\phi(t) - p}{1-p}. \end{aligned} \tag{6.10}$$

The mass p and the corresponding characteristic functions of different methods generated by (6.10), are listed in the following table.

Table 6.1: **CF for Continuous Part – Equal Weighting.** *Characteristic functions of the corresponding continuous distributions with different methods.*

	$\phi(t)$	p	$\phi_c(t)$
Binomial	$\left[1-\alpha+\alpha\frac{T}{it}(e^{\frac{it}{T}}-1)\right]^T$	$(1-\alpha)^T$	$\frac{\left[1-\alpha+\alpha\frac{T}{it}(e^{\frac{it}{T}}-1)\right]^T-(1-\alpha)^T}{1-(1-\alpha)^T}$
Poisson	$e^{-T\alpha[1-\frac{T}{it}(e^{\frac{it}{T}}-1)]}$	$e^{-T\alpha}$	$\frac{e^{-T\alpha[1-\frac{T}{it}(e^{\frac{it}{T}}-1)]}-e^{-T\alpha}}{1-e^{-T\alpha}}$
Truncated-P	$S(T)\sum_0^T\left[\frac{T^2\alpha}{it}(e^{\frac{it}{T}}-1)\right]^n\frac{e^{-T\alpha}}{n!}$	$S(T)e^{-T\alpha}$	$\frac{S(T)\sum_0^T\left[\frac{T^2\alpha}{it}(e^{\frac{it}{T}}-1)\right]^n\frac{e^{-T\alpha}}{n!}-S(T)e^{-T\alpha}}{1-S(T)e^{-T\alpha}}$

We will use the following relationship to remove the discontinuity and focus on the numerical methods only on continuous part of the distribution.

Given the characteristic functions of Binomial - exact, Poisson and Truncated Poisson methods for proposed test statistic, we can numerically estimate the distribution function by evaluating the Gil-Palaez inversion integral

$$F(x) \;=\; \frac{1}{2} - \int_{-\infty}^{\infty} \frac{\phi(t)}{2\pi i t} e^{itx} dt \tag{6.11}$$

by numerical quadrature. Fourier series are truncated for numerical computation. The fast Fourier transform (FFT) may be used to speed up the calculation.

Bohman (1975)[15] and Davies (1973) [20] originally proposed algorithms to numerically invert characteristic functions to obtain cumulative distribution functions. The results are generalized and extended to multivariate random variables by Shephard (1991) [53]. Davis (1973) obtains error bounds for the sampling and truncation error, but the bounds are not easily expressed in terms of the characteristic function. Hughett (1998) extends the previous work by obtaining error bounds for computing the probability density function and the cumulative distribution function from the characteristic function based on the decaying speeds of characteristic functions and density functions. We adopt the Huggett's approach to generate a numerical version of the CDF estimates, thus find the critical values for the test[4]. The detailed procedures are provided as follows.

Remark 6 *Numerical Inversion of a Characteristic Function*

[4]We will leave the error bounds discussion for the future work.

- Characteristic Function is obtained by

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$

- Construct an auxiliary function $f_c(x)$ which is well behaved in the sense that both it and its Fourier transform decay rapidly to zero and from which it is easy to compute and approximation to $F(x)$ over some selected interval $[-D/2, D/2]$. By Hughett (1998) [33], we choose

$$f_c(x) = F(x) - \frac{1}{2}F(x - D) - \frac{1}{2}F(x + D);$$

thus $F(x) \cong f_c(x) + \frac{1}{2}$. Observed that the right hand side of the equation approaches $F(x)$ as D increases for any fixed x, and $f_c(x)$ approaches 0 as $x \to \pm\infty$ for any fixed D.

- $g(x)$ is a lowpass-filtered approximation of $F(x)$ on $[-L/2, L/2]$ based on $f_c(x)$ by

$$g(x) = \sum_{k=-(N/2)}^{N/2} G[k] e^{i2\pi kx/L}$$

where $G[k]$ is the finite sequence defined by

$$G[k] = \begin{cases} \frac{1}{2} & \text{for } k = 0, \\ \frac{1-\cos(2\pi\lambda k)}{i2\pi k}\phi(-2\pi k/L) & \text{for } 0 < |k| < N/2, \\ 0 & \text{for } |k| = N/2 \end{cases}$$

where $\lambda = D/L$ is the fraction of the interval (-T/2, T/2) over which $g(x)$ accurately approximates the $F(x)$

- By Inverse Discrete Fourier transformation with R function `fft()` we get discrete approximation $\{g[k]\}_{k=0}^{N}$ of $g(x)$, where N is power of 2 in order to speed up the computing by FFT.

- Linear interpolation[5] is utilized to make the discrete point more useful for the critical value calculation at confidence level $\beta = 0.05, 0.10$. If there exists an n, $0 \leq n \leq N$, such that $g[n] \leq 1 - \beta$ and $g[n+1] > 1 - \beta$, then c_β can be achieved by

$$c_\beta \approx \left(\frac{1 - \beta - g[n]}{g[n+1] - g[n]} + n \right) \frac{L}{N}$$

A comparison of their performances can be viewed in the following graphs and tables.

[5]A band-limited interpolation method is proposed in the Hughett [33], but it is too ideal to implement. Linear interpolation is doing reasonably well, here.

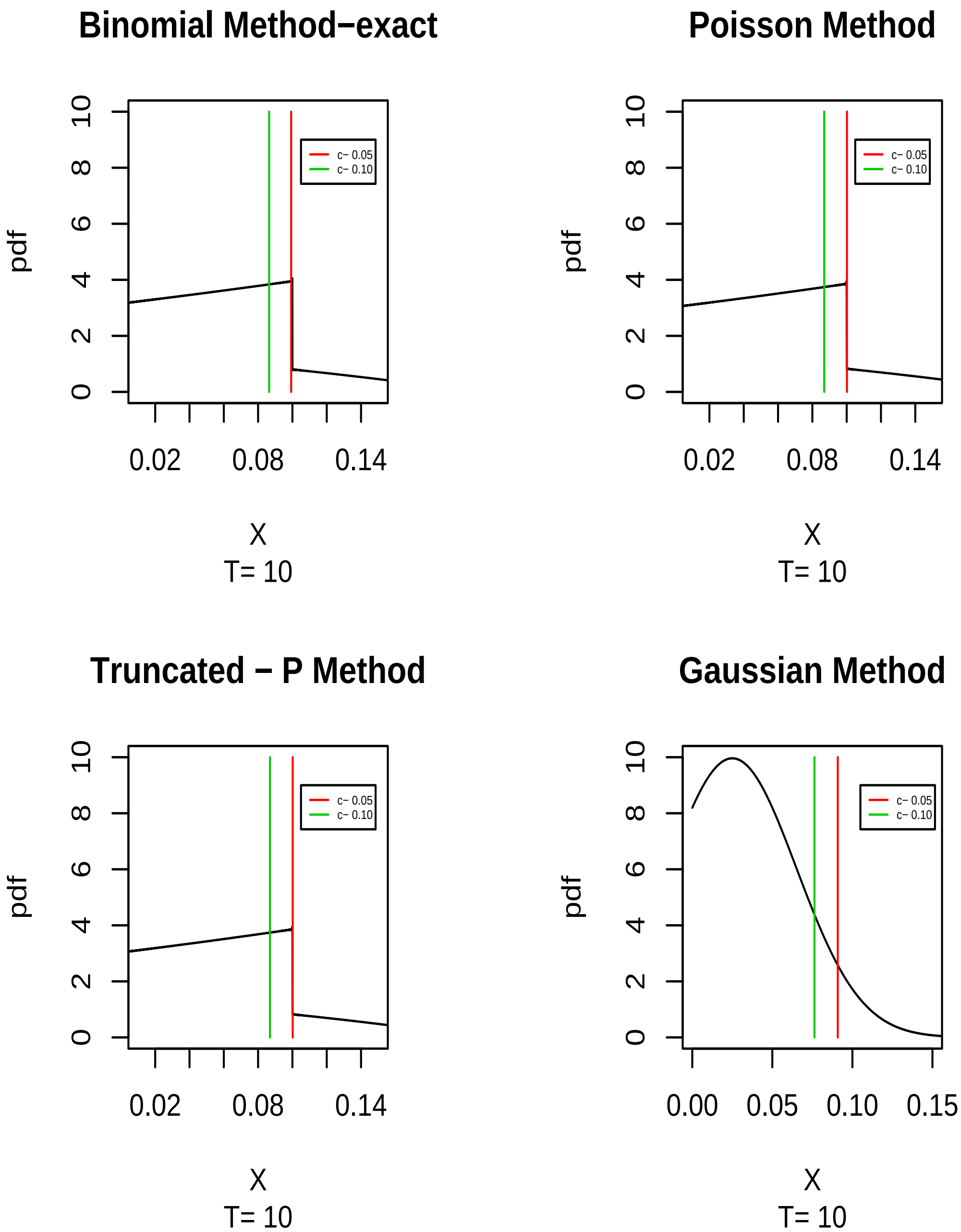

Figure 6.1: **Probability Density Functions – Equal Weighting.** $T = 10$, $\alpha = 0.05$

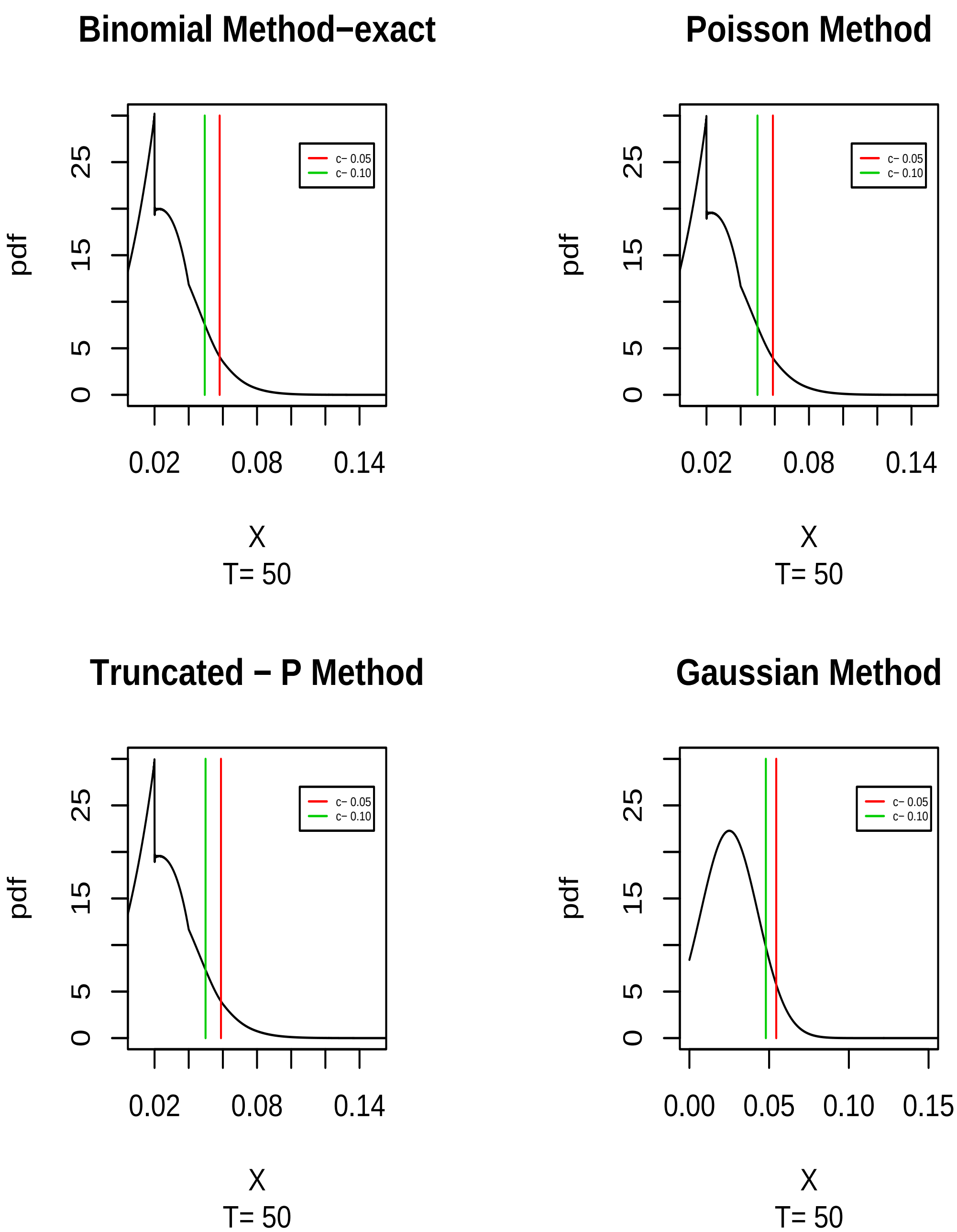

Figure 6.2: **Probability Density Functions – Equal Weighting.** $T = 50$, $\alpha = 0.05$

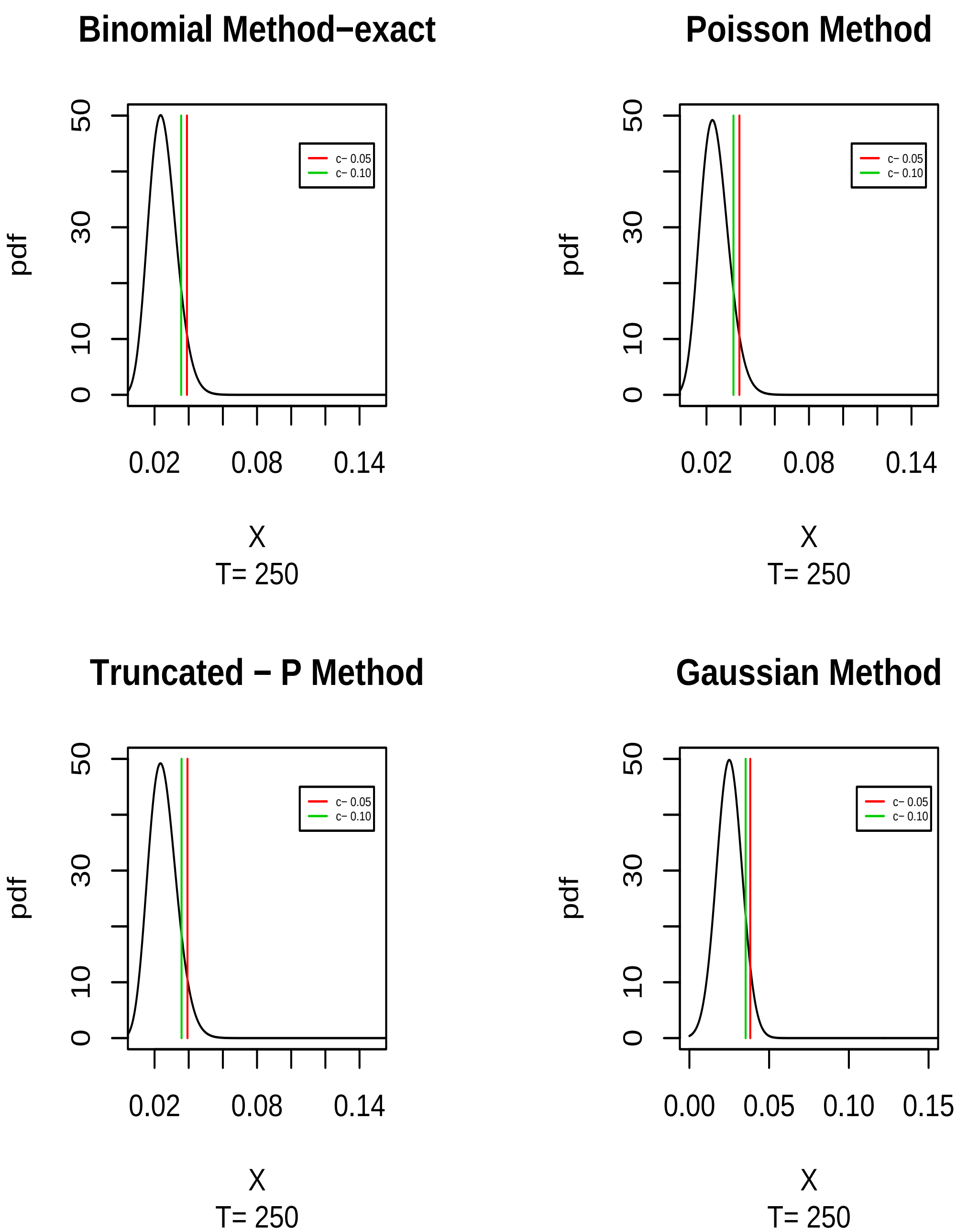

Figure 6.3: **Probability Density Functions – Equal Weighting.** $T = 250$, $\alpha = 0.05$

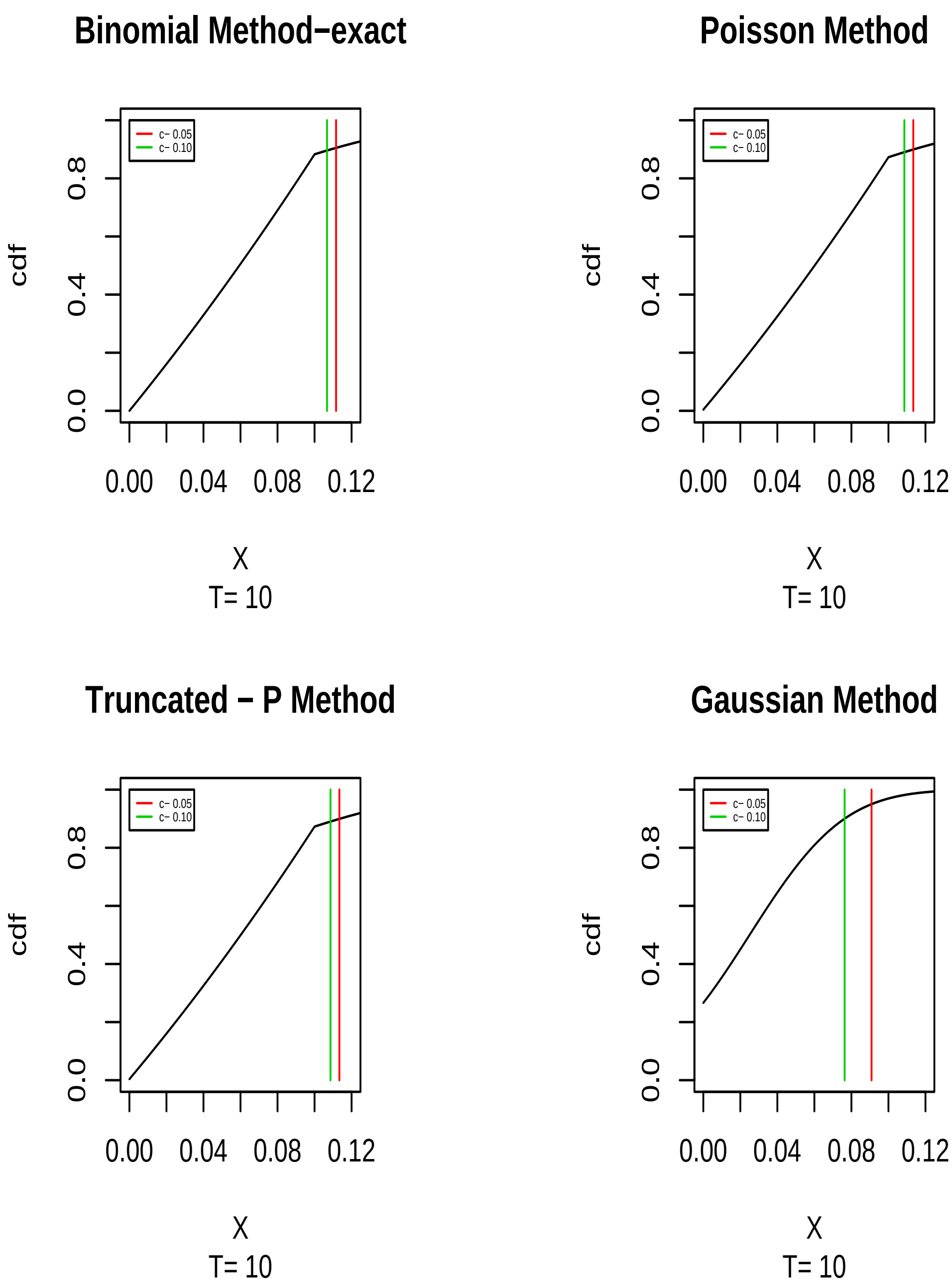

Figure 6.4: **Cumulative Distributional Functions – Equal Weighting.** $T = 10$, $\alpha = 0.05$

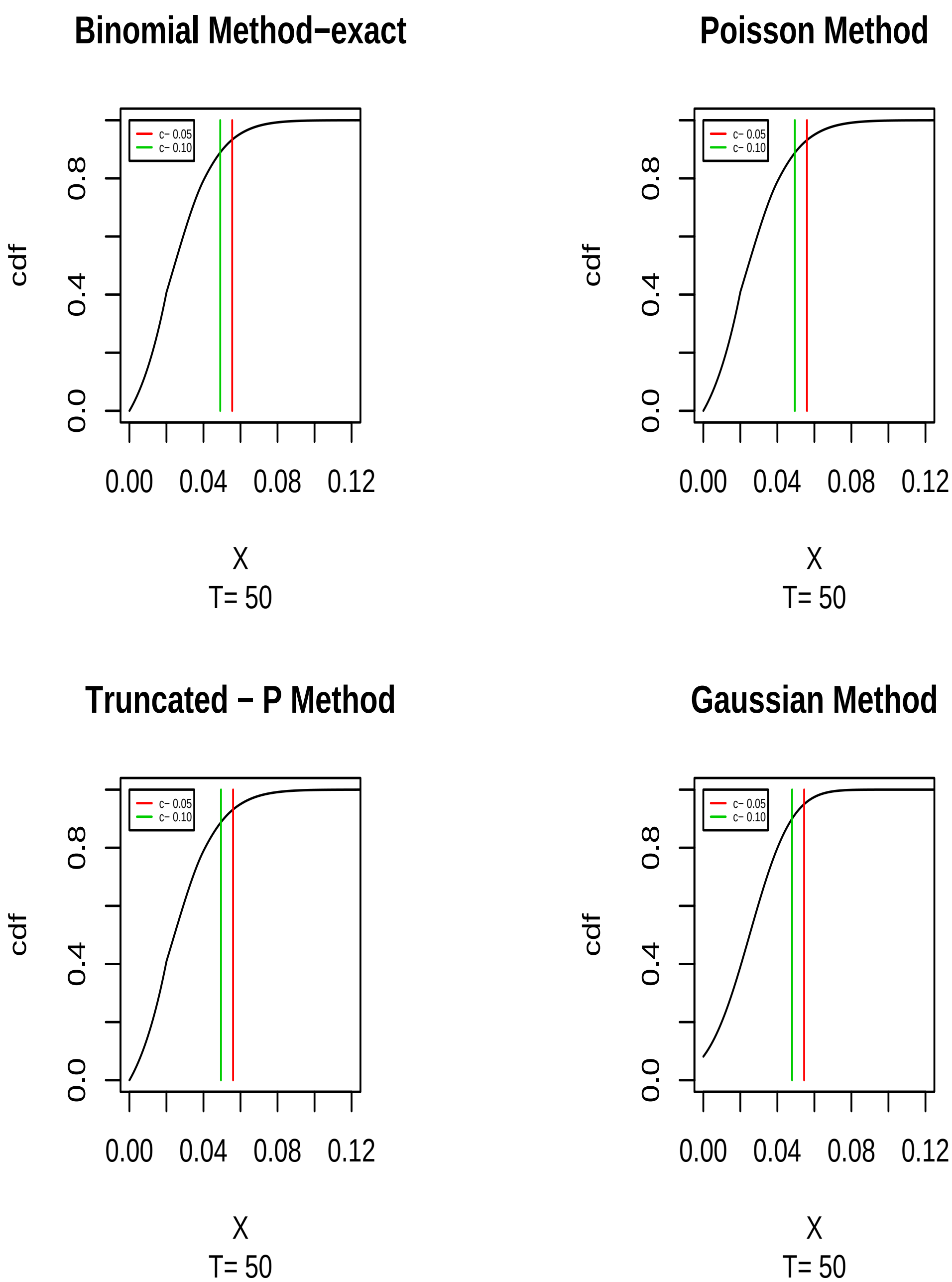

Figure 6.5: **Cumulative Distributional Functions – Equal Weighting.** $T = 50$, $\alpha = 0.05$

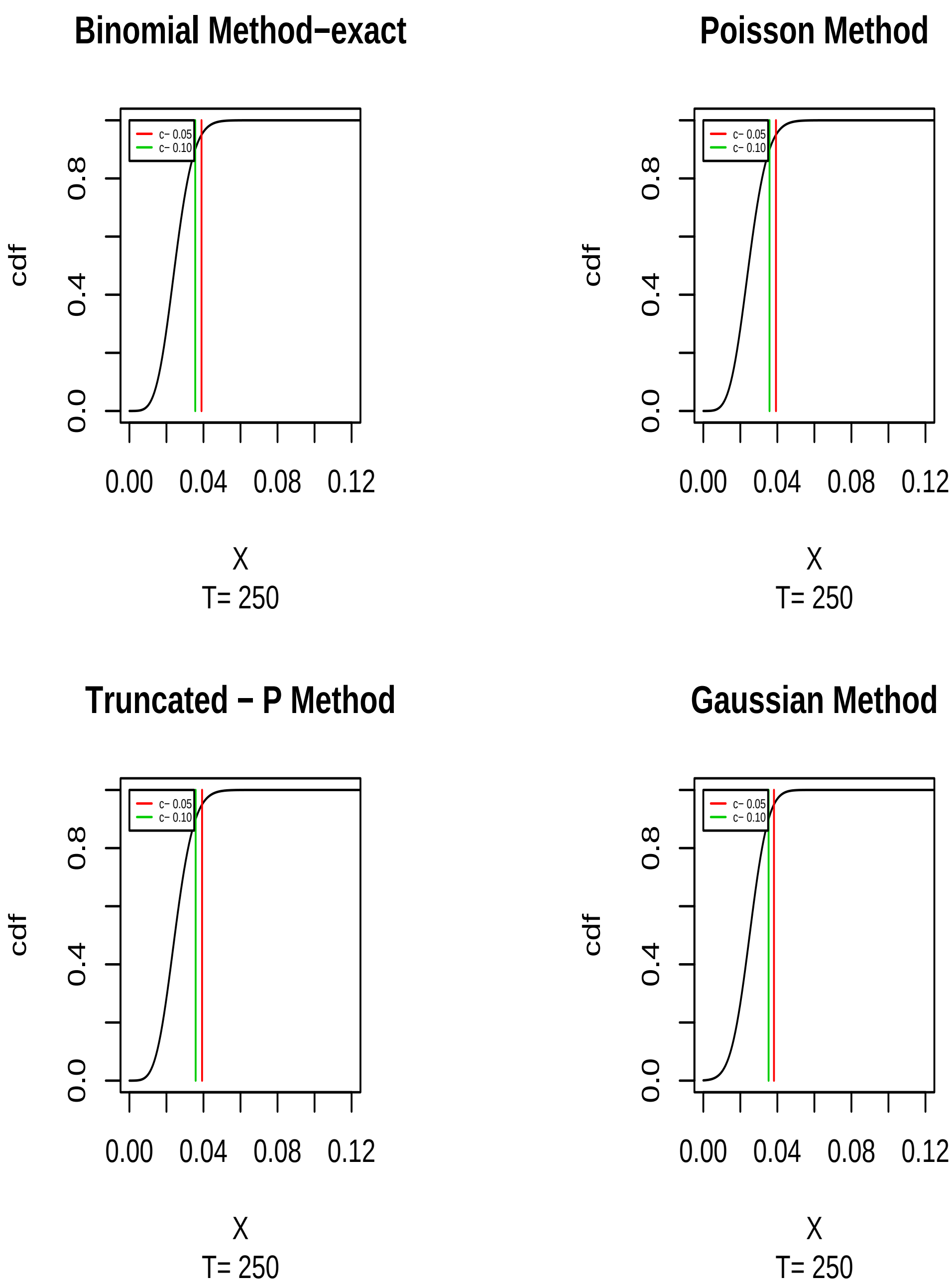

Figure 6.6: **Cumulative Distributional Functions – Equal Weighting.** $T = 250$, $\alpha = 0.05$

Table 6.2: **Critical Values – Equal Weighting.** *A numerical estimation with FFT, sample points* $= 2^{13}$, *interval* $[0, 1]$, *Linear Interpolation* $=$ *No.*

		Critical Values – c_β			
		Exact	**Approximate**	**Approximate**	**Asymptotic**
β	T	**Binomial**	**Poisson**	**Truncated P**	**Gaussian**
	2	0.24835200	0.24639890	0.24572750	0.17231160
	5	0.16577150	0.16693120	0.16693120	0.11816805
	10	0.09924316	0.10015869	0.10015869	0.09087976
0.05	50	0.05810547	0.05883789	0.05883789	0.05446232
	100	0.04772949	0.04821777	0.04821777	0.04583301
	250	0.03894043	0.03924560	0.03924560	0.03817595
	1000	0.03179932	0.03192139	0.03192139	0.03158798
	2	0.00000000	0.00000000	0.00000000	0.13977460
	5	0.12011720	0.11993400	0.11993400	0.09758984
	10	0.08636475	0.08685303	0.08685303	0.07632877
0.10	50	0.04937744	0.04980469	0.04980469	0.04795492
	100	0.04193115	0.04229736	0.04229736	0.04123158
	250	0.03558350	0.03576660	0.03576660	0.03526575
	1000	0.03021240	0.03033447	0.03033447	0.03013288

Table 6.3: **Critical Values – Equal Weighting.** *A numerical estimation with FFT, sample points* $= 2^{12}$*, interval* $[0, 1]$*, Linear Interpolation* $=$ *Yes.*

		Critical Values – c_β			
		Exact	**A**pproximate	**A**pproximate	**A**symptotic
β	T	**B**inomial	**P**oisson	**T**runcated P	**G**aussian
	2	0.00012173	0.24643676	0.24578471	0.17231160
	5	0.16579990	0.16699950	0.16699930	0.11816805
	10	0.09930414	0.10018474	0.10018474	0.09087976
0.05	50	0.05813861	0.05891749	0.05891749	0.05446232
	100	0.04775999	0.04827149	0.04827149	0.04583301
	250	0.03900192	0.03930324	0.03930324	0.03817595
	1000	0.03184441	0.03198380	0.03198380	0.03158798
	2	0.00000000	0.00012142	0.00012140	0.13977460
	5	0.12018720	0.11998240	0.11998210	0.09758984
	10	0.08645226	0.08692432	0.08692432	0.07632877
0.10	50	0.04941397	0.04988423	0.04988423	0.04795492
	100	0.04199696	0.04233696	0.04233696	0.04123158
	250	0.03561592	0.03582833	0.03582833	0.03526575
	1000	0.03026758	0.03037117	0.03037117	0.03013288

The above graphs gives us some idea about the shape of the distributions with different sample sizes. Gibbs's Phenomena are detected due to the discontinuity of the density at zero. Since we calculate the critical values on the right tail of the distribution, Gibbs's Phenomena should not affect our results very much.

The above two tables show that when sample size T is reasonably big, the three proposed approximation methods perform well - close to the exact critical values. When sample size is small, Gaussian method turns to be more conservative while Poisson and Truncated Poisson methods still perform reasonably well. It does not show much difference between Poisson and Truncated Poisson method[6], except for the extreme case of $T = 2$.

By comparing the above two tables, we know that using linear interpolation method for the discrete points on the estimated distribution curve is computationally more efficient. The grid points for integration can be reduced at least by half to achieve the same accuracy for the critical values.

[6]Truncated Poisson method did show some advantage in reducing the thickness of the right tail at $x = 1$, when sample size T is small.

Reciprocal Weighting By reciprocal weighting, we mean that $w(x) = \frac{1}{x}$ and $W(x) = \ln x$.

Thus

$$\hat{F}_T^R(z) = \frac{1}{Tz}\sum_{t=1}^{T} \mathbf{I}_{\{\gamma_{\mathbf{t+k}} \leq \mathbf{z}\}},$$

and

$$X_{T,R}^{\alpha} = \frac{1}{T\alpha}\sum_{t=1}^{T}(\ln\alpha - \ln\gamma_{t+k})^{+}. \tag{6.12}$$

We define $m_t' = (\ln\alpha - \ln\gamma_{t+k})^{+}$s, iid with the PDF of mixture of a mass, $1-\alpha$, at 0 and a α times Exponential (1) on $(0,\infty)$, so that we can omit the suffix t and denote its CDF as

$$F(m') = \Pr(M' \leq m') = \begin{cases} 0 & \text{when } m' < 0, \\ 1-\alpha & \text{when } m' = 0, \\ 1-\alpha e^{-m'} & \text{when } m' > 0 \end{cases}$$

If the sample size T is reasonably large, we can estimate the critical values for the test by an approximation. One example will be the following asymptotic normal approximation.

We notice that

$$E(M') = \alpha \leq 1, \quad \alpha \in (0,1];$$

and

$$\mathrm{VaR}_{\alpha}(M') = E(M'^2) - [E(M')]^2 = 2\alpha - \alpha^2 \leq 1, \quad \alpha \in (0,1].$$

By Central Limit Theory, we have

$$\sqrt{T}\left(X_{T,R}^{\alpha} - 1\right) \xrightarrow{\mathbf{d}} N\left(0, \frac{2-\alpha}{\alpha}\right). \tag{6.13}$$

For a large sample size T, the result of (6.13) can be used to construct the critical values and confidence intervals for the test statistics (6.12):

$$c_{\beta} = 1 + \frac{1}{\sqrt{T}} z_{\beta} \sqrt{\frac{2-\alpha}{\alpha}},$$

when β is the specified the confidence level, T and α are given, and z_{β} is the critical value of the standard normal distribution.

As we discussed before, $T = 250$ is a typical case and the above value may not be very close to the true one (A Monte Carlo study will illustrate this point later). Another approximation is considered here based on a Poisson distribution in hope of doing a better job.

Notice that we can rewrite the test statistic (6.12) as

$$X_{T,R}^{\alpha} = \frac{1}{T\alpha} \sum_{t=1}^{N} E_t. \tag{6.14}$$

where $N \sim \mathrm{Binomial}(T, \alpha)$ and $E_t \overset{iid}{\sim} \mathrm{Exponential}(1)$. N and E_t's are independent.

The characteristic function (c.f.) of $X^{\alpha}_{T,R}$ is

$$\begin{aligned}
\phi'(t) &= E(e^{itX^{\alpha}_{T,R}}) \\
&= \int_0^{\infty} e^{itx} d[G(x)] \\
&= E(e^{it\frac{1}{T\alpha}\sum_{t=1}^{N} E_t}) \\
&= E[E(e^{it\frac{1}{T\alpha}\sum_{t=1}^{N} E_t})|N] \\
&= E\left[\left(\frac{1}{1-i\frac{t}{T\alpha}}\right)^{N} |N\right]
\end{aligned} \tag{6.15}$$

where $G(x)$ is the CDF of $X^{\alpha}_{T,R}$.

Poisson $(T\alpha)$ can be a good approximation to N ~Binomial (T, α), with PDF $p_n = Pr(N = n) = e^{-T\alpha}\frac{(T\alpha)^n}{n!}$. Thus, the characteristic function (c.f.) of the Poisson approximation, denoted by $\phi'_P(t)$, is

$$\phi'(t) \approx \phi'_P(t) = \sum_{n=0}^{\infty} \frac{1}{(1-i\frac{t}{T\alpha})^n} e^{-T\alpha} \frac{(T\alpha)^n}{n!}$$

$$= \sum_{n=0}^{\infty} \left[\frac{T\alpha}{1 - i\frac{t}{T\alpha}}\right]^n \frac{e^{-T\alpha}}{n!}$$

$$= e^{-T\alpha\left[1 - \frac{1}{1 - it/(T\alpha)}\right]} \tag{6.16}$$

When the maximum sample size is T, the use of the truncated Poisson $(T\alpha)$ at T might be reasonable, that is

$$\phi'(t) \approx s(T) \sum_{n=0}^{T} \left[\frac{T\alpha}{1 - i\frac{t}{T\alpha}}\right]^n \frac{e^{-T\alpha}}{n!}, \tag{6.17}$$

where $s(T) = \frac{e^{T\alpha}}{\sum_{n=0}^{T} \frac{(T\alpha)^n}{n!}}$.

The exact or binomial c.f. is

$$\phi'(t) = \sum_{n=0}^{T} \frac{1}{(1 - i\frac{t}{T\alpha})^n} p_n$$

$$= (1-\alpha)^T \sum_{n=0}^{T} \binom{T}{n} \left[\frac{\alpha}{(1-\alpha)(1 - i\frac{t}{T\alpha})}\right]^n$$

$$= \left(1 - \alpha\left(1 - \frac{1}{1 - it/(T\alpha)}\right)\right)^T \tag{6.18}$$

where $p_n = \Pr(N = n) = \binom{T}{n} \alpha^n (1-\alpha)^{T-n}$.

Again, we will devote our effort to the estimation of the continuous part of the distribution functions. $\phi(t)$, p and the corresponding $\phi_c(t)$ are listed in the following table:

Table 6.4: **CF for Continuous Part – Reciprocal Weighting.** *Characteristic functions of the corresponding continuous distributions with different methods.*

	$\phi(t)$	p	$\phi_c(t)$
Binomial	$\left[1-\alpha+\frac{\alpha}{1-\frac{it}{T\alpha}}\right]^T$	$(1-\alpha)^T$	$\frac{\left[1-\alpha+\frac{\alpha}{1-\frac{it}{T\alpha}}\right]^T-(1-\alpha)^T}{1-(1-\alpha)^T}$
Poisson	$e^{-T\alpha\left[1-\frac{1}{1-\frac{it}{T\alpha}}\right]}$	$e^{-T\alpha}$	$\frac{e^{-T\alpha\left[1-\frac{1}{1-\frac{it}{T\alpha}}\right]}-e^{-T\alpha}}{1-e^{-T\alpha}}$
Truncated-P	$S(T)\sum_0^T\left[\frac{T\alpha}{1-\frac{it}{T\alpha}}\right]^n\frac{e^{-T\alpha}}{n!}$	$S(T)e^{-T\alpha}$	$\frac{S(T)\sum_0^T\left[\frac{T\alpha}{1-\frac{it}{T\alpha}}\right]^n\frac{e^{-T\alpha}}{n!}-S(T)e^{-T\alpha}}{1-S(T)e^{-T\alpha}}$

The same argument, numerical inversion of a characteristic function, as for the equal weighting test statistic is suffice to generate critical values for the reciprocal weighting case. Please refer to Remark 6, for the detailed discussion of the computing steps. Graphs and tables are provided for comparisons.

A comparison of their performances can be viewed in the following graphs and tables.

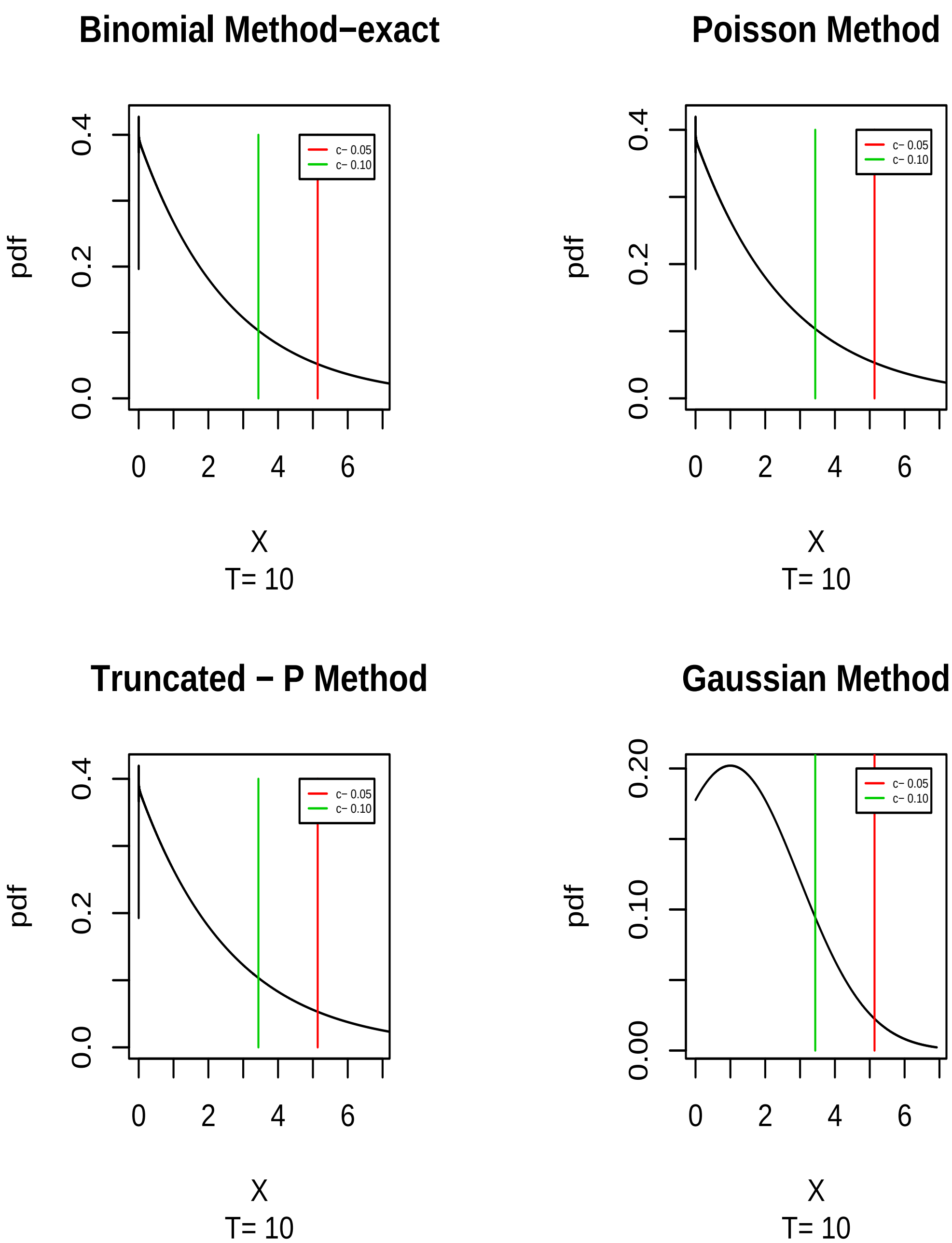

Figure 6.7: **Probability Density Functions – Reciprocal Weighting.** $T = 10$, $\alpha = 0.05$

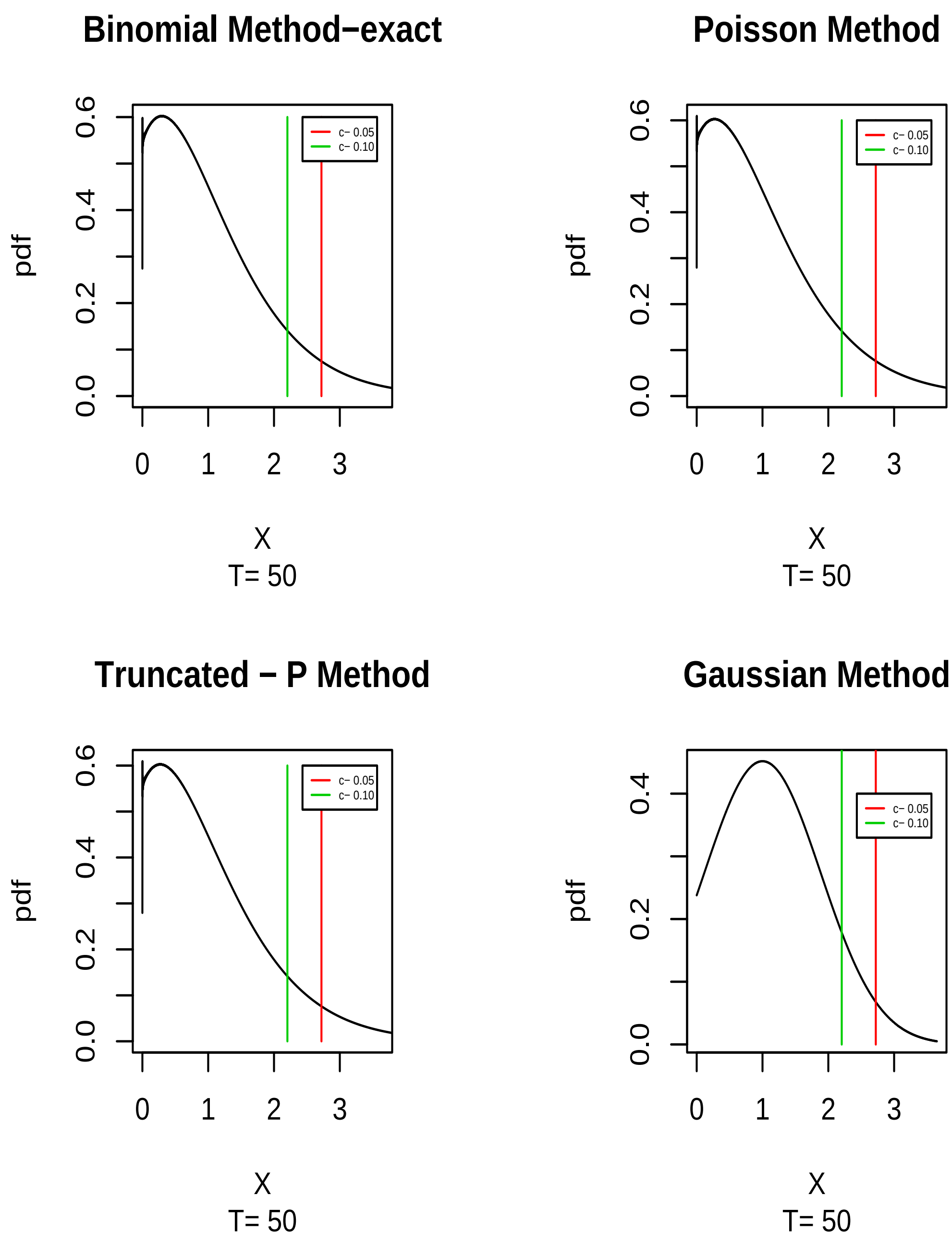

Figure 6.8: **Probability Density Functions – Reciprocal Weighting.** $T = 50$, $\alpha = 0.05$

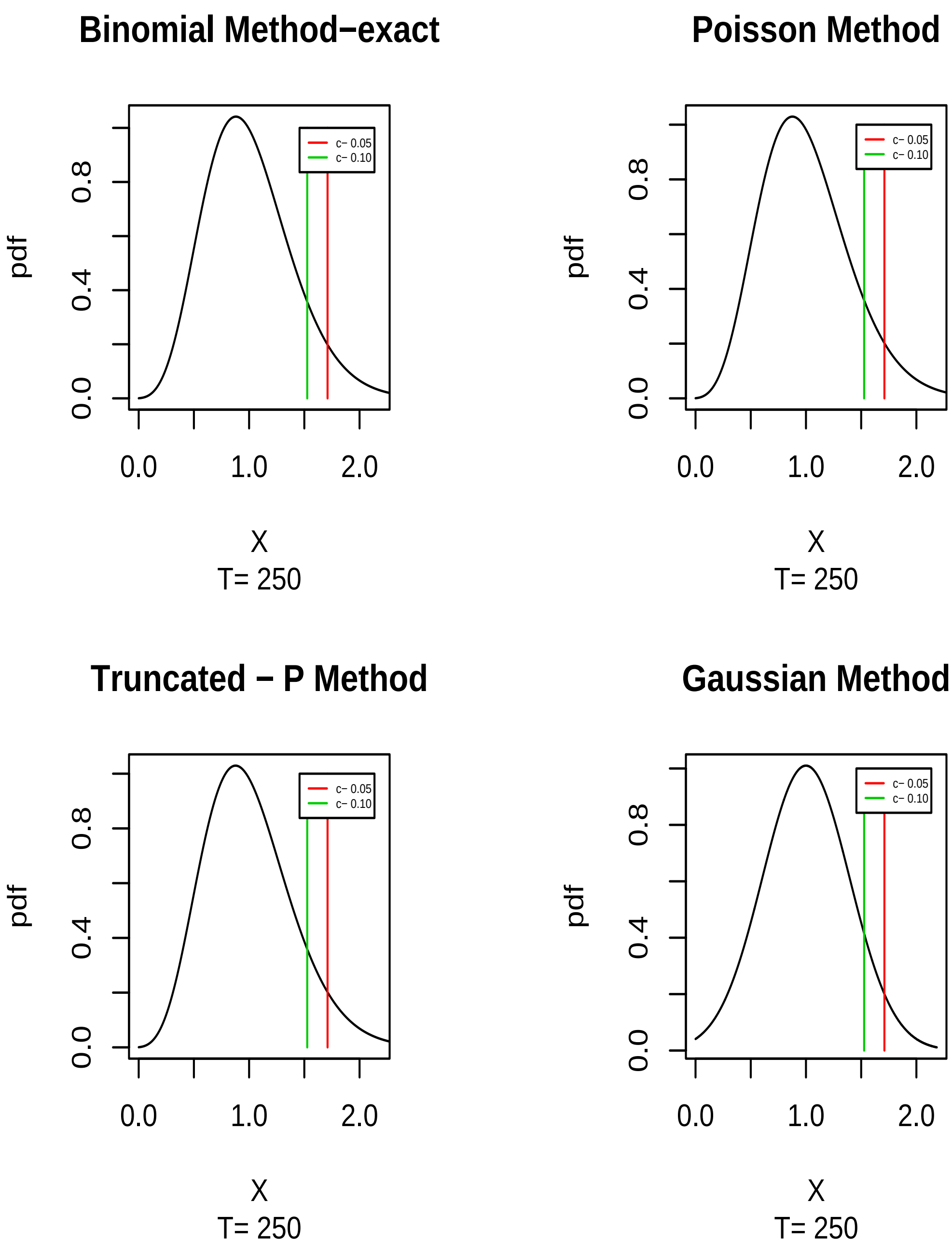

Figure 6.9: **Probability Density Functions – Reciprocal Weighting.** $T = 250$, $\alpha = 0.05$

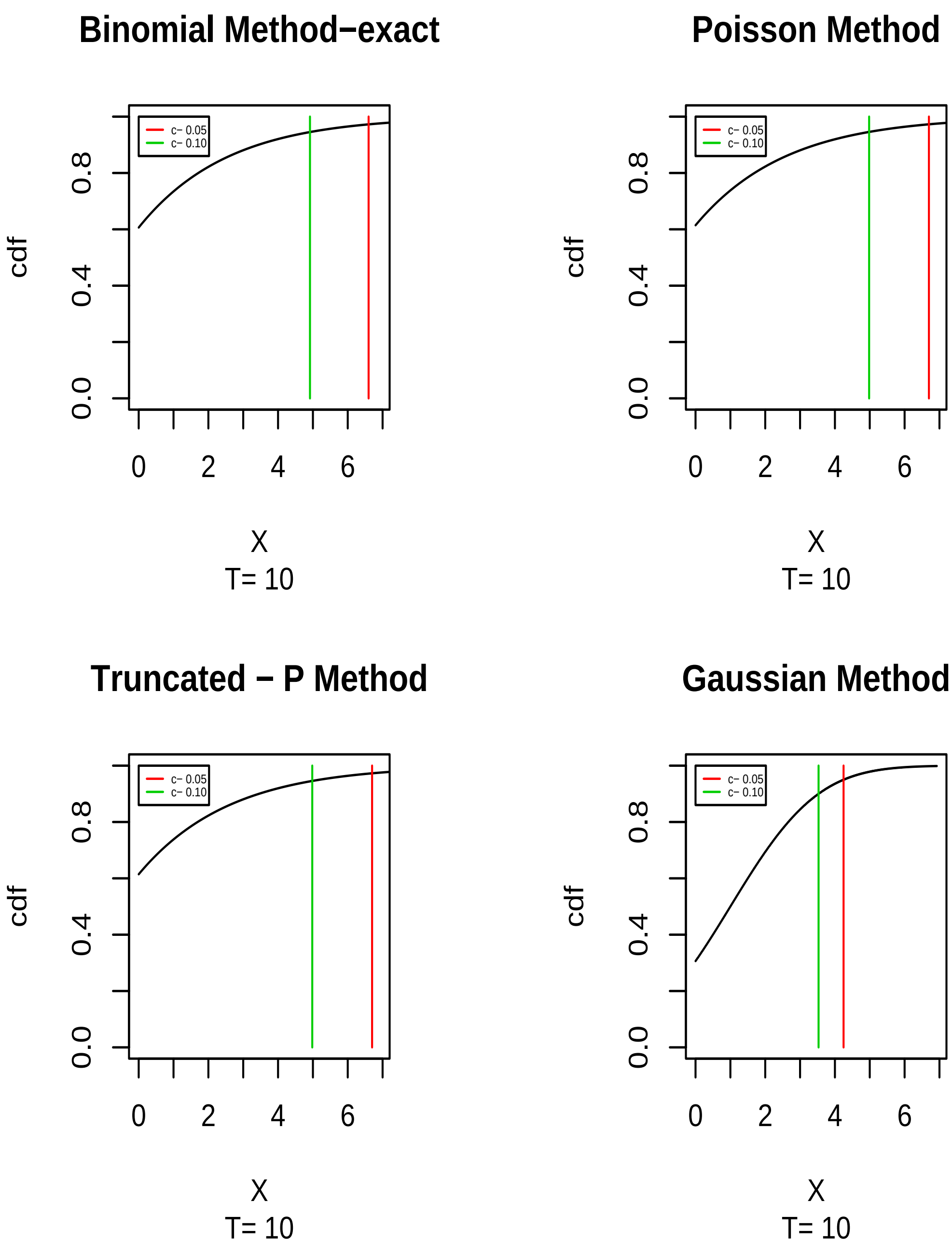

Figure 6.10: **Cumulative Distribution Functions – Reciprocal Weighting.** $T = 10$, $\alpha = 0.05$

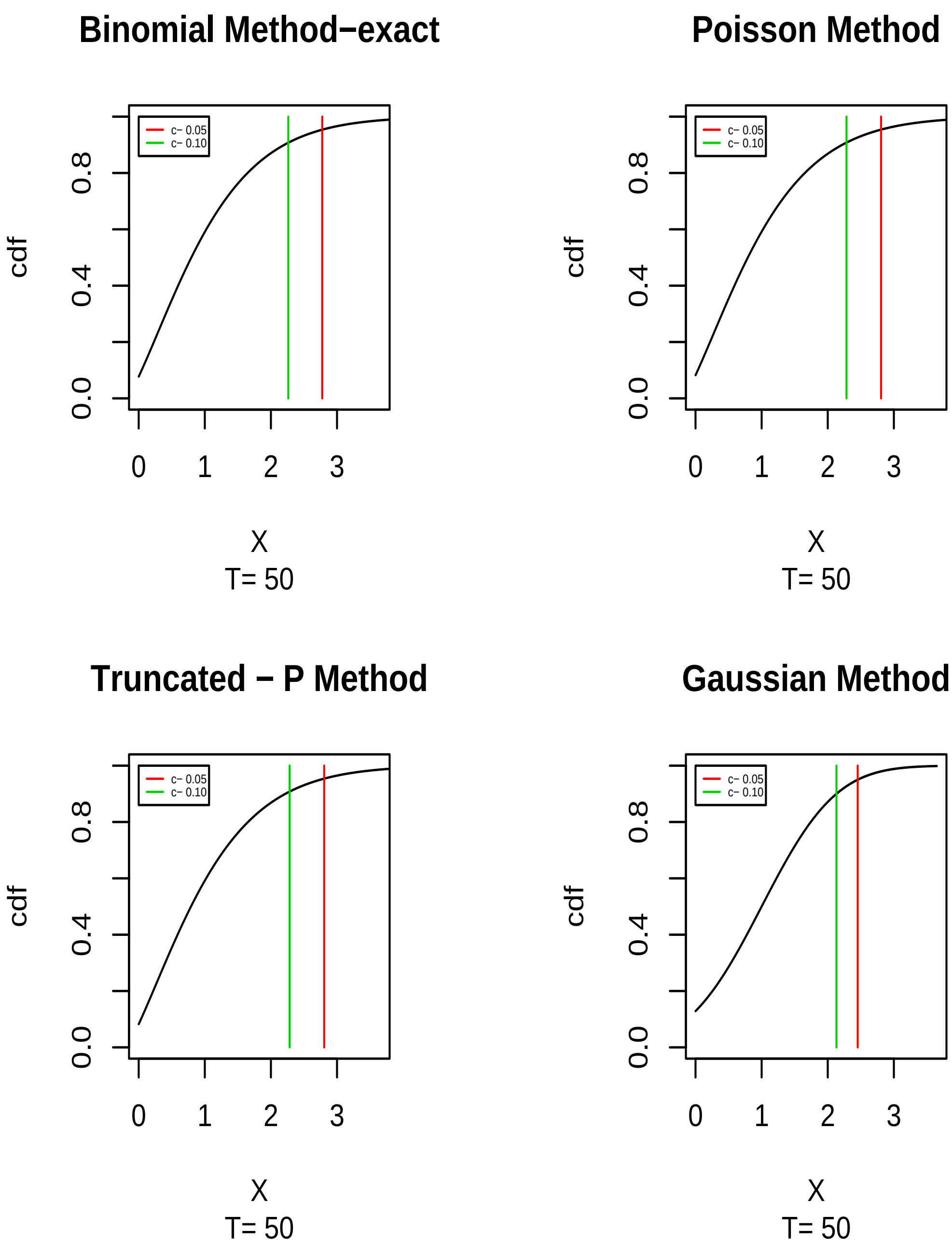

Figure 6.11: **Cumulative Distribution Functions – Reciprocal Weighting.** $T = 50$, $\alpha = 0.05$

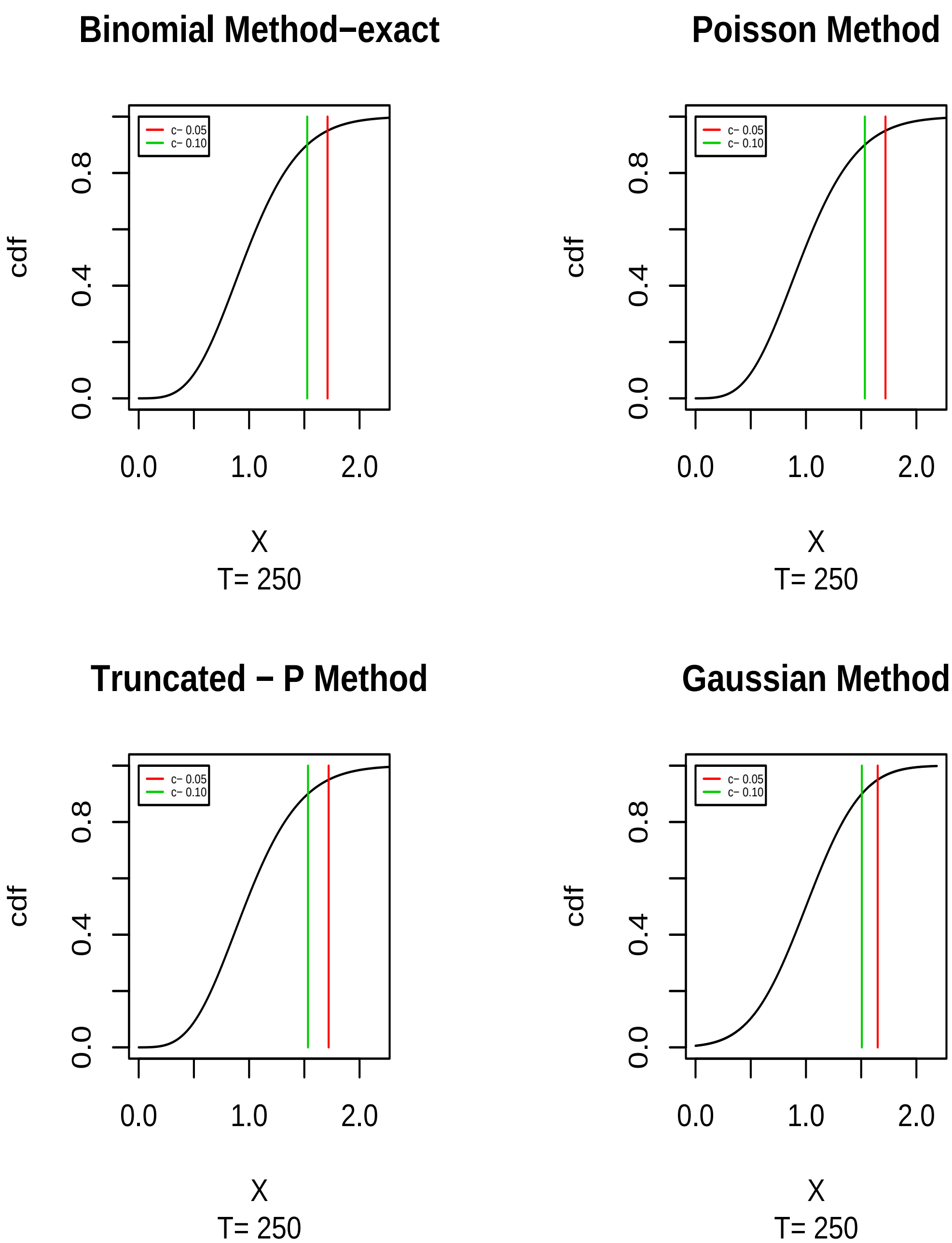

Figure 6.12: **Cumulative Distribution Functions – Reciprocal Weighting.** $T = 250$, $\alpha = 0.05$

Table 6.5: **Critical Values – Reciprocal Weighting.** *A numerical estimation with FFT, sample points* $= 2^{20}$, *interval* $[0, 3^4]$, *Linear Interpolation* = *No.*

		Critical Values – c_β			
		Exact	**Approximate**	**Approximate**	**Asymptotic**
β	T	**Binomial**	**Poisson**	**Truncated P**	**Gaussian**
	2	4.209300	3.998569	3.982347	8.263477
	5	6.316615	6.342339	6.342339	5.593826
	10	5.137044	5.187409	5.187409	4.248326
0.05	50	2.720970	2.745535	2.745535	2.452695
	100	2.170349	2.186803	2.186803	2.027211
	250	1.710185	1.719918	1.719918	1.649665
	1000	1.340632	1.345190	1.345019	1.324833
	2	0.000000	0.000000	0.000000	6.659178
	5	3.252511	3.208171	3.208171	4.579179
	10	3.435897	3.450728	3.450728	3.530862
0.10	50	2.203488	2.219092	2.219092	2.131836
	100	1.843978	1.855102	1.855102	1.800329
	250	1.526336	1.533288	1.533288	1.506172
	1000	1.258673	1.262072	1.262072	1.253086

Table 6.6: **Critical Values – Reciprocal Weighting.** *A numerical estimation with FFT, sample points* $= 2^{15}$, *interval* $[0, 3^4]$, *Linear Interpolation* $=$ *Yes.*

		Critical Values – c_β			
		Exact	**Approximate**	**Approximate**	**Asymptotic**
β	T	**Binomial**	**Poisson**	**Truncated P**	**Gaussian**
	2	4.209453	3.998739	3.982475	8.263477
	5	6.316839	6.342550	6.342537	5.593826
	10	5.137266	5.187574	5.187574	4.248326
0.05	50	2.721092	2.745672	2.745672	2.452695
	100	2.170441	2.186898	2.186898	2.027211
	250	1.710255	1.720000	1.720000	1.649665
	1000	1.340743	1.345274	1.345274	1.324833
	2	0.000000	0.000000	0.000000	6.659178
	5	3.252675	3.208335	3.208329	4.579180
	10	3.436002	3.450857	3.450857	3.530862
0.1	50	2.203603	2.219186	2.219186	2.131836
	100	1.844090	1.855171	1.855171	1.800329
	250	1.526458	1.533363	1.533363	1.506172
	1000	1.258784	1.262146	1.262146	1.253086

Again, the above graphs give us some idea about the shape of the distributions with different sample sizes. Gibbs's Phenomena are detected due to the discontinuity of the density at zero. Since we calculate the critical values on the right tail of the distribution, again, they should not affect our results very much.

The above two tables show that when sample size T is reasonably big, the Poisson and Truncated Poisson approximation methods perform well - close to the exact critical values. Gaussian method turns out to be loose, in the sense that it over-estimate the critical values, when T is small; it is conservative, in the sense that it under-estimate the critical values, when T is reasonably big. When T is big enough, all methods will converge to the asymptotic Gaussian, but the converging speed is much slower than the equal weighting case. Part of the reason was because the exponential distribution is a highly skewed one, the converging speed of the empirical distribution is slow. The related discussion can be found in consistency and power consideration.

When sample size is small, Poisson and Truncated Poisson methods still perform reasonably well. It does not show much difference between Poisson and Truncated Poisson method.

By comparing the above two tables, we know that using linear interpolation method for the discrete points on the estimated distribution curve is computationally more efficient . The grid points for integration can be reduced at least by seven eighthes to achieve the same accuracy for the critical values.

Asymptotic Properties

Remark 7 *Consistency Consideration of Proposed Test Statistics.*

Assume the true (accurate) conditional CDF of $\boldsymbol{\epsilon_t}$ as in Assumption 2 is $G_t(x)$, and model-defined $F_t(x)$[7]

$$\boldsymbol{\epsilon_t} \sim G_t(x)$$

thus by Rosenblatt's discussion in [46]

$$G_t(\boldsymbol{\epsilon_t}) \overset{iid}{\sim} \text{Uniform}(0,1), t = 1, \cdots, T.$$

Instead, we use model-defined CDF, $F_t(x)$, when the true CDF is not available. We want to explore the properties of $Y_t = F_t(\boldsymbol{\epsilon_t})$. If it is an iid sample, the asymptotic property of the class of the proposed test statistics is much more manageable.

We know that

$$G_t[F_t^{-1}(Y_t)] \overset{iid}{\sim} \text{Uniform}(0,1), t = 1, \quad \cdots, T,$$

thus,

$$Y_t = F_t[G_t^{-1}(U_t)]$$

In developing our consistent version of the proposed test statistic, we confine our attention to a random sample $\{\ \epsilon_1, \cdots, \epsilon_n\ \}$ from an unconditional distribution $G(\cdot)$ on $\mathbb{R}$ by omitting sub-index t. It seems possible to extend the results below to the

[7]We use simplified notation here to illustrate the basic idea.

heterogeneous and /or time series case - for instance, the following results hold for any $F_t(x)$ and $G_t(x)$ such that

$$F_t[G_t^{-1}(U_t)] \equiv F[G^{-1}(U_t)], \tag{6.19}$$

for some time-independent CDFs, $F(\cdot)$ and $G(\cdot)$.

We call a CDF $F(\cdot)$ under estimated for a CDF $G(\cdot)$ if and only if $F(x) \leq G(x)$, for all $x \in \mathbb{R}$, which is our alternative hypothesis of interest.

By Rosenblatt (1952) in [46] and the above assumption,

$$Y_t = F(\epsilon_t) \leq G(\epsilon_t) = \gamma_t \overset{iid}{\sim} \text{Uniform}(0,1), t = 1, \quad \cdots, T,$$

i.e. Y_t is randomly smaller than γ_t. Then we know that

$$F_Y(\cdot) \geq F_U(\cdot).$$

By the Glivenko-Cantelli theorem, as discussed in 3.3,

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{I}_{\{\gamma_t \leq z\}} \xrightarrow{P} z, \;\; z \in [0,1].$$

By the Dominance Convergence Theorem,

$$X_{T,W}^{\alpha} = \frac{1}{\alpha}\int_0^{\alpha} \hat{F}_T^W(z)dz \xrightarrow{P} \frac{1}{\alpha}\int_0^{\alpha} w(z)zdz$$

if $w(z)$ is bounded[8] or integrable on $[0, \alpha]$.

[8] The reciprocal weighting scheme is not satisfying this condition in general, because of the singular point at $z = 0$.

Under the alternative hypothesis,

$$\frac{1}{\alpha}\int_0^{\alpha} \hat{F}_T^W(z)dz \xrightarrow{P} \frac{1}{\alpha}\int_0^{\alpha} w(z)F_Y(z)dz \geq \frac{1}{\alpha}\int_0^{\alpha} w(z)zdz.$$

Since

$$F_Y(z) = GF^{-1}(z),$$

the proposed test statistic will be consistent by definition, if $GF^{-1}(Z) - Z \neq 0$ almost surely.

Remark 8 *Asymptotic Normality Under H_0 - a generalized case.*

We have individually explored the asymptotic normality of equal and reciprocal weighting cases in the first part of this subsection. A general result will be presented here[9], given that $w(z)$ is bounded[10]. Consider the proposed test statistic[11] and define

$$Z_T^* = \left[\int_0^{\alpha} w(z)\frac{1}{T}\sum_{t=1}^{T}\mathbf{I}_{\{\gamma_t \leq z\}}dz\right]$$

Notice that

$$T^{\frac{1}{2}}Z_T^* \quad \propto \quad T^{\frac{1}{2}}\left[X_{T,W}^{\alpha}\right]. \tag{6.20}$$

[9]A more general result may be developed, given any random weighting process $w_T(z)$, which is uniformly integrable (this is certainly the case when w_T is bounded uniformly by the same constant) on $[0,\alpha]$ and converges uniformly in z to the deterministic function $w(z)$ as $T \longrightarrow \infty$, i.e.,

$$\sup_z |w_T(z) - w(z)| \xrightarrow{P} 0,$$

and that $w(z)$ is uniformly bounded

[10]For the parallel case of "risk aversion function" generating a space of coherent risk measures, the weighting function is a normalized function on interval $[0,\alpha]$, or $[0,1]$.

[11]Notice that we can generalize the test statistics by the substitution of the deterministic weighting function $w(z)$ in the previous subsection to a stochastic process $w_T(z)$ on $[0,\alpha]$.

By Remark 7, it is not difficult to see that (6.20) has its first moment

$$\mu_T = T^{\frac{1}{2}} \int_0^\alpha w(z) z dz = T^{\frac{1}{2}} \left(\alpha W(\alpha) - \int_0^\alpha W(x) dx \right),$$

via integration by part, when $W(x)$ is integrable on $(0, \alpha)$ [12], under H_0. Now, its second moment will be our focus here. Without losing generality, we assume $\{Y_t\}_{t=1}^T$ is a random sample defined as in Remark 7. When both $W(x)$ and $W^2(x)$ are integrable on $(0, \alpha)$, we have

$$\begin{aligned}
\sigma^2 &= Var(T^{\frac{1}{2}} Z_T^*) \\
&= T[E(Z_T^*)^2 - E^2(Z_T^*)] \\
&= T\left(E\left[\int_0^\alpha w(z) \frac{1}{T} \sum_{t=1}^T \mathbf{I}_{\{\gamma_t \le z\}} dz \right]^2 - E^2(Z_T^*) \right) \\
&\xrightarrow{P} 2 \int_0^\alpha \int_0^y w(x) w(y) x(1-y) dx dy \\
&= (1-\alpha) \left[\alpha W^2(\alpha) - 2W(\alpha) \int_0^\alpha W(x) dx \right] + \int_0^\alpha W^2(x) dx - \left[\int_0^\alpha W(x) dx \right]^2
\end{aligned} \tag{6.21}$$

Thus, under H_0

$$\frac{T^{\frac{1}{2}} Z_T^* - \mu_T}{\sigma} \xrightarrow{D(H_0)} AN(0, 1) \tag{6.22}$$

[12] Notice that this result holds for equal weighting case, but not the reciprocal one because the integrability condition is not satisfied.

Power Considerations and Optimal Weights

We can construct some local alternatives against the null hypothesis, and talk about the power of proposed statistics. The exponential families[13] can be good candidates. Once we formulate every thing in terms of weighting functions, we can also talk about the optimal weighting schemes to achieve the optimal power for the specific local alternatives. Again, we confine our attention to the situation where (6.19) holds.

As you might imagine, the results will be very messy even for the exponential families, since they heavily rely on the inverse functions of the CDFs. We start with a simple example of 1-parameter exponential family.

Consider a family of pdf's or pmf's $\{g(x|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$, where

$$g(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp[b(\boldsymbol{\theta})t(x)].$$

In order to introduce the basic idea, we simplify the problem to a special case

$$g(x|\boldsymbol{\theta}) = c(\boldsymbol{\theta})t^{'}(x) \exp[b(\boldsymbol{\theta})t(x)],$$

so that the corresponding cdf's will have the form

$$G(x|\boldsymbol{\theta}) = 1 + a(\boldsymbol{\theta}) \exp[b(\boldsymbol{\theta})t(x)],$$

where $c(\boldsymbol{\theta}) = a(\boldsymbol{\theta})b(\boldsymbol{\theta})$, $a(\boldsymbol{\theta})$ and $b(\boldsymbol{\theta})$ are chosen such that $g(x|\boldsymbol{\theta})$ and $G(x|\boldsymbol{\theta})$ are valid pdf and cdf respectively. Obviously, $a(\boldsymbol{\theta})$ is required to be negative and $b(\boldsymbol{\theta})t(x)$ non-positive.

[13] The location-scale families are also of interest and we will leave it for future work.

Consider the null and alternative hypotheses as a sequence in T,

$$\begin{aligned} H_{0T}: \quad G(x|\boldsymbol{\theta}) &= 1 + a(\boldsymbol{\theta}) \exp[b(\boldsymbol{\theta}) t(x)] \\ H_{1T}: \quad F_T(x|\boldsymbol{\theta}) &= 1 + a(\boldsymbol{\theta}) \exp[(b(\boldsymbol{\theta}) + \beta_T d(\theta)) t(x)] \end{aligned}$$

where β_T converges to zero at the rate of $T^{\frac{1}{2}}$, that is $T^{\frac{1}{2}}\beta_T \longrightarrow \tau \neq 0$. Without losing generality we assume $d(\boldsymbol{\theta})t(x) > 0$, which will simplify our discussion later. The sign of $\beta_T d(\theta)$ is specified according to the sign of $t(x)$[14], such that

$$G(x) > F_T(x),$$

which means that we are only interested in detecting the power of the proposed test against the sequence of the alternatives under-estimating the true cdf's within the family specified by (6.23).

Remark 9 *Power Considerations.* Since β_T is converging to zero, the alternatives are considered local alternatives. The reason that we choose β_T to be order $T^{\frac{1}{2}}$ will be evident when we derive the asymptotic power of the test.

In order to find the asymptotic power, we need to find the limiting distribution of the proposed test statistic for the sequence of the local alternatives:

$$H_{1T}: \quad F_T(x|\boldsymbol{\theta}) = 1 + a(\boldsymbol{\theta}) \exp[(b(\boldsymbol{\theta}) + \beta_T d(\theta)) t(x)].$$

[14] A simple example will be that $\beta_T d(\theta) > 0$, given $t(x) = x$, but we will ignore the specification details here and will focus on the asymptotic properties of the test statistics under H_{1T} when no confusion occurs in the rest of the discussion.

Consider the test statistics, centralized by μ_T and normalized by $T^{\frac{1}{2}}$, i.e.

$$T^{\frac{1}{2}} Z_T^* - \mu_T = T^{\frac{1}{2}} \left[\int_0^{\alpha} w(z) \frac{1}{T} \sum_{t=1}^{T} (\mathbf{I}_{\{F[G^{-1}(U_t)] \leq z\}} - z) dz \right] \tag{6.23}$$

where U_t's are iid Uniform(1) samples, and

$$F[G^{-1}(U_t)] \leq z \Longleftrightarrow U_t \leq 1 + a(\boldsymbol{\theta}) \underbrace{[-a(\boldsymbol{\theta})]^{\frac{-b(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}}}_{h_T(\boldsymbol{\theta}) \longrightarrow -a(\boldsymbol{\theta})^{-1}} (1-z)^{\frac{b(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}}$$

so that we have

$$E_{F_T(\boldsymbol{\theta})}[T^{\frac{1}{2}} Z_T^* - \mu_T] = E_{F_T(\boldsymbol{\theta})} \int_0^{\alpha} w(z) [T^{\frac{1}{2}} (\mathbf{I}_{\{F[G^{-1}(U_t)] \leq z\}} - z)] dz \tag{6.24}$$

$$= \int_0^{\alpha} w(z) [T^{\frac{1}{2}} E_{F_T(\boldsymbol{\theta})} (\mathbf{I}_{\{F[G^{-1}(U_t)] \leq z\}} - z)] dz \tag{6.25}$$

$$\stackrel{\text{Uniform(1)}}{=} \int_0^{\alpha} w(z) T^{\frac{1}{2}} [1 + a(\boldsymbol{\theta}) h_T(\boldsymbol{\theta}) (1-z)^{\frac{b(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}} - z] dz$$

Notice that (6.24) holds because of iid assumption, and that some regularity conditions will be needed for exchanging integrals in (6.25).

Thus, we have

$$E_{F_T(\boldsymbol{\theta})}[T^{\frac{1}{2}} Z_T^* - \mu_T] \longrightarrow \int_0^{\alpha} w(z) \lim_{T \longrightarrow \infty} T^{\frac{1}{2}} [1 + a(\boldsymbol{\theta}) h_T(\boldsymbol{\theta}) (1-z)^{\frac{b(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}} - z] dz \tag{6.26}$$

$$\longrightarrow \int_0^{\alpha} w(z)(1-z) \lim_{T \longrightarrow \infty} T^{\frac{1}{2}} [1 - (1-z)^{\frac{b(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})} - 1}] dz$$

$$= \int_0^{\alpha} w(z)(1-z) \lim_{T \longrightarrow \infty} \underbrace{T^{\frac{1}{2}} [1 - (1-z)^{\frac{-\beta_T d(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}}]}_{\psi_T(z)} dz,$$

Again, some regularity conditions are needed for exchanging limit and integral in (6.26).

By Maclaurin Series, we have

$$\begin{aligned}\psi_T(z) &= T^{\frac{1}{2}}[\frac{-\beta_T d(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}z + o(\beta_T^2)] \\ &\longrightarrow \frac{-T^{\frac{1}{2}}\beta_T d(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}z \\ &\longrightarrow \frac{-\tau d(\boldsymbol{\theta})}{b(\boldsymbol{\theta})}z > 0 \qquad (6.27)\end{aligned}$$

Thus,

$$E_{F_T(\boldsymbol{\theta})}[T^{\frac{1}{2}}Z_T^* - \mu_T] \longrightarrow \underbrace{-\tau\frac{d(\boldsymbol{\theta})}{b(\boldsymbol{\theta})}\int_0^\alpha w(z)z(1-z)dz}_{\tau A(\boldsymbol{\theta},w)}$$

Consider empirical process defined below

$$H_T(z) \equiv \sqrt(T)\left[\frac{1}{T}\sum_1^n \mathbf{I}_{\{U_t\leq z\}} - z\right],$$

with mean zero, variance $z(1-z)$, and

$$cov(H_T(x), H_T(z)) = x(1-z).$$

By weak convergence theory, we know that $H_t(z)$ has a limiting Gaussian process, i.e. Brownian Bridge. Since linear functionals, such as integrals, will not change the Gaussian property, we know that (6.23) has a limiting Gaussian distribution under suitable regularity conditions. We will find the limiting second moment based on the property of Brownian Bridge.

For our case, notice that

$$H_T\left(1-(1-z)^{\frac{b(\boldsymbol{\theta})}{b(\boldsymbol{\theta})+\beta_T d(\boldsymbol{\theta})}}\right) \xrightarrow{D} H_T(z),$$

so that

$$\begin{aligned} Var_{F_T(\boldsymbol{\theta})}[T^{\frac{1}{2}}Z_T^* - \mu_T] &\xrightarrow{P} 2\int_0^{\alpha}\int_0^{y} w(x)w(y)x(1-y)dxdy \\ &= \sigma^2. \end{aligned}$$

Let us define

$$Q_T(w) \equiv \frac{T^{\frac{1}{2}}Z_T^* - \mu_T}{\sigma},$$

and we know that

$$Q_T(w) \xrightarrow{D(H_{0T})} N(0,1)$$

$$Q_T(w) \xrightarrow{D(H_{1T})} N\left(\frac{\tau A(\boldsymbol{\theta}, w)}{\sigma}, 1\right). \tag{6.28}$$

The asymptotic power is illustrated by the following graph,

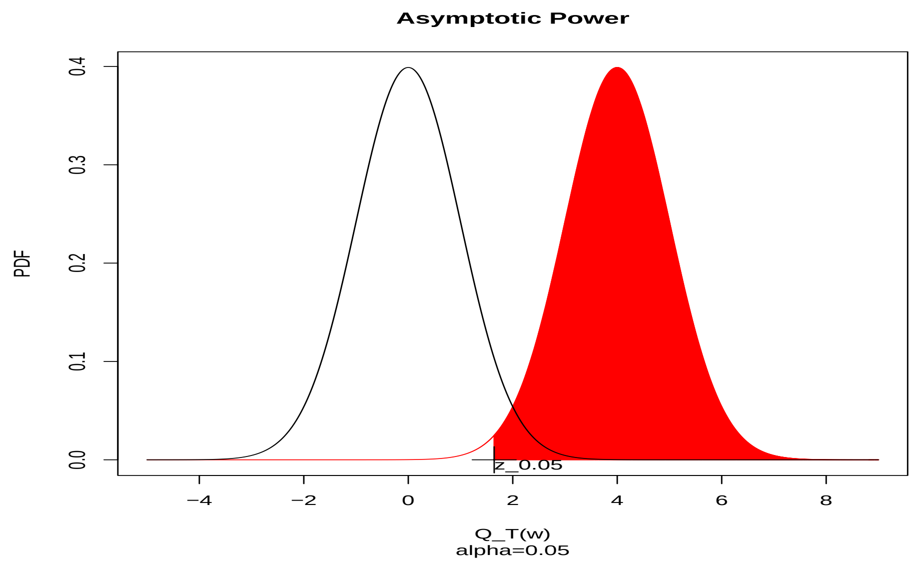

Figure 6.13: **Graphical illustration of Asymptotic Power.**

An asymptotic level α test (one-sided) rejects H_{0T}, when

$$|Q_T(w)| \geq z_\alpha.$$

Thus, we know that the asymptotic power to detect the alternative H_{1T} will coverage to

$$\Phi\left(\frac{\tau A(\boldsymbol{\theta}, w)}{\sigma} - z_\alpha\right),$$

where $\Phi(\cdot)$ denotes the CDF of a standard normal distribution.

Since the power increases as the non-centrality parameter $\frac{\tau A(\boldsymbol{\theta},w)}{\sigma}$, therefore the weighting function $w(z)$ which maximize $\frac{\tau A(\boldsymbol{\theta},w)}{\sigma}$ or

$$\frac{-\tau\frac{d(\boldsymbol{\theta})}{b(\boldsymbol{\theta})}\int_0^\alpha w(z)z(1-z)dz}{\left[2\int_0^\alpha\int_0^y w(x)w(y)x(1-y)dxdy\right]^{\frac{1}{2}}},$$

will generate the optimal asymptotic power.

6.2 Other Approaches

Beyond hypothesis testing method, tests based on the idea of minimizing economic loss function, was proposed, too. See Lopez (1997) [35] and some other literatures for details. This kind of approaches are generally considered outside the hypothesis-testing paradigm, but can also be a reasonable criteria to test the model validity or forecasting accuracy.

Part IV

Summarizing Remarks

Chapter 7

Summarizing Remarks

7.1 Conclusion

This dissertation offers a systematic study on the coherent risk management problems in a decision-under-risk framework, where the distributions of the uncertain outcomes are always known or estimable to the decision makers. Both economic and statistical aspects of the answers to the risk management problems are provided with great details, although the main focus is always on the statistical solutions.

In the early part of this dissertation research, we investigate the economic reasons why a coherent risk measure, such as ES_α, is an interesting quantity. We show that ES_α is not only a reasonable screening device in ranking the the portfolio choices but also a natural benchmark measure for the risk capital assessment, comparing to a non-coherent risk measure VaR_α. As a risk measure in the two-parameter portfolio selection rules, ES_α is consistent with the Expected Utility Hypothesis for all the decision makers with nondecreasing and convex utilities, thus consistent with the SSD. As a risk measure indexing the level of risk capital to protect institutions from extreme losses, ES_α is a natural remedy for VaR_α, given that VaR_α is not sub-additive but well-accepted as an industrial standard for risk capital calculations. Disaggregating the risk capital down to the business level also catches enough attention in this research, because we believe that fairly allocating risk capital is essential to keep a stable incentive structure among the trading desks or alternatively locally signaling the better performed assets in the given portfolio.

The later part of this dissertation research is dedicated to the statistical methods in solving the coherent risk management problems. Both modeling methods for VaR_α, ES_α and SES_α, and back-testing methods based on ES_α are discussed. We offer consistent estimators in the non-parametric modeling framework. Analytic expressions and simulation studies in the parametric framework are provided for the multivariate Gaussian, the fat-tailed multivariate t-distribution and the general elliptical families. Back-testing methods based on ES_α in a hypothesis test framework are proposed. A class of consistent test statistics with different weighting schemes on the tail part are proposed and studied carefully. We tabulate the critical values to the three digit accuracy for two test statistics with particular weighting functions, in four estimation methods. Three of the estimation methods involve the numerical inversion of characteristic functions. The asymptotic normality of the test statistics is proved in a general setup. The asymptotic power function is provided for a class of exponential families, with optimal weighting scheme discussed.

7.2 Future Research

As we have already book-marked in the body of the dissertation, there are quite a few topics that we would like to leave for future work. Meanwhile, we can not enumerate them all here but list a few major issues as examples.

One is to model the ES_α with spline methods and compare the accuracy loss and gain with the existing kernel methods or other nonparametric methods for the purpose of estimating SES_α. As Scaillet (2002) [52] also mentioned that spline method is anther option to achieve smooth estimators for CDFs and PDFs, so that VaR_α, ES_α and SES_α are easily calculated. We do not devote any time on this topic in the dissertation, partially because we have been focusing on other issues so that we do not have a chance to really work on it.

Another one is to study some real portfolios (P/L) behaviors. Dr. Kyle personally has been suggesting some real data from different markets as a starting point for the ES_α modeling, since it will be very interesting to see some good applications from our theoretical discussions. The potential applicable methods include the methods for the fat-tailed return data, and ES_α based back-testing methods for model validation. Again, this line of research is computational intensive and a careful selection of empirical data is always crucial.

The last potential future topic in this list is the continuing study on the asymptotic power properties for proposed test statistics. We offer a power function for an exponential family in the dissertation and talk about the optimal weighting schemes. Unfortunately, we do not offer the analytic/numerical solution to it. This line of research requires solving integral equations analytically or numerically. Since this is down to a pure mathematical problem and is computationally intensive, we will do it later on. We would be able to give, at least, a numerical solution in the future. Another distributional family we would like to work on is the location-scale family. Comparing to the exponential family, the location-scale family is harder to deal with because its inverse CDF is not easy to express analytically. The messy part will be the first order or second order approximation for the calculation, but we view it as a workable problem and would like to further investigate it.

Bibliography

[1] Acerbi, Carlo. (2002) Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion. *Journal of Banking & Finance*, 26, 1505-1518.

[2] Acerbi, C., C. Nordio, C. Sirtori. (2001) Expected Shortfall as a Tool for Financial Risk Management. *Working Paper*, www.gloriamundi.org/var/wps.html

[3] Acerbi, c., D. Tasche. (2002) On the coherence of Expected Shortfall. *Working Paper*, http://www.ma.tum.de/stat

[4] Anderson, T.W. (1993) Nonnormal Multivariate Distribution: Inference Based on Elliptically Contoured Distributions. *Multivariate Analysis: Future Directions*, C.R. Rao, eds, Elsevier Science, Amsterdam.

[5] Artzner, P., F. Delbaen, J. Eber, and D. Heath. (1997) Thinking Coherently. *Risk*, 10, 68-71.

[6] Artzner, P., F. Delbaen, J. Eber, and D. Heath. (1999) Coherent Measures of Risk. *Mathematical Finance*, 9, 203-228.

[7] Artzner, P.. (1999) Application of Coherent Risk Measures to Capital Requirements in Insurance. *North American Actuarial Journal*, 3, 11-26.

[8] Aumann, Robert J.. (1964) Markets with A Continuum of Traders. *Econometrica*, 32, 39-50.

[9] Aumann, Robert J.. (1975) Values of Markets with A Continuum of Traders. *Econometrica*, 43, 611-646.

[10] Basel Committee on Banking Supervision. (1996) *Amendment to the Capital Accord to Incorporate Market Risks*, http://www.bis.org/publ/.

[11] Bawa, Vijay S.. (1976) Admissible Portfolios for All Individuals. *Journal of Finance*, 31, 1169-1183.

[12] Bawa, Vijay S., Eric B. Lindenberg. (1977) Capital Market Equilibrium in a Mean-Lower Partial Moment Framework. *Journal of Financial Economics*, 5, 189-200.

[13] Bawa, Vijay S.. (1978) Safty-First, Stochastic Dominance, and Optimal Portfolio Choice. *Journal of Financial and Quantitative Analysis*, 13, 255-271.

[14] Bertsimas, D., G.J. Lauprete, A. Samarov.(2000) Shortfall as a Risk Measure: Properties, Optimization and Applications. *Working Paper*, Sloan School of Management, MIT, Cambridge.

[15] Bohman, H.. (1975) Numerical Inversions of Characteristic Functions. Scandinavian Actuarial Journal, 121-124.

[16] Boos, D.D.. (1984) Using Extreme Value Theory to Estimate Large Percentiles. Technometrics, 26, 33-39.

[17] Bradley, B.O., Murad S. Taqqu. (2002) Financial Risk and Heavy Tails. Heavy-tailed Distribution in Finance, Svetlozar T. Rachev, editor, North Holland.

[18] Chavez-Demoulin, V., P. Embrechts. (2001) Smooth Extremal Models in Finance and Insurance. Working Paper.

[19] Chistoffersen, P.F.. (1998) Evaluating Interval Forecasts. International EconomicReview, 39, 841-862.

[20] Davies, R.B. (1973) Numerical Inversion of Characteristic Function. Biometrica, 60, 415-417.

[21] Delbaen, F.. (2000) Coherent Risk Measures on General Probability Spaces. Working Paper, ETH Zurich, available at http://www.math.ethz.ch/ delbaen/.

[22] Denault, Michel. (2001) Coherent Allocation of Risk Capital. Journal of Risk, 4, 7-21.

[23] Doukhan, P.. (1994) Mixing: Properties and Examples. Lecture Notes inStatistics, Springer-Verlag, New-York.

[24] Embrechts, P., C. Klüppelberg, T. Mikosch. (1997) *Modeling ExtremalEvents for Insurance and Finance*, Berlin:Springer.

[25] Embrechts, P.. (1996) A Survival Kit on Quantile Estimation. *Working Paper*, Swiss Federal Institute of Technology.

[26] Fishburn, P.C.. (1964) *Decision and Value Theory*, New York: Wiley.

[27] Friedman, E. J., Moulin, H. (1999) Three Methods to Share Joint Costs or Surplus. *Journal of Economic Theory*, 87, 275-312.

[28] Friedman, E. J. (1999) optimization Based Characterizations of Cost Sharing Methods. *Working Paper*.

[29] Gallant A.R., Douglas W. Nychka. (1987) Semi-nonparametric Maximum Likelihood Estimation. *Econometrica*, 55, 363-390.

[30] Glasserman, P., Philip Heidelberger and Perwez Shahabuddin. (2002) Portfolio Value-at-Risk with Heavy-Tailed Risk Factors. *Mathematical Finance*, 12, 239-269.

[31] Harlow, W.V., Ramesh K.S. Rao. (1989) Asset Pricing in a Generalized Mean-Lower Partial Moment Framework: Theory and Evidence. *Journal ofFinancial and Quantitative Analysis*, 24, 285-311.

[32] Hosking, J., J. Wallis. (1987) Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 339-349.

[33] Hughett, Paul. (1998) Error Bounds for Numerical Inversion of a Probability Characteristi Function. *SIAM Journal on Mumerical Analysis*, 35, 1368-1392.

[34] Kamdem, Jules S..(2004) Value-at-Risk and Expected Shortfall for Linear Portfolios with Elliptically Distributed Risk factors. *Working Paper*, http://econwpa.wustl.edu/eps/ri/papers/0403/0403001.pdf

[35] Lopez, J.A.. (1997) Regulatory Evaluation of Value-at-Risk Models. *Working Paper*.

[36] Kupiec, P.H.. (1995) Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, 73-84.

[37] Kusuoka, S.. (1997) On law invariant coherent risk measures. *Journalof Risk and Uncertainty*, 14, 103-127.

[38] Machina, Mark J., John F. Pratt. (2001) Increasing Risk: Some Direct Constructions. *In Advances in Mathematical Economics*, 3, 83-95.

[39] Matten, C.. (1996) *Managing Bank Capital*, Wiley, Chichester.

[40] Markowitz, H.. (1952) Portfolio Selection. *Journal of Finance*, 7, 77-91.

[41] McNeil, A.. (1997) Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin*, 27, 117-137.

[42] Pritsker, M.. (1997) Evaluating Value at Risk Methodologies. *Journal ofFinancial Services Research*, 12, 201-242.

[43] Quirk, J.P., R. Saposnik. (1962) Admissibility and Measurable Utility Functions. *Review of Economic Studies*, 29, 140-146.

[44] Rockafellar, R. Tyrrell, Stanislav Uryasev. (2000) Optimization of Conditional Value-at-Risk. *Journal of Risk*, 3, 21-41.

[45] Rockafellar, R. Tyrrell, Stanislav Uryasev. (2002) Conditional Value-at-Risk for General Loss Distribution. *Journal of Banking & Finance*, 26, 1443-1471.

[46] Rosenblatt, M.. (1952) Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23, 470-472.

[47] Rosenmüller, J., Walter Trockel. (2001) Game Theory. *Working Paper*.

[48] Ross, S.A.. (1976) The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13, 341-360.

[49] Rothschild, Michael, Joseph E. Stiglitz. (1970) Increasing Risk: I. A Definition. *Journal of Economic Theory*, 2, 225-243.

[50] Roy, A.D.. (1952) Safty-First and the Holding of Assets. *Econometrica*, 20, 434-449.

[51] Samuelson, P.. (1958) The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means Variances and Higher Moments. *Review of Economic Studies*, 25, 65-86.

[52] Scaillet, Olivier. (2004) Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall. forthcoming in *Mathematical Finance*, 14, 115-129.

[53] Shephard, N.G. (1991) From Characteristic Function to Distribution Function: a simple Framework for the Theory. *Econometric Theory*, 7, 519-529.

[54] Tasche, Dirk. (2002) Expected Shortfall and Beyond. *Journal of Banking& finance*, 26, 1519-1533.

[55] Tasche, Dirk. (2000) Risk Contributions and Performance Measurement. *Working Paper*,

www-m4.mathematik.tu-muenchen.de/m4/Papers/Tasche/riskcon.pdf

[56] Uryasev, Stanislav. (2000) Conditional Value-at-Risk: Optimization Algorithms and Applications. *Financial Engineering News*.

[57] Von Neumann, J., O. Morgenstern. (1947) *Theory of Games andEconomic Behavior (Second ed.)*, Princeton: Princeton University press.

[58] Waller, L.A., Turnbull, B.W., Hardin, J.M. (1995) Obtaining Distribution

Functions by Numerical Inversion of Characteristic Functions with applications. *The American Statistician*, 49, 346-350.