5-7-2008

Moment Problems with Applications to Value-At-Risk and Portfolio Management

Ruilin Tian

Follow this and additional works at: https://scholarworks.gsu.edu/rmi_diss

Part of the Insurance Commons

Recommended Citation
Tian, Ruilin, "Moment Problems with Applications to Value-At-Risk and Portfolio Management."
Dissertation, Georgia State University, 2008.
doi: https://doi.org/10.57709/1059053

This Dissertation is brought to you for free and open access by the Department of Risk Management and Insurance at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Risk Management and Insurance Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
Permission to Borrow

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, or to publish this dissertation may be granted by the author or, in his/her absence, the professor under whose direction it was written or, in his absence, by the Dean of the Robinson College of Business. Such quoting, copying, or publishing must be solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential gain will not be allowed without written permission of the author.

__________________________________
signature of author
Notice to Borrowers

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:

Ruilin Tian  
Department of Risk Management & Insurance  
J. Mack Robinson College of Business  
Georgia State University  
Atlanta, GA 30302-4036

The director of this dissertation is:

Samuel H. Cox, FSA, CFA  
Professor of Actuarial Science  
Department of Risk Management & Insurance  
J. Mack Robinson College of Business  
Georgia State University  
Atlanta, GA 30302-4036

Users of this dissertation not regularly enrolled as students at Georgia State University are required to attest acceptance of the preceding stipulations by signing below. Libraries borrowing this dissertation for the use of their patrons are required to see that each user records here the information requested.

Name of User    Address    Date
MOMENT PROBLEMS WITH APPLICATIONS TO VALUE-AT-RISK AND
PORTFOLIO MANAGEMENT

BY

Ruilin Tian

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy
in the Robinson College of Business
of
Georgia State University

GEORGIA STATE UNIVERSITY
ROBINSON COLLEGE OF BUSINESS

May 2008
Copyright by
Ruilin Tian

May 2008
ACCEPTANCE

This dissertation was prepared under the direction of the RuiLin Tian’s Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor in Philosophy in Business Administration in the Robinson College of Business of Georgia State University.

H. Fenwick Huss, Dean
Robinson College of Business

Dissertation Committee:

Dr. Samuel H. Cox, Chair

Dr. Luis F. Zuluaga

Dr. Shaun Wang

Dr. Adam Speight

Dr. Eric Ulm
To my Mom.
Contents

List of Tables viii
List of Figures ix
Acknowledgements xi
Abstract xii

1 Introduction and Overview 1
  1.1 Moment Problems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
  1.2 Portfolio Optimization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2 Univariate Moment Problem 5
  2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
  2.2 Preliminaries . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
    2.2.1 Moment Problems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
    2.2.2 Dual Problems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
    2.2.3 Sum of Squares . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
    2.2.4 SOS Programming . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
    2.2.5 Optimal Bounds on Value at Risk . . . . . . . . . . . . . . . . . . . . . . . 17
  2.3 Moment Bounds for Arbitrary Distributions . . . . . . . . . . . . . . . . . . . . . 19
    2.3.1 SOS Approach . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
    2.3.2 Smith’s approach . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
  2.4 Moment Bounds for Unimodal Distributions . . . . . . . . . . . . . . . . . . . . . 25
    2.4.1 SOS Approach . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
    2.4.2 Smith’s approach . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
  2.5 The Maximum-Entropy Method . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
  2.6 Numerical Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
    2.6.1 Special Distribution Analysis . . . . . . . . . . . . . . . . . . . . . . . . . 37
    2.6.2 Empirical analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
# CONTENTS

2.7 Stability Experiments ................................................. 45
2.8 Conclusion .............................................................. 48

3 Bivariate Moment Problem .............................................. 49
  3.1 Introduction .......................................................... 49
  3.2 Preliminaries and Notation ........................................ 51
  3.3 SOS Programming Formulations ....................................... 54
    3.3.1 Extreme probability bounds ................................... 54
    3.3.2 VaR probability bounds ....................................... 57
    3.3.3 Bounds on Stop-Loss payments ............................... 67
  3.4 Numerical Examples ................................................ 72
    3.4.1 Example of Extreme Probability Bounds ....................... 72
    3.4.2 Example of VaR Probability Bounds .......................... 78
    3.4.3 Example of Stop-loss Payments .............................. 82
  3.5 Conclusions .......................................................... 85

4 Portfolio Optimization with CVaR-like Constraints ................. 86
  4.1 Introduction .......................................................... 87
  4.2 Portfolio and Efficient Frontier: Descriptions .................... 89
    4.2.1 Definition and Notation ...................................... 89
    4.2.2 Asset-Liability Portfolio .................................... 90
    4.2.3 Optimization Problem Description ........................... 92
  4.3 Improving Skewness of Mean-Variance Portfolio with CVaR .......... 93
    4.3.1 Optimization with CVaR Objective Function .................. 95
    4.3.2 Optimization with CVaR-like Constraints .................... 97
  4.4 Other Portfolio Optimization Approaches .......................... 100
    4.4.1 Boyle-Ding Approach ....................................... 101
    4.4.2 Mean-Absolute Deviation Approach .......................... 104
  4.5 Empirical Illustration: Multiple Assets and Lines of business ........ 106
  4.6 Conclusion .......................................................... 113

A Bounds on $\Pr(X_1 \geq t_1, X_2 \geq t_2)$ ............................... 115

B Bounds on $\Pr(w_1 X_1 + w_2 X_2 \geq a)$ .............................. 119

C Obtain bounds on stop-loss payments from a transformed problem ........ 122

D Proof of CVaR Expression Transformation: Equation (4.9) .............. 123

E Assets and Lines of Business ........................................... 125
List of Tables

2.1 Descriptive Statistics of three lines of business from 1980 to 2005 . . . . . . . . . 43
4.1 Descriptive Statistics of assets and lines of business from 1980 to 2005 . . . . . . . 108
List of Figures

2.1 Contact set of \( h(x) \). ................................................................. 14
2.2 Relationship between bounds and VaR ................................................. 18
2.3 Bounds on cumulative distribution when \( t \leq \mu_1 \), given the first two moments. ... 25
2.4 Bounds on cumulative distribution when \( t \geq \mu_1 \), given the first two moments. ... 25
2.5 Obtain bounds on cumulative distribution given four moments from stair functions. 26
2.6 Use maximum-entropy distribution to match gamma distribution given the first four moments. ................................................................. 36
2.7 Arbitrary and unimodal bounds on the beta distribution with \( a = 2 \), \( b = 3 \) and \( \theta = 5 \), given 2, 4, 6, 8, or 10 moments and the mode. ......................... 38
2.8 Arbitrary and unimodal bounds on beta distribution with \( a = 2 \), \( b = 3 \) and \( \theta = 5 \), given 14 and 16 moments and the mode. ................................. 39
2.9 Arbitrary and unimodal bounds on standard normal distribution, given 2, 4, 5, 8 or 10 moments and the mode. ......................................................... 40
2.10 Arbitrary and unimodal bounds on standard normal distribution, given 12 and 14 moments and the mode. ......................................................... 41
2.11 Arbitrary and unimodal bounds on lognormal distribution with parameters \( \mu = 0.05 \) and \( \sigma = 0.1 \), given 2 or 4 moments and the mode. ................. 41
2.12 Histograms of Allied, PPauto, and Comp ............................................. 42
2.13 Bounds on \( F(t) = \Pr(X \leq t) \) for three lines of business, Allied, PPauto and Comp. 42
2.14 Comparison of arbitrary and unimodal bounds on \( F(t) = \Pr(X \leq t) \) for the line Allied given 4 moments. ......................................................... 43
2.15 Comparison of arbitrary and unimodal bounds on \( F(t) = \Pr(X \leq t) \) for the line PPauto given 4 moments. ......................................................... 44
2.16 Comparison of arbitrary and unimodal bounds on \( F(t) = \Pr(X \leq t) \) for Comp given 2 moments and the mode. Maximum-entropy distributions of Comp given 2 or 4 moments. ......................................................... 44
2.17 Arbitrary bounds of \( \Pr(X \leq t) \) for \( t \geq E(X) + 2\sqrt{\text{Var}(X)} \) on Pareto distribution with \( \alpha = 5 \), \( \theta = 10 \), given 4 moments. ......................................................... 46
2.18 Unimodal bounds of \( \Pr(X \leq t) \) for \( t \geq E(X) + 2\sqrt{\text{Var}(X)} \) on Pareto distribution with \( \alpha = 5 \), \( \theta = 10 \), given 4 moments. ......................................................... 47
2.19 Arbitrary bounds of $Pr(X \leq t)$ for $t \geq E(X) + 2\sqrt{\text{Var}(X)}$ on Pareto distribution with $\alpha = 1, \theta = 10$, given 2 or 4 moments. 48

3.1 3D plots of the upper bound on joint probability $Pr(R \leq t_1, M \leq t_2)$ and the ratio of the upper bound to the bivariate normal with the same moments. 77

3.2 2D plots of the upper bound on the joint probability $Pr(R \leq t_1, M \leq t_2)$ by fixing $t_2$ at $E(M) - k\sqrt{\text{Var}(M)}$ with $k = 0.25, 0.50, 0.75, 1, 1.25$ and $1.5$. 78

3.3 Upper and lower bounds for the probability $Pr(0.5R_{sp} + 0.5R_{nk} \leq a)$. 80

3.4 Comparison of VaR probability bounds with and without exchange option information. 82

3.5 Upper and lower bounds on the expected stop-loss payment by fixing $a$ at $0, 0.25, 0.5, 0.75, 1$ or $1.5$. 84

3.6 Upper and lower bounds on the expected stop-loss payment by fixing $b$ at $0, 0.25, 0.5, 0.75, 1$ or $1.5$. 85

4.1 The efficient frontiers of 5-asset portfolios. 110

4.2 Skewness-variance graph of 5-asset portfolios. 111

4.3 The 5-asset mix for the efficient portfolios. 112

4.4 The efficient frontiers of 20-asset portfolios. 113

4.5 Skewness-variance graph of the 20-asset portfolios. 113

4.6 The efficient frontiers of 14-line and 5-asset portfolios. 114

4.7 Skewness-variance graph of the 14-line and 5-asset portfolios. 114

A.1 Bounds on $Pr(X_1 \geq t_1, X_2 \geq t_2)$ with covariance of $X_1$ and $X_2$ equals 0.5 and -1. 119
Acknowledgements

While a Ph.D. dissertation is primarily an individual effort, I owe much to the help and guidance received from many people. My first greatest appreciation goes to Professor Samuel H. Cox, my advisor during my Ph.D. studies at Georgia State University. I am so lucky to be a student of Dr. Cox, from whom I have learnt far more than I expected. As an advisor, Dr. Cox not only provides unsurpassed guidance on my Ph.D. work, but also teaches me a sense of performing academic research. More than an advisor, Dr. Cox is also a respectable friend, a mentor of my personal life and a model of my future career. Besides my advisor, Dr. Cox, I would like to thank Professor Luis F. Zuluaga, who helps me find my way to numerical linear and quadratic programming. I am particularly grateful to Dr. Yijia Lin, who is not only my colleague and coauthor, but also my trustworthy friend. Thanks for her kindness to help me target academia as my life-time career and perform as a researcher. It has been privilege to do my research under their direction, and I hope that we may continue to work together in the future.

I also show my great appreciation to Dr. Jim Backman’s encouragement in the early stages of this research: my interest in this area of research was kindled while doing a research project for him at the General Re-New England Asset Management, Inc. (GR-NEAM). I am also grateful to Professors Shawn Wang, Eric Ulm and Adam Speight, for their insightful comments and advice, and for serving on my reading committee.

It has been an honor and a pleasure to be a student in the Department of Risk Management and Insurance (RMI) at Georgia State University. Over the years, I have received help and encouragement from many members of the RMI “community”. Among many, I would like to particularly thank Dr. Richard Phillips, Dr. Daniel Bauer, etc. I also highly appreciate all the help from Ms. Libby Crawley to keep me on the road to graduation.

Without a doubt, my deepest gratitude is to my mother, who has passed away last year, my father, and my parents-in-law, who have been so generous to help me take care of my sweet son during the last four years. And finally, I leave my special thanks to my dear husband, Long Wu, who has been with me to pursue our dreams for almost ten years and will do so in the rest of our lives. Without their support and encouragement, I never would have made it.
Abstract

MOMENT PROBLEMS WITH APPLICATIONS TO VALUE-AT-RISK AND PORTFOLIO MANAGEMENT

BY

Ruilin Tian

May 2008

Committee Chair: Samuel H. Cox
Major Academic Unit: Risk Management and Insurance

My dissertation provides new applications of moment theory and optimization to financial and insurance risk management. In the investment and managerial areas, one often needs to determine some measure of risk, especially the risk of extreme events. However, complete information of the underlying outcomes is usually unavailable; instead one has access to partial information such as the mean, variance, mode, or range. In Chapters 2 and 3, we find the semiparametric upper and lower bounds for the value-at-risk (VaR) with incomplete information, that is, moments of the underlying distribution.

When a single variable is concerned, bounds on VaR are computed to obtain a 100% confidence interval. When the sample financial data have a global maximum, we show that unimodal assumption tightens the optimal bounds. Next we further analyze a function of two correlated random variables. Specifically, we find bounds on the probability of two joint extreme events. When three or more variables are involved, the multivariate problem can sometimes be converted to a single variable problem. In all cases, we use the physical measure rather than the commonly used equivalent pricing probability measure. In addition to solving these problems using the traditional approach based on the geometry of a moment problem, a more efficient method is proposed to
solve a general class of moment bounds via semidefinite programming.

In the last part of the thesis, we apply optimization techniques to improve financial portfolio risk management. Instead of considering VaR, we work with a coherent risk measure, the conditional VaR (CVaR). As an extension of Krokhmal et al. (2002), we impose CVaR-related functions to the portfolio selection problem. The CVaR approach sets a $\beta$-level CVaR as the objective function and maximizes the worst case on the tail of the distribution. The CVaR-like constraints approach adds a set of CVaR-like constraints to the traditional Markowitz problem, reshaping the portfolio distribution. Both methods greatly increase the skewness of portfolios, although the CVaR approach may lose control of the variance. This capability of increasing skewness is very attractive to the investors who may prefer higher probability of obtaining higher returns. We compare the CVaR-related approaches to some other popular portfolio optimization methods. Our numerical analysis provides empirical support for the superiority of the CVaR-like constraints approach in terms of portfolio efficiency.

**Key words:** moment problem, semidefinite programming, semiparametric bounds, maximum entropy, portfolio management, VaR, CVaR
Chapter 1

Introduction and Overview

This is a multi-essay dissertation about moment problems and optimization with applications to risk management, insurance and finance. We compute semiparametric upper and lower bounds on probabilities of rare events, value at risk (VaR) and expected payoffs, subject to empirical moment information. In the first essay (Chapter 2), we analyze moment problems involving one random variable. The second essay (Chapter 3) extends the bound problems to two variables cases. In the third essay (Chapter 4), instead of focusing on the VaR, we make use of the conditional VaR (CVaR) to incorporate the portfolio’s third moment into the traditional mean-variance portfolio selection system. We follow Krokhmal et al. (2002)’s suggestion to improve the skewness of the classical Markowitz portfolios by considering CVaR-related functions, either as an objective function or as one or more additional constraints of the portfolio optimization problem.

In risk management, financial engineering and actuarial science applications, one often needs to determine some measure of risk. The major risk measures people choose include variance, value-at-risk (VaR), expected shortfall and condition VaR (CVaR). Variance denotes the data dispersion through the whole distribution without differentiating the left and the right tails. VaR is a tail risk measurement which is widely applied in quantitative risk management for many types of risk. It is the maximum possible loss over a specified period at a given confidence level. However, VaR does not give any information about the severity of loss by which it is exceeded. In contrast, another tail risk measure, CVaR, designates the magnitude of the tail events by calculating the expected loss that exceeds the VaR. Moreover, compared with VaR, CVaR and expected shortfall are coherent measures which satisfy the properties of monotonicity, sub-additivity, homogeneity and translational invariance.
In this thesis, we focus on analyzing VaR and CVaR. The former is the standard risk measure sanctioned by the Basle Committee although it has certain undesirable theoretical properties. The latter conveys more information about the tail of the distribution and it can be used to manage the third moment of the distribution.

1.1 Moment Problems

In risk assessment, one frequently encounters the situation that the distribution of the interested random variables is unknown. Instead, one only has partial information such as the mean, variance, covariance, skewness, kurtosis, mode and range. In the thesis, we analyze how to use the moment method to measure the tail risk, for example, we obtain a 100% confidence interval on the VaR. Given the moment information and the corresponding support, we find the semiparametric upper and lower bounds on the tail probability. This is the best one can do when the incomplete information consists of estimates of moments. In classical probability theory, these problems are known as "moment problems". They generalize Tchebyshev’s inequality. These types of bounds are usually called semiparametric bounds in the recent related literature. The calculation is based on the physical measure, rather than a pricing or risk neutral measure. That is, bounds for actual or physical probabilities are found. This method offers potential improvements in accuracy and efficiency over the standard approximate methods.

When a single variable is concerned, we provide an optimization framework for computing upper and lower bounds on functional expectations of distributions given moments constraints. These bounds form a 100% confidence interval in which any feasible distribution with same moments is inside. The inverse of the bounds problem solves the value-at-risk (VaR) problem, which finds the upper and lower bounds on $t$ where $\Pr(X \leq t) = \alpha$, subject to moment information on $X$. When financial insurance sample data have a unique global maximum, we can use the unimodal assumption to tighten the optimal bounds. For the univariate moment problems, we use two approaches. In the first, we investigate the mathematics behind the bound problems and solve the problems using the geometry of moment problems. Second, we also provide an efficient method for solving a very general class of moment bounds via semidefinite programming, using some newly developed software such as SOSTOOLS. Furthermore, we use a moment-related method, the maximum-entropy method, to find a representative distribution satisfying the given moment requirements.

Then we go further to analyze bounds on a function of two corrected random variables. The bounds depend on not only the means and variance, but also their covariance. We demonstrate the methodology using three specific applications. The first finds bounds on the probability of a joint extreme events, when two random variables simultaneously take extreme values. We also investigate the bounds on the tail probability of a portfolio consisting of two components. As
the third application, we apply the moment problem to stop-loss payments. The pattern of stop-loss payoffs embraces a class of options such as the call and put options. When more than two variables are involved, a set of random variables is considered as a portfolio and the corresponding semiparametric bound problem is solved by converting it to a one variable problem. In all these applications, we reformulate the corresponding semiparametric bound problem as a sum of squares (SOS) program and use the readily available SOS programming solvers to numerically solve the problems.

The potential usefulness of the moment method is that the incomplete knowledge of distributions is very common, especially the information about the rare events in the tail. Rare events may occur only one or two times in a lifetime – leaving little room to learn from experience. However, in many cases, extreme events contribute a lot to the risks. The extreme events, no matter how rare, could have a profound impact on an individual, a company or even the whole country. Therefore, even in some cases when there are plenty of observations available (e.g., daily price observations), assuming a particular distribution is still perilous if people lack of observations on the extreme events. Moreover, when the distributions of the random variables are assumed to be known, this approach can be implemented to measure the sensitivity of the given probabilities or VaR to model misspecification. That is, the moment method provides not only an initial estimate for cumulative probabilities regardless of any model specifications, but also a mechanism for checking the consistency of models.

1.2 Portfolio Optimization

As we discussed in Section 1.1, the moment method provides a prospective scenario of finding robust bounds which embraces all feasible distributions with specified moments. This approach helps people measure potential risk, especially the tail risk under the condition of incomplete data information. On the other hand, when the starting step of the investment is concerned, one is asked to determine the optimal investment strategy, finding a way to use up the potential of the mean-variance tradeoff and take investor’s risk tolerance into account at the same time. That is, one should consider the third moment (or skewness) of the portfolio. In the last part of this thesis (Chapter 4), we extend the linear programming (LP) and quadratic programming (QP) techniques to improve portfolio risk management.

In 1952, Markowitz (1952) pointed out the tradeoff between the mean and variance of a portfolio. Since then, especially recently, much attention has been focused on asymmetric distributions of the portfolio to fulfill the investors’ special skewness preferences. To address this issue, we extend Krokhmal et al. (2002)’s approach to improve portfolio selection in a three-moment world using a coherent risk measure, the conditional VaR(CVaR). We first analyze the CVaR approach,
which shifts the portfolio distribution to the right by maximizing the conditional VaR of the return. Then we investigate the CVaR-like constraints approach. It reshapes the portfolio distribution by adding CVaR-like constraints to the mean-variance portfolio optimization problem. Adding CVaR-like constraints makes it is possible to increase skewness without significant sacrifice of the tradition Markowitz mean-variance frontier. The CVaR optimization technique has the advantage of reshaping either the left or right tail of a distribution while not significantly affecting the other. When these two approaches are compared with the traditional Markowitz approach, the Boyle-Ding approach, and the mean-absolute deviation (MAD) approach. Our numerical analysis provides empirical support for the superiority of the CVaR-like constraints approach in terms of skewness improvement of mean-variance portfolios. This is very attractive to investors who may prefer higher skewness, or in other words, higher probability of obtaining higher returns.

In a three-moment world of portfolio selection, we avoid solving a double objective optimization problem which minimizes variance and maximizes skewness simultaneously, by setting CVaR as the objective function or by adding CVaR-like constraints. Furthermore, these CVaR-related approaches do not add any additional non-linear constraint to the traditional mean-variance Markowitz portfolio problem. This provides a big advantage in the numerical computation.

In addition to analyzing the classical asset portfolio, we extend our portfolio risk management to the asset-liability portfolio which considers both the asset return of investments and the liability of the financial institutions.

My thesis is organized as follows. In Chapter 2, we introduce the background of the moment problems and some prerequisites, such as the geometry of moment problems, the sum of squares (SOS) programs, the positive semidefinite (PSD) programs, and the duals of the primal, etc. We then solve the univariate moment problems via semidefinite programming as well as the Smith’s approach. Chapter 3 extends our analysis to the semiparametric upper and lower bounds on joint distributions as well as the payoffs with two components. We only focus on solving these problems by reformulating them as sum of squares (SOS) programs and using a SOS programming solver. In Chapter 4, we investigate portfolio risk management in a three-moment world. We utilize a coherent risk measure, CVaR, to impose the investor’s skewness requirements into the traditional mean-variance optimization.
Chapter 2

Optimal Bounds on Value-at-Risk as Solutions to Univariate Moment Problems

The purpose of this chapter is to analyze one variable moment problems. We provide an optimization framework for computing optimal upper and lower bounds on functional expectations of distributions with special properties, given moment constraints. We find the optimal bounds on the “value-at-risk” probability $\Pr(X \leq t) = E[\phi(X)]$, where $\phi(X) = I(-\infty, t](X)$ is the indicator function for the event $X \leq t$, subject to moment constraints $E[X^i] = \mu_i$ for all $i = 1, 2, \ldots, n$. The inverse problems solve $\xi_1$ for the upper bound $\bar{p}(\xi_1) = t$ and $\xi_2$ for the lower bound $\underline{p}(\xi_2) = t$. Then the value-at-risk, $\text{VaR}_t$, has to fall between $\xi_1$ and $\xi_2$, i.e., $\xi_1 \leq \text{VaR}_t \leq \xi_2$. To analyze the sensitivity of the bound estimations with respect to the changes of moments, robustness tests are performed by altering sample sizes. In addition, we use the maximum-entropy technique to obtain a representative distribution based only on the moments and no other information.

2.1 Introduction

In financial engineering and actuarial applications, institutions are interested in the probabilities of extreme events such as catastrophic losses or dramatic price decreases, which can be expressed as value-at-risk of the variables of interest. They frequently encounter situations involving random variables $X$ (with distribution functions $F$) for which they need to determine some measure of risk such as value-at-risk. However, sometimes complete information of the underline distribution or full-possible-range empirical data about the variable of interest is not available, instead one has partial information such as estimates of mean, variance, mode, or range. Therefore, based on incomplete information, one must settle for an approximation of the measure of risk. A feasible effort is to incorporate moment methodology into analysis without distribution assumptions. That
is, calculate the semiparametric upper and lower bounds on distributions or confidence intervals of value-at-risk.

Information about real events is more likely to be incomplete since these things occur just once or twice in a lifetime. However, extreme events may have dramatic influence on the world, which results in increasing interest in tail risk management including managing investment downside risk and insurance catastrophe risk. An example of an extreme event in financial markets comes from the Asian currency crisis in 1997, largely attributed to over-expansion of corporate credit with unhedged short-term borrowing from abroad, large amounts of unproductive capital investment, and speculation on overvalued assets and large trade deficits (Hong, 1998). In the insurance market, insurers are also not free from the impact of catastrophic events, especially large-scale, extreme ones. The total loss of the tragic September 11 terrorist attacks exceeded $80 billion with the insured losses amounting $40.2 billion (Yu and Lin, 2007). In the recent two decades, managing extreme losses caused by catastrophic events like U.S. stock market crash in 1929, hurricanes and earthquakes has been a major concern for market participants. Thus, developing statistical techniques to model extreme events in the area of risk management/insurance and finance is certainly a major task for risk managers.

One of the many problems encountered in forecasting extreme losses is the availability of corresponding loss data. By definition, catastrophic events occur infrequently, and thus, any statistical analysis related to extreme events must deal with tail probability or extreme quantiles of the underlying loss distribution, using only the scare historical data. Traditional statistical methods do not work for such tasks because these methods typically produce a good fit in those regions in which most of the data reside but at the expense of good fit in the tails (Hsieh, 2004).

Accurate determination of tail risk measures based on incomplete data information is impossible. However, one can use the information to obtain bounds on the risk measure. There are many recent approaches applying this approach to the value-at-risk. In classical probability theory, this leads to a Generalized Tchebyshev Inequality (Karlin and Studden (1966); Zuluaga and Peña (2005); and Vandenberghe et al. (2007)). Moment problems generalize the Tchebyshev’s inequalities and provide bounds, given moment information. These bounds are called semiparametric bounds in the recent related literature.

Among the first applications of this approach to practical problems were done by Scarf (1958) (inventory management) and Lo (1987) (mathematical finance). Applications in finance focus on option pricing in the well-known Black and Scholes (1973) setting (Merton, 1973; Levy, 1985; Ritchken, 1985; Schepper and Heijnen, 2007) and other asset pricing and portfolio problems (Ferson and Siegel, 2001, 2003). For example, Lo (1987) gives a closed-form upper bound on the payoff of a European call option when only second-order moment information (i.e., mean and variance) about the underlying asset price at maturity is available. Brockett and Cox (1985), Cox
Brockett et al. (1996) and Roos (2007) apply moment methods in insurance. Bertsimas and Popescu (2005) give a review of the literature and historical perspective on this method, which covers developments from Tchebyshev and Markov in the late 1800s to breakthroughs in the last 10 years.

The most important recent work involves solving the problems using new results on semidefinite programming (Parrillo, 2000; Wolkowicz et al., 2005) to derive semiparametric upper and lower bounds on value at risk (VaR). Following the work of Smith (1990), Cox (1991), Brochett et al. (1995), Zuluaga (2004a), Popescu (2005) and Bertsimas and Popescu (2005), we obtain a range of possible values which contains the risk measure corresponding to every distribution that satisfies the partial information. This range can be considered as a 100% confidence interval.

The common theme here is the use of moments (mean, variance, etc.) as a summarizing description of a probability distribution. In this chapter, we show how to compute the semiparametric upper and lower bounds on \( \Pr(X \leq t) \), where \( X \) is a single random variable, given the moments of the distribution of \( X \). In section 2.3, we consider the case that no additional constraint is added. We call them arbitrary bounds. In order to numerically solve for the semiparametric bounds, we reformulate the corresponding semiparametric bound problem as a sum of squares (SOS) program and use the readily available SOS programming solvers such as SOSTOOLS (Prajna et al. (2002)), GloptiPoly (Henrion and Lasserre (2003)), or YALMIP (Löfberg (2004)). Smith (1990) developed an alternative approach based on the geometry of the moment problem. This approach involves the construction of a discrete distribution with the given moments.

The arbitrary bounds may be improved if we have more information. For example, when the underlying distribution is unimodal, we get better bounds. Therefore, in Section 2.4, we add the unimodal assumption and find the narrower upper and lower bounds given the same moments.

In Section 2.5, we discuss the method of constructing representative distributions to match given moments using the maximum-entropy approach (N. Agmon and Levine (1979)). Finally, in Section 2.7, we test the sensitivity of the bounds with respect to the data sample size. Section 2.8 concludes the chapter.

### 2.2 Preliminaries

The analytic foundation for the methods developed here comes from the classical moment problem. The moment problem was first studied by Tchebyshev, Markov and Stieltjes in the 1870’s. They formulated and solved many variations on what Stieltjes called the “problem of moments”.

problem is to determine a distribution function $F(x)$ with a prescribed set of moments:

$$\int_{x=-\infty}^{+\infty} x^i dF(x) = \mu_i, \text{ for all } i = 0, 1, \ldots,$$

where the values of $\mu_1, \mu_2, \ldots$ are moments. In the rest of this chapter, we consider the problem of determining bounds on $E[\phi(X)]$, the expectation of an arbitrary function given some moments of the underlying distribution.

### 2.2.1 Moment Problems

A moment problem is an optimization problem with the form:

$$\max (\text{or } \min) \ E[\phi(X)]$$

where $X$ is a set of random variables with specified support and moments.

For example, the Tchebyshev’s inequality can be considered as a moment problem (Lindgren, 1993, p.132). If $X$ has mean $\mu$ and variance $\sigma^2$, then $\Pr(|x - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$. This can be restate as an optimization problem and its solution. The problem is stated as follows:

$$\max_X \ E[\phi(X)]$$

where the support is over all $X$ subject to

$$E[X] = \mu,$$
$$E[(X - \mu)^2] = \sigma^2,$$

where

$$\phi(x) = \begin{cases} 
1 & \text{if } |x - \mu| \geq k\sigma, \\
0 & \text{if } |x - \mu| < k\sigma.
\end{cases}$$

The solution is $\frac{1}{k^2}$. This means that $\Pr(|x - \mu| \geq k\sigma) = E(\phi(X)) \leq \frac{1}{k^2}$.

Smith (1990, p.23) provides the following summaries to determine whether or not a given sequence of numbers are the moments of some probability distribution.

Define the $(n + 1) \times (n + 1)$ moment matrix $M_{2n}$ as follows:

$$M_{2n} = \begin{bmatrix}
\mu_0 & \mu_1 & \ldots & \mu_n \\
\mu_1 & \mu_2 & \ldots & \mu_{n+1} \\
\vdots & \vdots & \ddots & \vdots \\
\mu_n & \mu_{n+1} & \ldots & \mu_{2n}
\end{bmatrix}_{(n+1) \times (n+1)}$$

(2.2)
(i) The sequence \( \{ \mu_0, \mu_1, \ldots, \mu_{2n} \} \) represents moments of some probability distribution if and only if the moment matrix \( M_{2n} \) is positive semidefinite (PSD). That is, there is a random variable with the given moments if and only if \( M_{2n} \) is PSD.

(ii) In order for the sequence \( \{ \mu_0, \mu_1, \ldots, \mu_{2n} \} \) to be the moments of a distribution with more than \( n \) points of its support \(^1\), it is necessary and sufficient that \( M_{2n} \) is positive definite. In this case, there are infinite number of distributions that match the given sequence of moments and we call the sequence \( \{ \mu_0, \mu_1, \ldots, \mu_{2n} \} \) non-degenerate.

(iii) In addition, \( \{ \mu_0, \mu_1, \ldots, \mu_{2n} \} \) are the moments of a distribution with exactly \( n \) points of support if and only if \( |M_{2i}| > 0 \) for all \( i = 1, 2, \ldots, n - 1 \) and \( |M_{2i}| = 0 \) for \( i = n \). In this case, the distribution is uniquely determined and the sequence is \( \{ \mu_0, \mu_1, \ldots, \mu_{2n} \} \) degenerate.

Notice that \( \mu_0 \) is always 1 because the possibilities sum to 1. Therefore, we can test whether the problem has solution by checking whether the moment matrix \( M_{2n} \) is PSD. When a moment problem has solution, we say it is feasible.

In this chapter, we consider the moment problem of finding optimal bounds on \( \mathbb{E}[\phi(X)] \) subject to constraints \( \mathbb{E}[g_i(X)] = \mu_i \) for \( i = 1, 2, \ldots, n \). In general, \( X \) is a vector of random variables, but here we are only considering the univariate case, i.e., \( X \) is a single random variable. In Chapter 3, bounds on joint distributions with two random variables will be discussed. In addition, we are considering only the “classical” case for which \( g_i(x) = x^i \) for all \( i = 1, 2, \ldots, n \). To calculate the bounds on cumulative distribution function (CDF) \( \Pr(X \leq t) \), \( \phi(x) \) is set as the indicator function \( \mathbb{I}_{(-\infty,t]}(x) \) for a fixed \( t \in \mathcal{I} \), where \( \mathcal{I} \) is the support of \( x \). That is:

\[
\phi(x) = \begin{cases} 
1 & \text{for all } x \leq t \\
0 & \text{for all } x > t 
\end{cases} \quad x \in \mathcal{I}.
\]

(2.3)

The general primal problem for the upper bound can be expressed as follows:

\[
\overline{p} = \max \mathbb{E}_F[\phi(X)]
\]

where the support is over all \( X \) subject to
\[
\mathbb{E}_F[X^i] = \mu_i \quad \text{for all } i = 1, \ldots, n,
\]

and (perhaps) \( X \) is unimodal with mode \( m \),

\( F(x) \) a probability distribution on \( \mathcal{I} \),

where \( \overline{p} \) denotes the optimal solution of the problem. The support \( \mathcal{I} \) and the moments \( \mu = \)

\(^1\)A point \( x \) is a point of support of a distribution if it is a point of increase of \( F \), i.e., if for any \( a \) and \( b \) with \( a < x < b \), then \( F(a) < F(b) \).
[µ₁, µ₂, ..., µₙ] are given. In our work, the support interval I can be one of the following four choices, i.e., I = (−∞, a], I = [a, b], I = [b, +∞) and I = (−∞, +∞). In addition to moments, sometimes there is an additional constraint such as X is unimodal distributed with a given mode or X is symmetric about a given value (Popescu, 2005). It turns out that the numerical methods (semidefinite or linear programming) apply when gi(x) is a piecewise polynomial. Piecewise polynomial means there is a decomposition of I into a finite number of disjoint subintervals and gi(x) is a polynomial on each subinterval. We are considering only the classical case and we are focusing our attention on the value-at-risk (VaR). However, the setting in which gi(x) and φ(x) are more general is worth keeping in mind.

Write the primal problems in (2.4) with classical moment constraints as follows:

\[
p = \max \int_{I} \phi(x) \, dF(x)
\]
\[
\text{subject to } \int_{I} x^i \, dF(x) = \mu_i, \text{ for all } i = 1, 2, \ldots, n
\]  
(2.5)

The lower bound problem is an analogue, except for that the objective function is:

\[
p = \min \int_{I} \phi(x) \, dF(x),
\]  
(2.6)

with the same constraints as (2.5).

### 2.2.2 Dual Problems

Since the primal problem (2.5) (or (2.6)) is difficult to solve directly, we try to solve its complementary problem, the dual problem. A solution to either the primal or dual determines a solution to both.

Karlin and Studden (1966, Chapter XII, p.476) prove that the dual problem of the program (2.5) can be written as follows:

\[
d = \min \sum_{i=0}^{n} a_i \mu_i
\]
\[
\text{subject to } p(x) \geq \phi(x), \text{ for all } x \in I,
\]  
(2.7)

and correspondingly, the dual of the lower bound problem is

\[
d = \max \sum_{i=0}^{n} a_i \mu_i
\]
\[
\text{subject to } p(x) \leq \phi(x), \text{ for all } x \in I,
\]  
(2.8)

where p(x) = ∑₀ⁿ aᵢxᵢ.
It is easy to see that weak duality holds between $\bar{p}$ and $\bar{d}$ (or $\underline{p}$ and $\underline{d}$) (Chvatal, 1983, p.139); that is, the feasible solution to the dual yields a bound on the optimal value of the primal:

$$\bar{p} \leq \bar{d} \text{ (or } \underline{p} \geq \underline{d}).$$

If problems (2.5) is feasible and there exist $a_0, a_1, \ldots, a_n$ such that

$$\sum_{i=0}^{n} a_i x^i > \phi(x), \text{ for all } x \in I,$$

then the strong duality holds; that is, $\bar{p} = \bar{d}$. In the analog for the problem (2.6), we reverse the inequality and replace $\bar{p} = \bar{d}$ with $\underline{p} = \underline{d}$. Zuluaga and Peña (2005, Proposition 3.1) show that this follows the convex duality.

Throughout the whole chapter, $\phi(x)$ is an indicator function bounded in $[0, 1]$. Therefore, for the upper bound problem, the dual solution $a_0 > 2$, and $a_i = 0$ for all $i \neq 0$ strictly satisfies (i.e., with $>$) the constraint in (2.7) for all $x \in I$. And for the lower bounds problem, the dual solution $a_0 < 0$, and $a_i = 0$ for all $i \neq 0$ strictly satisfies (i.e., with $<$) the constraint in (2.8) for all $x \in I$. So as long as the problem (2.5) (or (2.6)) is feasible, $\bar{p} = \bar{d}$ (or $\underline{p} = \underline{d}$).

Now, let’s provide some geometric explanation to the conversion from the primal problems to their dual problems. As a special case of the development of Kemperman (1987), given a non-degenerate sequence of moments $\{\mu_0, \mu_1, \ldots, \mu_n\}$, one can construct a discrete distribution to match the given moments of any continuous distribution. After that, according to Smith (1990), one can calculate the value of the objective function (in general, the expectation of a piecewise polynomial), based on the support points and their corresponding probabilities that are determined by the discrete distribution. In general, one is given the freedom to choose $k$ points and $k$ probabilities to satisfy the $2k$ conditions posed by the requirement of matching $\mu_0, \mu_1, \ldots, \mu_{2k-1}$. Therefore, to match $n$ moments, a polynomial of degree $\frac{n+1}{2}$ should be constructed for the purpose of finding $\frac{n+1}{2}$ points of support. The construction of that polynomial is not unique, but one can construct the same degree orthogonal polynomials to guarantee uniqueness. Details of constructing orthogonal polynomials are discussed in Section 2.3.2. Here we only focus on the existence of the polynomials with the required degree.

For the upper bound problem, suppose there is a polynomial of degree $l = \frac{n+1}{2}$, $h(x) = \sum_{i=0}^{l} a_i x^i$, for which $h(x) \geq \phi(x)$ for all $x \in I$. Let $Z$ denote the contact set of $h(x)$, which is defined by

$$Z = \{x \in I : h(x) = \phi(x)\}.$$

Given the support $I$ and moments $\mu = [\mu_1, \mu_2, \ldots, \mu_n]$, let $\pi(\mu)$ denote the set of all cumulative
distributions \( F \) with support in \( I \) for which \( E_F[g_i(X)] = \mu_i \) for all \( i = 1, \ldots, n \). Now, assume that there is a cumulative distribution \( G \) in \( \pi(\mu) \) with its support entirely within the contact set \( Z \); that is:

\[
\int_Z dG(x) = 1.
\]

For any cumulative distribution \( F \in \pi(\mu) \), we have the following relations (Cox, 1991):

\[
E_G[\phi(X)] = \int_I \phi(x) dG(x) = \int_Z \phi(x) dG(x) = \int_Z h(x) dG(x) = \sum_{i=0}^l a_i \int_Z x^i dG(x) = \sum_{i=0}^l a_i \mu_i
\]

\[
\geq \int_I \phi(x) dF(x)
\]

(2.9)

Therefore, \( E_G[\phi(X)] \) is the smallest upper bound, i.e., \( \overline{p} = E_G[\phi(X)] \). Kemperman (1987, p. 36) shows that such a polynomial \( h(x) \) always exists. So to calculate the upper bound \( \overline{p} \), one only needs to determine \( h(x), Z \) and \( G \). Similarly, to determine the lower bound \( \underline{p} \), one should analyze the polynomial \( h(x) \) for which \( h(x) \leq \phi(x) \) on the support.

In our problem, since the objective function is \( E[\phi(X)] \) where \( \phi(x) \) is the indicator function \( \mathbb{I}_{(-\infty,t]}(x) \), we are interested in constructing discrete distributions that include one particular point of support \( t \). In general, if \( m \) points of support are included in advance, one should construct a \( k \) points (\( m \) pre-given points included) discrete distribution to match the first \( 2k - m - 1 \) moments, \( \mu_1, \mu_2, \ldots, \mu_{2k-m-1} \). Therefore, when we are calculating bounds on the probability \( \Pr(X \leq t) \), we construct \( k \) points (\( t \) is included) of a discrete distribution to match \( 2k - 2 \) moments. Let us analyze the following example to illustrate the geometry of the relationship between the primal and dual problems. For example below, \( 2k - 2 = 4 \), so we need to construct a polynomial with \( k = 3 \) support points.

**Example 1.** Consider the upper bound on arbitrary distributions (without any additional assumption such as unimodality, symmetry, etc.) given the first four raw moments, i.e., the non-degenerate
sequence \( \{\mu_0, \mu_1, \ldots, \mu_4\} \), maximizing the value of

\[
\int_{\mathcal{I}} \phi(x) \, dF(x)
\]

over all cdf \( F(x) \) with support in \( \mathcal{I} \) subject to

\[
\int_{\mathcal{I}} x^i \, dF(x) = \mu_i, \quad \text{for } i = 0, 1, \ldots, 4,
\]

where \( \phi(x) = 1 \) for \( x \leq t \) and \( \phi(x) = 0 \) for \( x > t \) and \( \mathcal{I} = (-\infty, +\infty) \).

Recall that the dual (Chvatal, 1983, p.140) of the classical primal problem with equality constraints

\[
\begin{align*}
\text{max} & \quad c^T x \\
\text{subject to} & \quad G_i x = b_i, \quad \text{for all } i = 1, 2, \ldots, m;
\end{align*}
\]

is defined to be the problem

\[
\begin{align*}
\text{min} & \quad b^T y \\
\text{subject to} & \quad G_i^T y = c_j, \quad \text{for all } j = 1, 2, \ldots, n,
\end{align*}
\]

where \( c \in \mathbb{R}^{n \times 1}, b \in \mathbb{R}^{m \times 1}, x \in K = \mathbb{R}^{n 	imes 1}, y \in K^* = \mathbb{R}^{m \times 1}, G \in \mathbb{R}^{m \times n} \) with \( G_i \) the \( i \)-th row of the \( G \)-matrix, and \( G_j^T \) is the \( j \)-th row of \( G^T \). Here, \( K \) is a closed convex cone and \( K^* \) denotes the dual cone of \( K \).

For Example 1, let us construct an orthogonal polynomial\(^3\) of degree 3, \( h(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \), for which \( h(x) \geq \phi(x) \) for all \( x \in \mathcal{I} \). From Figure (2.1), we can see that there are 3 support points, which are the roots of the equation \( h(x) = \phi(x) \). Let \( x_1, x_2, x_3 \) denote these three support points and let \( p_1, p_2, p_3 \) denote their respective probability masses. Note that the support point \( x_2 \) is \( t \); that is \( x_2 = t \). Then problem (2.10) can be written as

\[
\begin{align*}
\text{max} & \quad \phi(x_1)p_1 + \phi(x_2)p_2 + \phi(x_3)p_3 \\
\text{subject to} & \quad x_1^i p_1 + x_2^i p_2 + x_3^i p_3 = \mu_i, \quad \text{for } i = 0, 1, \ldots, 4.
\end{align*}
\]

Set \( x^T = [p_1, p_2, p_3], c^T = [\phi(x_1), \phi(x_2), \phi(x_3)], b^T = [\mu_0, \mu_1, \ldots, \mu_4], y^T = [a_0, a_1, \ldots, a_4] \)

\(^3\)The cone in this thesis means specifically a convex cone; that is, a subset of a vector space that is closed under linear combinations with positive coefficients. Let \( C \subset V \) be a convex cone in a real vector space \( V \) equipped with a scalar product. A dual cone to \( C \) is a set

\[ \{v \in V \mid \text{for all } w \in C, (w, v) > 0\} . \]

This is also a convex cone.

\(^3\)Any polynomial of degree 3 is fine. The requirement of orthogonality guarantees its uniqueness.
Figure 2.1. \( h(x) \) is a cubic polynomial. \( h(x) = \phi(x) \) at exactly three points, \( x_1, x_2, x_3 \), and \( h(x) \geq \phi(x) \) on the support \( \mathcal{I} \).

and \( G \) as follows

\[
G = \begin{pmatrix}
1 & 1 & 1 \\
x_1 & x_2 & x_3 \\
\vdots & \vdots & \vdots \\
x_1^4 & x_2^4 & x_3^4
\end{pmatrix}_{5 \times 3}
\]

The dual of problem (2.13) can be written as follows: Minimize

\[
\sum_{i=0}^{4} a_i \mu_i
\]
over all \( a_0, a_1, a_2, a_3, a_4 \) subject to

\[
\begin{bmatrix}
1 & x_1 & \ldots & x_1^4 \\
1 & x_2 & \ldots & x_2^4 \\
1 & x_3 & \ldots & x_3^4 \\
\end{bmatrix}
\begin{bmatrix}
a_0 \\
a_1 \\
a_2 \\
a_3 \\
a_4
\end{bmatrix}
= \begin{bmatrix}
h(x_1) + a_4 x_1^4 \\
h(x_2) + a_4 x_2^4 \\
h(x_3) + a_4 x_3^4
\end{bmatrix} = \begin{bmatrix}
\phi(x_1) \\
\phi(x_2) \\
\phi(x_3)
\end{bmatrix},
\]

for all \( x_1, x_2, x_3 \) in \( \mathcal{I} \).

According to our construction process, in the contact set \( \mathcal{Z} \), any support point \( x_i \) for \( i = 1, 2, 3 \) satisfies \( h(x_i) = \phi(x_i) \). So we get \( a_4 = 0 \) since not all roots of \( h(x) \) are zero. The constraints of (2.14) hold for the points in the contact set \( \mathcal{Z} \) for which \( h(x) = \phi(x) \). In general, \( h(x) \geq \phi(x) \). Therefore, if we replace \( x \) with \( x_i \) for \( i = 1, 2, 3 \), we get the same constraint as in problem (2.7);
that is:
\[
\sum_{i=0}^{4} a_i x^i \geq \phi(x).
\]

Therefore, we can summarize the relationship between \( p(x) \) in the dual problems and \( h(x) \) as follows:

Given the first \( n \) moments, we denote \( p(x) = \sum_{i=0}^{n} a_i x^i \) as the polynomial on the left hand side of the dual problem (2.7) and (2.8). \( h(x) \) is constructed as a polynomial of degree \( k \) with \( k = \frac{n+1}{2} \) if no pre-specified point is included or \( k = \frac{n+m+1}{2} \) if the problem is given \( m \) points in advance. When \( p(x) \) and \( h(x) \) are properly constructed \(^4\), we have \( p(x) = h(x) \) with \( a_i = 0 \) for all \( k < i \leq n \).

**2.2.3 Sum of Squares**

Denote
\[
p(x) = p(x_1, \ldots, x_n) = \sum_{i_1, \ldots, i_n \in \mathbb{N}} a_{(i_1, \ldots, i_n)} x_1^{i_1} \cdots x_n^{i_n}
\]
a polynomial of degree \( m \), where \( \max \{ \sum_{j=1}^{n} i_j, m \} = m \). Given a cone \( D \subseteq \mathbb{R}^n \), if \( p(x) \) satisfies \( p(x) \geq 0 \) for all \( x = [x_1, \ldots, x_n] \in D \), then \( p(x) \) is a positive semidefinite (PSD) polynomial on \( D \). If
\[
p(x) = \sum_i [q_i(x)]^2
\]
for some polynomials \( q_i(x) = q_i(x_1, \ldots, x_n) \), then \( p(x) \) is a sum of squares (SOS) polynomial on \( D \) with \( D \subseteq \mathbb{R}^n \). Obviously, SOS is a sufficient condition of PSD.

More than a century ago, David Hilbert proved that not every PSD polynomial is SOS. To check whether a polynomial is a sum of square polynomial, one applies the sum of square decomposition. It is recently presented as the “Gram matrix” method. The method is implemented as follows (Powers and Wörmann, 1998): Express the given polynomial as a quadratic form in some new variables \( z \). These new variables are original \( x \) ones, plus all the monomials of degree less than or equal to \( \frac{m}{2} \) given by the different products of the \( x \) variables. Therefore, \( p(x) \) can be represented as:
\[
p(z) = z^T Q z, \tag{2.15}
\]
where \( Q \) is a constant matrix. If \( Q \) is positive semidefinite, \( p(x) \) is positive semidefinite polynomial. Since the variables \( z_i \) are not independent, the representation (2.15) might not be unique,

\(^4\) The construction of \( p(x) \) and \( h(x) \) is not unique, but the contract sets \( \{ x \in I : h(x) = \phi(x) \} \) and \( \{ x \in I : p(x) = \phi(x) \} \) are equal.
and \( Q \) may be PSD for some representations but not for others. Actually, there is a linear subspace of matrices \( Q \) that satisfy (2.15). If the intersection of this subspace with the positive semidefinite cone is nonempty, then the original polynomial \( p(x) \) is guaranteed to be SOS (and therefore PSD).

For some special cases, the equality between PSD and SOS holds. Hilbert (1888) gave the following theorem:

**Theorem 1** (Hilbert (1888)). A PSD polynomial on \( \mathbb{R}^n \) is SOS if and only if one the following conditions is satisfied:

1. Polynomial with one or two variables \((n \leq 2)\);
2. Quadratic polynomial \((m = 2)\), where the sum of squares decomposition follows from eigenvalue/eigenvector factorization;
3. Quartic polynomial with three variables \((m = 3, n = 4)\).

We will apply case (1) of Hilbert’s Theorem (Theorem 1) to solve univariate moment problems by using a SOS programming solver. Note that Theorem 1 holds on \( \mathbb{R}^n \). When a moment problem with support \( \mathcal{D} \subseteq \mathbb{R}^n \) is considered, we will use the concept of copositive matrix to convert it to some solvable SOS programs.

A matrix \( Q \in \mathbb{R}^{n \times n} \) is copositive if \( x^T Q x \geq 0 \) for all \( x \in \mathbb{R}^n, x_i \geq 0 \). Equivalently, the quadratic form is nonnegative on the closed nonnegative orthant. If \( x^T Q x \) takes only positive values on the closed orthant (except the origin, of course), then \( Q \) will be strictly copositive. Parrilo (2000, p.62) points out that to check copositivity of \( Q \), one can consider the change of variables \( x_i = z_i^2 \), and study the global nonnegativity of

\[
p(z) = z^T Q z = \sum_{i,j} m_{ij} z_i^2 z_j^2,
\]

where \( z = [z_1^2, z_2^2, \ldots, z_n^2]^T \). \( Q \) is copositive if and only if \( p(z) \) is PSD. Therefore, a sufficient condition for \( Q \) to be copositive is that \( p(z) \) can be written as a SOS.

The theorem below proposed by Diananda (1962) is relevant to our following discussion. Here, we present the theorem in a form that will be suitable for our purposes, instead of presenting it in its original form. Parrilo (2000) and Zuluaga (2004b) prove the equivalence between the original version of Diananda’s Theorem and Theorem 2 below.

**Theorem 2** (Diananda (1962)). Let \( p(x_1, \ldots, x_n) \) be a polynomial of degree \( m \) with the quadratic form \( z^T Q z \), where \( z \) contains original \( x \) and all the monomials of \( x \) of degree less than and equal to \( \frac{m}{2} \). If the number of variables \( n \leq 3 \), then \( p(x_1, \ldots, x_n) \geq 0 \), for all \( x_1, \ldots, x_n \geq 0 \) if and only if \( p(x_1^2, \ldots, x_n^2) \) is a SOS polynomial.

Therefore, to check if a univariate polynomial \( p(x) \) is positive on \( \mathcal{I} \) with
(1) $\mathcal{I} = (-\infty, a]$,  
(2) $\mathcal{I} = [a, b]$,  
(3) $\mathcal{I} = [b, \infty)$,

one can substitute $x$ with 
(1) $x = a - x'$,  
(2) $x = a + x'$ for all $x \geq a$, and $x = b - x'$ for all $x \leq b$,  
(3) $x = b + x'$,

to check whether $p(x'^2)$ is a SOS polynomial.

### 2.2.4 SOS Programming

For our univariate moment problem, case (1) of Hilbert Theorem (Theorem 1) applies. If $p(x)$ is PSD on $\mathbb{R}$, it is SOS on $\mathbb{R}$ as well. Note that the constraints of the dual problems (2.7) and (2.8) are PSD (and therefore SOS) constraints. The upper bound problem has the constraint $p(x) - \phi(x) \geq 0$, for all $x \in \mathbb{R}$ and the lower bound problem requires $\phi(x) - p(x) \geq 0$, for all $x \in \mathbb{R}$. Therefore the upper bound (or low bound) problem reduces to solving a semidefinite program, so long as the problem has a solution (feasible).

This semiparametric bound problem is a sum of squares (SOS) program and can be solved by SOS programming solvers such as SOSTOOLS, GloptiPoly, or YALMIP. A SOS program is an optimization program where the variables are coefficients of polynomials, the objective is a linear combination of the variable coefficients, and the constraints are given the polynomials being SOS.

It is worth mentioning that any SOS program can be reformulated as a semidefinite program (SDP) (Todd (2001), Parrillo (2000)). Semidefinite optimization problems are linear programs with linear matrix inequality (LMI) constraints, i.e., positive semidefinite constraints on matrices of variables. Bertsimas and Popescu (2002) provide an efficient method for solving a very general class of moment bounds via semidefinite programming. In fact, SOS programming solvers work by reformulating the SOS program as a SDP, and then using SDP solvers such as SeDuMi (Sturm (1999)) to solve it. However, SDP formulations of SOS programs are typically fairly involved. Thus for clarity purposes and to make it easy to reproduce our results, throughout our work we use SOS programming tools instead of directly reformulating the problem as a SDP.

### 2.2.5 Optimal Bounds on Value at Risk

The value at risk (VaR) problem is to find the upper and lower bounds on $t$ where $\Pr(X \leq t) = \alpha$, subject to moment information on $X$. We connect this to a semiparametric probability problem by finding bounds on $\Pr(X \leq t)$ for enough values of $t$’s to solve the inverse problem.
CHAPTER 2. UNIVARIATE MOMENT PROBLEM

Figure 2.2. Relationship between bounds and VaR. The upper curve represents the upper bound, i.e., \( \overline{p}(x) \). The lower curve represents the lower bound, i.e., \( \underline{p}(x) \). The curve in the middle is the true empirical distribution, i.e., \( F(x) \).

The solution \( d = p \) (or \( d = \underline{p} \)), which equals \( a_0 + a_1 \mu_1 + \cdots + a_n \mu_n \) is the upper bound (or lower bound) on \( \Pr(X \leq t) \) for all random variable \( X \) with the given moments and support. As \( t \) varies over the support \( \mathcal{I} \) of \( X \), the solution values \( \{a_0, a_1, \ldots, a_n\} \) varies as well. Therefore, the bounds are functions of \( t \), i.e., \( \overline{p}(t) \) (or \( \underline{p}(t) \)). Both \( \overline{p}(t) \) and \( \underline{p}(t) \) are actual distribution functions. They are increasing functions of \( t \) which tend to be 1 as \( t \to \infty \) and tend to 0 as \( t \to -\infty \).

As showed in Figure 2.2, the bounds \( \overline{p}(x) \) and \( \underline{p}(x) \) on the cumulative distribution function correspond to bounds on the value at risk. Consider \( \Pr(X \leq t) = 0.8 \). \( F(\xi_1) \leq \overline{p}(\xi_1) = 0.8 = \underline{p}(\xi_2) \leq F(\xi_2) \). Therefore, \( \text{VaR}_{0.8} \), which is 80% value at risk, is within the 100% confidence interval \([\xi_1, \xi_2]\). In general, for a given probability \( \alpha \), the corresponding value at risk is bounded by \( \xi_1 \) at the \( \alpha \)-th percentile of \( \overline{p} \) and by \( \xi_2 \) at the \( \alpha \)-th percentile of \( \underline{p} \). That is, for any distribution \( F \in \pi(\mu) \), the \( \alpha \)-level VaR, which is denoted \( \text{VaR}_\alpha \), of \( F(x) \) is between \( \xi_1 \) and \( \xi_2 \). That is

\[
F(\xi_1) \leq \overline{p}(\xi_1) = \alpha = \underline{p}(\xi_2) \leq F(\xi_2)
\]

Solving the inverse functions for \( \overline{p}(x) \) and \( \underline{p}(x) \), we have \( \xi_1 \leq \text{VaR}_\alpha \leq \xi_2 \), a 100% confidence interval for the \( \alpha \)-level VaR.
2.3 Moment Bounds for Arbitrary Distributions

As for arbitrary distributions, we mean we are considering distributions with given moments and support, with no additional information.

If the primal problem is feasible and strong duality holds (see page 11), the dual problem is equivalent to its primal in the sense that the numerical solution to the dual is equal to that of its primal. Therefore instead of solving problems (2.5) and (2.6) directly, we solve their duals (2.7) and (2.8).

In this section, we use two different approaches to solve the arbitrary bounds problems. The first method from Bertsimas and Popescu (2002) can solve a very general class of moment bounds via semidefinite program (SDP). In our paper, we use SOS program solvers with more friendly interface to compute the semiparametric bounds. We call this the SOS approach. The second method constructs discrete approximations to match given moments and calculates bounds based on those moment-matching discrete distributions. The later approach is based on the geometry of moment problems (Kemperman, 1987). A distinguished example is Smith (1990)’s Ph.D. thesis. He presents a method for discretizing distributions to match as many moments as possible and applies it to decision analysis. In the rest of this chapter, we will call the second method Smith’s approach.

2.3.1 SOS Approach

The dual problem falls in a class of optimization problems called semidefinite problems. This class is analogous to linear programming problems, but the inequality constraint applies over a continuum rather than a finite set.

First, let us consider the upper bound problem (2.7). The inequality constraint $p(x) \geq \phi(x)$ with $\phi(x) = 1$ for $x \leq t$ and $\phi(x) = 0$ for $x > t$ is equivalent to two simultaneous inequalities:

$$
p(x) - 1 \geq 0 \quad \text{for all } x \in (-\infty, t]
$$
$$
p(x) \geq 0 \quad \text{for all } x \in (t, \infty),
$$

(2.16)

where $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$.

Applying the substitution $x \rightarrow t - x', x \rightarrow t + x'$ to the first and second constraints of (2.16) respectively, it follows that (2.16) is equivalent to:

$$
p(t - x') - 1 \geq 0 \quad \text{for all } x' \geq 0
$$
$$
p(t + x') \geq 0 \quad \text{for all } x' \geq 0
$$

(2.17)

Now applying Diananda’s Theorem (Theorem 2), problem (2.7) is equivalent to the following
SOS program:
\[
\tilde{d} = \min \sum_{i=0}^{n} a_i \mu_i \\
\text{subject to} \\
[p(t - x^2) - 1] \text{ is a SOS polynomial} \\
p(t + x^2) \text{ is a SOS polynomial.}
\] (2.18)

Notice that above we drop the primes in the variable labels. The SOS program (2.18) can be readily solved with a SOS programming solver. Thus, as long as the problem (2.5) is feasible (page 8), we can obtain the semiparametric upper bound \( \tilde{p} \) by numerically solving problem (2.18) with a SOS solver.

Now we solve the lower bound problem (2.8). The inequality constraint is equivalent to two simultaneous inequalities:
\[
1 - p(x) \geq 0 \quad \text{for all } x \in (-\infty, t] \\
-p(x) \geq 0 \quad \text{for all } x \in (t, \infty),
\] (2.19)

The equivalent constraints are
\[
\tilde{d} = \max \sum_{i=0}^{n} a_i \mu_i \\
\text{subject to} \\
1 - p(t - x^2) \text{ is a SOS polynomial} \\
-p(t + x^2) \text{ is a SOS polynomial.}
\] (2.20)

Once again one can consider them as SOS constraints and solve the problem by SOS programming solver. We also can solve the lower bound problem by considering its upper bound counterpart. That is, the lower bound can be obtained by solving a transformed upper bound problem. Although we state the upper and lower bound problems as two problems, a complete solution to one is sufficient to solve the other. By this we mean that, if we have a method of solving all upper bound problems, we can solve all lower bound problems.

The lower bound on \( \Pr(X \leq t) \) can be obtained from the upper bound on \( \Pr(X > t) \). Specifically, the lower bound \( \underline{p}(t) \) on \( \Pr(X \leq t) = \mathbb{E}[[(-\infty,t](X)] \) can be found by finding the upper bound on its complement \( \psi(X) = 1 - \phi(X) \). If \( \overline{p} = \max \{ \mathbb{E}_F[\psi(X)] : F \in \pi(\mu) \} \) where \( \psi(x) = \mathbb{I}_{(t,\infty)}(x) \), then we have \( p(t) = 1 - \overline{p}(t) \)

\(^5\)If the SOS programming solver SOSTOOLS (which calls semidefinite program solvers such as SDP or SeDuMi) is used, we can avoid the process of reformulating the constraints in (2.16) to copositive constraints.
2.3.2 Smith’s approach

Smith’s approach constructs discrete distributions to match the moments of the underlying distribution. This method is a result of the duality between the moments of a distribution and the polynomials whose expectations are defined by these moments. The usefulness of moments as a summarizing description of a probability distribution is related to the effectiveness of polynomial interpolation and polynomial approximation.

With moment matrix $M_{2n}$ defined as in (2.2), the “pseudo-expectation” can be defined as follows:

$$\langle h(x), q(x) \rangle = [a_0, a_1, \ldots, a_n] \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

where $\{a_0, a_1, \ldots, a_n\}$ and $\{b_0, b_1, \ldots, b_n\}$ denote the coefficients of the polynomials $h(x)$ and $q(x)$ respectively. Notice that the expectation $\langle h(x), q(x) \rangle$ is not the real expectation under the measure of the real distribution. It is an expectation defined only by the moments, therefore, we call it the pseudo-expectation.

Now let us define orthogonal and orthonormal polynomials with respect to pseudo-expectation. Two polynomials $h(x)$ and $q(x)$ are said to be orthogonal if $\langle h(x)q(x) \rangle = 0$. A polynomial $h_k(x)$ of degree $k$ is called a rank $k$ orthogonal polynomial if it is orthogonal to all polynomials of degree less than $k$. An orthogonal polynomial $h_k^*(x)$ is called orthonormal if $\langle h_k^*(x) \rangle = 1$ and the leading coefficient of $h_k^*(x)$ is positive.

Given a non-degenerate sequence of moments $\{\mu_0, \mu_1, \ldots, \mu_n\}$, a sequence of orthogonal polynomials $h_0(x), h_1(x), \ldots, h_k(x)$, where $k = \frac{n}{2}$, and a new sequence of orthonormal polynomials $h_0^*(x), h_1^*(x), \ldots, h_k^*(x)$ is uniquely determined as follows

$$h_{i+1}(x) = (x - \langle xh_i^*(x), h_i^*(x) \rangle) h_i^*(x) - \langle h_i(x), h_i(x) \rangle^{1/2} h_{i-1}^*(x) \quad \text{for } 1 \leq i \leq k - 1,$$

where

$$h_i^*(x) = \frac{h_i(x)}{\langle h_i(x), h_i(x) \rangle^{1/2}}.$$

The orthogonal polynomial $h_k(x)$ can also be obtained as the determinant of the following
matrix:

\[
\begin{vmatrix}
\mu_0 & \mu_1 & \cdots & \mu_k \\
\mu_1 & \mu_2 & \cdots & \mu_{k+1} \\
\vdots & \vdots & \ddots & \vdots \\
\mu_{k-1} & \mu_k & \cdots & \mu_{2k-1} \\
1 & x & \cdots & x^k
\end{vmatrix}
\]

(2.23)

According to the construction process in (2.23), the definition of \( h_k(x) \) is based on knowledge of the first \( n-1 \) moments rather than the full sequence of \( \{\mu_0, \mu_1, \ldots, \mu_n\} \). The value of \( \mu_n \) affects only the scaling of the \( k \)-th orthonormal polynomial \( h^*_k(x) \).

**Theorem 3** (Smith (1990)). Given a non-degenerate sequence of moments, \( \mu_0, \mu_1, \ldots, \mu_n \), the orthogonal polynomial \( h_k(x) \) with \( k = \frac{n}{2} \) defines a unique \( k \)-point discrete probability distribution whose first \( n-1 \) \( (i.e., 2k - 1) \) moments match \( \mu_0, \mu_1, \ldots, \mu_{n-1} \). The support of this distribution are the roots of \( h_k(x) \) and the masses are given by the following equation:

\[
p_i = \left< \prod_{j=1, j \neq i}^k \frac{x - x_j}{x_i - x_j} \right> = \frac{1}{h_k'(x)} \left< \frac{h_k(x)}{(x - x_i)} \right>.
\]

(2.24)

In many situations, one is interested in constructing a discrete distribution that includes some particular points of support. In problems (2.7) and (2.8), the objective function \( \phi(x) \) is set as the indicator function \( I_{(-\infty, t]}(x) \) for a fixed \( t \in \mathcal{I} \). So \( t \) is one support point, \( i.e., t \) is a root of the polynomial \( h_k(x) \), which is pre-given. That means we want to construct a discrete distribution that includes \( t \) in advance as a point of support. For instance, in the example discussed in Section 2.2.2 (page 14), we find the bounds on distribution given four moments. Figure (2.1) shows that \( t \) is a support of polynomial \( h(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \). So the problem becomes: given \( n - 2 \) moments, \( \mu_0, \mu_1, \ldots, \mu_{n-2} \), construct a \( k \)-th degree \( (k = \frac{n}{2}) \) orthogonal polynomial \( h_k(t) \) as a function of the pre-specified support point \( t \) to match the first \( n - 2 \) moments as well as the “appropriate” made moment \( \mu_{n-1} \). Here, \( \mu_{n-1} \) is chosen such that \( t \) is a root of \( h_k(x; \mu_{n-1}) \), where \( h_k(x; \mu_{n-1}) \) denotes the \( k \)-th orthogonal polynomial given by taking the \((n - 1)\)-th moment to be \( \mu_{n-1} \).

Let the roots of \( h_k(x) \) be \( X = \{x_1, \ldots, x_k\} \). Since \( t \in X \), an “appropriate” made moment \( \mu_{n-1} \) can be obtained by solving the condition \( h_k(t) = 0 \). Another method to construct the appropriate \( \mu_{n-1} \) is to set

\[
\mu_{n-1} = -\frac{h_k(t; \overline{\mu}_{n-1})}{\rho_{k-1}^2 h^*_k(x)}
\]
where \( h_k(x; \overline{\mu}_{n-1}) \) denotes the \( k \)-th orthogonal polynomial defined by taking the \((n-1)\)-th moment to be some other value (perhaps zero) and \( \rho_{k-1} \) is the leading coefficient of the \((k-1)\)-th orthonormal polynomial \( h^*_k(x) \). One can prove that the \((n-1)\)-th moment \( \mu_{n-1}(t) \), a function of \( t \), continuously varies over all of its possible values as \( t \) varies over any interval \((x_i, x_{i+1})\), for all \( x \in X \).

Once \( \mu_{n-1}(t) \) is determined, the polynomial \( h_k(x) \) and its roots \( x \)'s are known, so is the discrete probability distribution. According to Theorem 3, these roots are support of the distribution and the corresponding probabilities can be calculated from formula (2.24). Alternatively, the probabilities can be obtained by solving the first \( k-1 \) moment conditions, i.e., \( \sum_{i=1}^{k} p_i x_i^j = \mu_j \) for all \( j = 0, 1, 2, \ldots, k-1 \), although there are \( 2k-2 \) (i.e., \( n-2 \)) moment constraints specified in the problem. With this alternative method, the probabilities are explicitly calculated as:

\[
\begin{bmatrix}
  p_1 \\
  p_2 \\
  p_3 \\
  \vdots \\
  p_k
\end{bmatrix}
= 
\begin{bmatrix}
  1 & 1 & \ldots & 1 \\
  x_1 & x_2 & \ldots & x_k \\
  x_1^2 & x_2^2 & \ldots & x_k^2 \\
  \vdots & \vdots & \ddots & \vdots \\
  x_1^{k-1} & x_2^{k-1} & \ldots & x_k^{k-1}
\end{bmatrix}^{-1}
\begin{bmatrix}
  1 \\
  \mu_1 \\
  \mu_2 \\
  \vdots \\
  \mu_{k-1}
\end{bmatrix}
\]

(2.25)

For any piecewise polynomial \( \phi(x) \) of degree \( n - 2 \) or less, the pseudo-expectation, which is the “accurate” expectation based on only moment information, is calculated as follows:

\[
\langle \phi(x) \rangle = \sum_{i=1}^{k} p_i \phi(x_i).
\]

(2.26)

As analyzed in Section 2.2.2, if the orthogonal polynomial \( h_k(x) \) is constructed subject to the constraint that \( h(x) \geq \phi(x) \), all points of support of \( h_k(x) \) are within the contact set \( Z \) for which \( Z = \{ x \in \mathcal{I} : h(x) = \phi(x) \} \). Therefore, we have

\[
\mathbb{E}[\phi(x)] \leq \sum_{i=1}^{k} p_i \phi(x_i) = \overline{\mu}
\]

Matching 2 or 4 moments

Now, let us illustrate how to calculate bounds on \( \Pr(X \leq t) \) via Smith’s approach, given the estimation of the first two or four raw moments.

1. Given the first two moments \( \mu_0, \mu_1, \mu_2, n-2 = 2 \), so \( n = 4 \). One should construct \( h_2(x) = a_0 + a_1 x + a_2 x^2 \), a quadratic orthogonal polynomial of degree 2 \((k = \frac{n}{2} = 2)\) to determine
the discrete distribution. Notice that in the moment matrix $M_2$

$$M_2 = \begin{bmatrix}
1 & \mu_1 & \mu_2 \\
\mu_1 & \mu_2 & \mu_3 \\
\mu_2 & \mu_3 & \mu_4
\end{bmatrix},$$

(2.27)

$\mu_3$ and $\mu_4$ are unknown. But we only need to choose an “appropriate” $\mu_3(t)$ to guarantee $t$ is a root of $h_2(x)$; that is:

$$h_2(t) = \text{Det} \begin{vmatrix}
\mu_0 & \mu_1 & \mu_2 \\
\mu_1 & \mu_2 & \mu_3(t) \\
1 & t & t^2
\end{vmatrix} = 0.$$

(2.28)

When constructing the discrete distribution, the following two cases should be considered.

(i) If $t \leq \mu_1$ (Figure 2.3), the upper bound is obtained from

$$\begin{bmatrix} 1 & 1 \\ t & x_1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix}$$

and the lower bound is always 0. When lower bound is considered, the quadratic polynomial has only one support point for which $h(x) = \phi(x)$. In this case, one can easily prove that this contact point is $x = \mu_1$.

(ii) If $t \geq \mu_1$ (Figure 2.4), the upper bound is always 1 and the lower bound is obtained by solving two similar equations as in (2.29):

$$\begin{bmatrix} 1 & 1 \\ x_1 & t \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix}$$

(2.30)

2. If the first four moments $\mu_0, \mu_1, \ldots, \mu_4$ are given, $n - 2 = 4$, so $n = 6$. One should construct an orthogonal polynomial of degree $k = \frac{n}{2} = 3$, $h_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$. First, we get the range of $t$ according to the roots of quadratic polynomial $h_2(x)$, i.e., $t \in (x_1, x_2)$.

With the same construction process, one can determine a unique discrete distribution for which three points of support (in which $t$ is included) and the corresponding probabilities are known. Varying $t$ in its range $(x_1, x_2)$, the upper bound $p(t)$ and lower bound $p(t)$, functions of $t$, define a 100% confidence interval in which includes all feasible distributions with the given moments.

Figure 2.5 shows an example of the bounds on a distribution given 4 moments. It illustrates how to build the upper and lower bounds from the stair functions constructed by the roots $x_i$.
and the corresponding probabilities $p_i$ ($i = 1, 2, 3$) of the polynomial $h_3(x)$. Here, $t = x_2$ and $t \in (x_1, x_2)$. The bounds form an envelop containing all distributions with the given moments. The upper bound is constructed from the points $(x_1(t), p_1)$ and $(t, p_1 + p_2)$ and lower bound is constructed by the points $(t, p_1)$ and $(x_2(t), p_1 + p_2)$.

Compared with the SOS approach, the Smith’s approach is faster and more transparent. In Section 2.6, we will discuss numerical examples which show that both methods obtain exactly the same solutions to the moment problems, confirming each other.

2.4 Moment Bounds for Unimodal Distributions

In this section, in addition to moment constraints, we assume the underline distribution is unimodal with pre-specified mode $m$. This additional assumption helps us to narrow the optimal bounds.

A continuous-type random variable $X$ is unimodal with mode $m$ if it satisfies one of the fol-
Figure 2.5. Obtain bounds on cumulative distribution given four moments from stair functions. The upper and lower curves are the upper and lower bounds, respectively, developed from an envelope outline of stair functions.

Following two equivalent conditions:

(i) \((m - x)f'(x) \geq 0\) for all \(x\) in its support \(I\), where \(f(x)\) is the pdf of \(X\).

(ii) Khintcine’s Representation: There are independent random variables \(U\) and \(Y\) such that \(X = m + UY\), where \(U\) is uniformly distributed on \((0, 1)\).

The condition (i) implies the usual definition: The pdf \(f(x)\) has a global maximum at \(m\). This usual definition is equivalent to the second condition (ii), which applies without regard to continuity.

Now, let’s consider the classical moment problems with unimodal constraints. The idea here is to transfer the unimodal bounds problem to its equivalent arbitrary bounds problem and solve the transferred problem using the methods discussed in Section 2.3. The objective function \(\phi(x)\) for the bounds on \(\Pr(X \leq t)\) is transferred to \(\phi^*(y)\) for the bounds on \(\Pr(Y \leq t^*)\), where \(t^* = t - m\),
as follows:

\[
\phi^*(y) = E[\phi(X)|Y = y] \\
= E[\phi(m + UY)|Y = y] \\
= \int_0^1 \phi(m + uy) \, du \\
= \begin{cases} 
\frac{1}{y} \int_{m}^{m+y} \phi(s) \, ds & y \neq 0 \\
\phi(m) & y = 0
\end{cases}
\]

\[
\phi(x) = \begin{cases} 
1 & x \leq t - m \\
0 & x > t - m
\end{cases}
\]

where \( \phi(x) = \begin{cases} 
1 & x \leq t - m \\
0 & x > t - m
\end{cases} \)

Considering the relationship between \( t \) and \( m \), the function \( \phi^*(y) \) has the following two possible expressions:

1. In this case of \( t \geq m \),

\[
\phi^*(y) = \begin{cases} 
1 & y \leq t - m \\
\frac{t - m}{y} & y \geq t - m.
\end{cases}
\]  \[(2.31)\]

2. In this case of \( t < m \),

\[
\phi^*(y) = \begin{cases} 
1 - \frac{t - m}{y} & y \leq t - m \\
0 & y \geq t - m.
\end{cases}
\]  \[(2.32)\]

With the representation \( X = m + UY \), we calculate the moments \( \mu^*_i \) of \( Y \) from the moments \( \mu_i \) of \( X \), using the independence of \( U \) and \( Y \) and the moments of \( X \) and \( U \).

\[
E[(UY)^i] = E[(X - m)^i] \\
E[U^i]E[Y^i] = E\left[ \sum_{j=0}^{i} \binom{i}{j} X^j (-m)^{i-j} \right] \\
\frac{1}{1+r} E[Y^i] = \sum_{j=0}^{i} \binom{i}{j} E[X^j] (-m)^{i-j} \\
\mu^*_i = E[Y^i] = (i + 1) \sum_{j=0}^{i} \binom{i}{j} \mu_j (-m)^{i-j}
\]

Therefore, a set of moment constraints for a unimodal random variable \( X \), \( E(X^i) = \mu_i \) for \( i = \)
1, 2, . . . , n is equivalent to a set of moment constraints for the corresponding random variable Y, 
\( E(Y^i) = \mu_i^* \) for \( i = 1, 2, \ldots, n \).

With the objective function and constraints converted, we transfer the upper bound problem for unimodal variable \( X \) to an equivalent problem for another variable \( Y \). The problem is

\[
\begin{align*}
\bar{p}^* &= \max \int_{\mathcal{I}^*} \phi^*(y) \, dF^*(y) \\
\text{subject to } \int_{\mathcal{I}^*} y^i \, dF^*(y) &= \mu_i^*, \quad \text{for all } i = 1, 2, \ldots, n,
\end{align*}
\]

where \( \mathcal{I}^* \) is the support of \( Y \).

Similarly, the lower bound problem for variable \( Y \) can be obtained by setting the objective function as

\[
\underline{p}^* = \min \int_{\mathcal{I}^*} \phi^*(y) \, dF^*(y),
\]

with the same constraints as (2.33).

### 2.4.1 SOS Approach

The dual of problem (2.33) with unimodal assumption is

\[
\begin{align*}
\bar{d}^* &= \min \sum_{i=0}^{n} a_i^* \mu_i^* \\
\text{subject to } p^*(y) &\geq \phi^*(y), \quad \text{for all } y \in \mathcal{I}^*,
\end{align*}
\]

and, correspondingly, the dual of the lower bound problem is

\[
\begin{align*}
\underline{d}^* &= \max \sum_{i=0}^{n} a_i^* \mu_i^* \\
\text{subject to } p^*(y) &\leq \phi^*(y), \quad \text{for all } y \in \mathcal{I}^*,
\end{align*}
\]

where \( p^*(y) = \sum_{i=0}^{n} a_i^* y^i \).

To write the problem (2.34) as a sum of squares problem suitable to the SOS programming solvers such as SOSTOOLS, we need to write the constraints in an equivalent way, but in terms of polynomials. The inequality constraint of (2.34) is equivalent to the following two simultaneous polynomial inequalities:

1. \( t \geq m \):
   
   \[ \phi^*(y) = 1 \text{ for } y \leq t - m, \text{ so } p^*(y) \geq \phi^*(y) \text{ on } (-\infty, t - m] \text{ is equal to} \]
   
   \[ p^*(y) - 1 \geq 0, \quad y \in (-\infty, t - m] \]

   \[ \phi^*(y) = \frac{t - m}{y} \text{ for } y \geq t - m, \text{ so } p^*(y) \geq \phi^*(y) \text{ on } [t - m, \infty) \text{ is equal to} \]

   \[ p^*(y) - \frac{t - m}{y} \geq 0, \quad y \in [t - m, \infty) \]
CHAPTER 2. UNIVARIATE MOMENT PROBLEM

\[ yp^*(y) - (t - m) \geq 0, \quad y \in [t - m, \infty) \]

2. \( t < m: \)
\[ \phi^*(y) = 1 - \frac{t - m}{y} \text{ for } y \leq t - m, \text{ so } p^*(y) \geq \phi^*(y) \text{ on } (-\infty, t - m] \text{ is equal to} \]
\[ y[1 - p^*(y)] - (t - m) \geq 0, \quad y \in (-\infty, t - m] \]
\[ \phi^*(y) = 0 \text{ for } y \geq t - m, \text{ so } p^*(y) \geq \phi^*(y) \text{ on } [t - m, \infty) \text{ is equal to} \]
\[ p^*(y) \geq 0, \quad y \in [t - m, \infty) \]

By our earlier discussion with Diananda’s Theorem, the dual (2.34) is equivalent to the following SOS program:

\[
\overline{d}^* = \min \sum_{i=0}^{n} a_i^* \mu_i^* \]
subject to

1. \( t \geq m: \)
\[ p^*(t - m - y^2) - 1 \quad \text{is a SOS polynomial} \]
\[ (t - m + y^2)p^*(t - m + y^2) - (t - m) \quad \text{is a SOS polynomial} \]

2. \( t < m: \)
\[ (t - m - y^2)[1 - p^*(t - m - y^2)] - (t - m) \quad \text{is a SOS polynomial} \]
\[ p^*(t - m + y^2) \quad \text{is a SOS polynomial}. \]

The SOS program (2.36) can be solved with a SOS programming solver. As for the dual problem of lower bound, problem (2.35), we can write its inequality constraint as the following two simultaneous polynomial inequalities:

1. \( t \geq m: \)
\[ 1 - p(y) \geq 0 \quad y \in (-\infty, t - m] \]
\[ (t - m) - yp(y) \geq 0 \quad y \in [t - m, \infty) \]

2. \( t \leq m: \)
\[ (t - m) - y[1 - p(y)] \geq 0 \quad y \in (-\infty, t - m] \]
\[ -p(y) \geq 0 \quad y \in [t - m, \infty) \]

The similar process applied to the dual of the lower bound problem. Alternatively, we can obtain lower bound by finding the upper bound on the expectation of \( \psi^*(Y) = 1 - \phi^*(Y) = \mathbb{I}_{(t-m,\infty)}(Y) \). The lower bound on \( \mathbb{E}[\phi^*(Y)] \) equals 1 minus the upper bound of \( \mathbb{E}[\psi^*(Y)] \).

In Section 2.6, the numerical analysis shows that the bounds are tightened when the unimodal
condition is added. Although the addition of the unimodal condition improves the bounds dramatically, the improvement from adding more moment constraints wears out after four or five moments, as Popescu (2005) observes.

2.4.2 Smith’s approach

In this part, as an illustration, we show how to use Smith’s method to compute bounds on a unimodal distribution given the first four moments. Smith recommends employing a modified Newton algorithm to obtain the optimal solution. Newton’s method works when the function of interest is twice continuously differentiable at each support point. Notice that \( \phi^*(y) \) in (2.31) and (2.32) does not satisfy this assumption. Since \( \phi^*(y) \) is not twice continuously differentiable at the point \( t^* = t - m \), we will drop \( t^* \) as a choice variable. This trick does cost us anything since \( t^* \) is given beforehand.

In addition, notice that \( \phi^*(y) \) is not a polynomial. One basic assumption to construct discrete approximations for continuous probability distribution is that \( \phi^*(y) \) must be a polynomial, otherwise (2.26) does not hold.

To streamline our discussion, we define some more compact notations as follows:

\[
\begin{align*}
z(\mu) &\equiv (p_1, p_2, p_3, y_1, y_3), \quad t^* \text{ is assumed to be } y_2 \\
f(z(\mu)) &\equiv p_1 \phi^*(y_1) + p_2 \phi^*(t^*) + p_3 \phi^*(y_3) \\
g_i(z(\mu)) &\equiv p_1 y_1^i + p_2 t^* y_i + p_3 y_3^i = \mu_i \\
g(z(\mu)) &\equiv (g_0(z(\mu)), \ldots, g_5(z(\mu))) = (\mu_0, \ldots, \mu_5)
\end{align*}
\]

Note that \( \mu_5 \) is not pre-specified. As a function of the support point \( t^* = t - m, \mu_5(t^*) \) is properly constructed to guarantee \( t^* \) is a root of \( h_3(y) \), the orthogonal polynomial of degree 3 based on the given moments \( \mu_1^*, \mu_2^*, \mu_3^* \) and \( \mu_4^* \).

Under this notation, our goal is to develop a procedure to compute

\[
\max_{\mu_5} f(z(\mu)) \tag{2.37}
\]

subject to

\[
g_i(z(\mu)) = \mu_i, \quad \text{for all } i = 1, 2, \ldots, 4.
\]

We begin by computing the gradient and Hessian of \( f(z(\mu)) \) with respect to the moments \( \mu = (\mu_0, \ldots, \mu_5) \).

\[
\begin{align*}
\nabla_{\mu} f(z(\mu)) &= \nabla_z f(z(\mu)) \nabla_{\mu} z(\mu) \\
\nabla_{\mu} g(z(\mu)) &= \nabla_z g(z(\mu)) \nabla_{\mu} z(\mu) \tag{2.38}
\end{align*}
\]
Consider that

\[ g_0(z(\mu)) \equiv p_1 + p_2 + p_3 = \mu_0 \]
\[ g_1(z(\mu)) \equiv p_1y_1 + p_2t^* + p_3y_3 = \mu_1 \]
\[ g_2(z(\mu)) \equiv p_1y_1^2 + p_2t^{*2} + p_3y_3^2 = \mu_2 \]
\[ g_3(z(\mu)) \equiv p_1y_1^3 + p_2t^{*3} + p_3y_3^3 = \mu_3 \]
\[ g_4(z(\mu)) \equiv p_1y_1^4 + p_2t^{*4} + p_3y_3^4 = \mu_4 \]
\[ g_5(z(\mu)) \equiv p_1y_1^5 + p_2t^{*5} + p_3y_3^5 = \mu_5 \]

(2.38) can be written in detail as follows:

\[
\left( \frac{\partial f}{\partial \mu_0}, \ldots, \frac{\partial f}{\partial \mu_5} \right)_{1 \times 6} = \left( \frac{\partial f}{\partial p_1}, \ldots, \frac{\partial f}{\partial y_3} \right)_{1 \times 5} \begin{bmatrix} \frac{\partial p_1}{\partial \mu_0} & \frac{\partial p_1}{\partial \mu_2} & \cdots & \frac{\partial p_1}{\partial \mu_5} \\ \frac{\partial p_2}{\partial \mu_0} & \frac{\partial p_2}{\partial \mu_2} & \cdots & \frac{\partial p_2}{\partial \mu_5} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_5}{\partial \mu_0} & \frac{\partial p_5}{\partial \mu_2} & \cdots & \frac{\partial p_5}{\partial \mu_5} \end{bmatrix}_{5 \times 6}
\]

\[
\begin{bmatrix} \frac{\partial y_1}{\partial \mu_0} & \frac{\partial y_1}{\partial \mu_2} & \cdots & \frac{\partial y_1}{\partial \mu_5} \\ \frac{\partial y_2}{\partial \mu_0} & \frac{\partial y_2}{\partial \mu_2} & \cdots & \frac{\partial y_2}{\partial \mu_5} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_5}{\partial \mu_0} & \frac{\partial y_5}{\partial \mu_2} & \cdots & \frac{\partial y_5}{\partial \mu_5} \end{bmatrix}_{6 \times 6} = \begin{bmatrix} \frac{\partial y_1}{\partial \mu_0} & \frac{\partial y_1}{\partial \mu_2} & \cdots & \frac{\partial y_1}{\partial \mu_5} \\ \frac{\partial y_2}{\partial \mu_0} & \frac{\partial y_2}{\partial \mu_2} & \cdots & \frac{\partial y_2}{\partial \mu_5} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_5}{\partial \mu_0} & \frac{\partial y_5}{\partial \mu_2} & \cdots & \frac{\partial y_5}{\partial \mu_5} \end{bmatrix}_{6 \times 5} \begin{bmatrix} \frac{\partial p_1}{\partial \mu_0} & \frac{\partial p_1}{\partial \mu_2} & \cdots & \frac{\partial p_1}{\partial \mu_5} \\ \frac{\partial p_2}{\partial \mu_0} & \frac{\partial p_2}{\partial \mu_2} & \cdots & \frac{\partial p_2}{\partial \mu_5} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_5}{\partial \mu_0} & \frac{\partial p_5}{\partial \mu_2} & \cdots & \frac{\partial p_5}{\partial \mu_5} \end{bmatrix}_{5 \times 6}
\]

Since \( g(z(\mu)) = \mu \), we see that \( \nabla_\mu g(z(\mu)) \) is an identity matrix and \( \nabla_\mu z(\mu) = [\nabla_z g(z(\mu))]^{-1} \). So we have \( \nabla_\mu f(z(\mu)) = \nabla_z f(z(\mu))[\nabla_z g(z(\mu))]^{-1} \), or equivalently

\[
\nabla_z g(z(\mu))^T \nabla_\mu f(z(\mu))^T = \nabla_z f(z(\mu))^T
\]

(2.39)

The \( i \)-th component of the gradient, \( \nabla_\mu f(z(\mu)) \), is the partial derivative of \( f(z(\mu)) \) with respect to \( \mu_i \). Denote the gradient of \( f(z(\mu)) \) with respect of \( \mu \) as \((\lambda_0, \ldots, \lambda_5)\), equation (2.39) can be written as follows:

\[
\begin{bmatrix} 1 & y_1 & y_1^2 & y_1^3 & y_1^4 & y_1^5 \\ 1 & t^* & t^{*2} & t^{*3} & t^{*4} & t^{*5} \\ 1 & y_3 & y_3^2 & y_3^3 & y_3^4 & y_3^5 \\ 0 & p_1 & 2p_1y_1 & 3p_1y_1^2 & 4p_1y_1^3 & 5p_1y_1^4 \\ 0 & p_3 & 2p_3y_3 & 3p_3y_3^2 & 4p_3y_3^3 & 5p_3y_3^4 \end{bmatrix} \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} = \begin{bmatrix} \phi^*(y_1) \\ \phi^*(t^*) \\ \phi^*(y_3) \\ p_1\phi^*(y_1) \\ p_3\phi^*(y_3) \end{bmatrix}
\]

(2.40)

In addition to defining the partial derivatives \((\lambda_0, \ldots, \lambda_5)\), the first 3 rows of \( \nabla_z g(z(\mu))^{-T} \) thus indicate the sensitivity of the probability \((p_1, p_2, p_3)\) to changes in the moments \((\mu_0, \mu_1, \ldots, \mu_5)\).
From equation (2.40), we have

$$(p_1, p_2, p_3, 0, 0) \nabla z g(z(\mu))^T = (\mu_0, \mu_1, \ldots, \mu_5)$$ (2.41)

Note that $\nabla z g(z(\mu))^T$ is not a square matrix, so if we calculate $\nabla z g(z(\mu))^{-T}$ directly, we should calculate its pseudo-inverse 6. Below, we will show that $\nabla z g(z(\mu))^{-T}$ can be obtained indirectly without finding its pseudo-inverse.

First, note that the matrix $\nabla z g(z(\mu))$ has an important dual interpretation. The partial derivative $\lambda_i$ may be interpreted as the coefficient of a polynomial

$$p^*(y) = \sum_{i=0}^{5} \lambda_i y^i$$ (2.42)

The dual interpretation of $\nabla z g(z(\mu))$ allows us to use the Lagrange interpolation formula (Davis, 1975, p.35-37) to write an explicit formula for its pseudo-inverse. Taking $h(y)$ to be the orthogonal polynomial

$$h(y) = \prod_{j=1}^{3} (y - y_j)$$

The first 3 rows of $\nabla z g(z(\mu))^{-T}$ are given by the coefficients of the polynomial

$$[1 - \frac{h''(y_j)}{h'(y_j)} (y - y_j)] \left( \frac{h(y)}{h'(y_j)(y - y_j)} \right)^2$$

The 4th and 5th rows are calculated as follows:

$$p_1(y - y_1) \left( \frac{h(y)}{h'(y_1)(y - y_1)} \right)^2$$

$$p_3(y - y_3) \left( \frac{h(y)}{h'(y_3)(y - y_3)} \right)^2$$

Thus once we have $z(\mu)$, the inverse of $\nabla z g(z(\mu))$ can be easily and accurately computed.

Follow Smith’s approach, the Hessian of $f(z(\mu))$ can be written as

$$\nabla^2 f(z(\mu)) = \nabla z g(z(\mu))^{-T} \begin{bmatrix} 0_{3\times 3} & 0_{3\times 2} \\ 0_{2\times 3} & C_{2\times 2} \end{bmatrix} \nabla z g(z(\mu))^{-1}$$ (2.43)

---

6For a matrix $A_{n\times m}$ whose columns are linearly independent, its pseudo-inverse $A_{n\times m}^+$ equals $(A_{n\times n}^* A_{n\times m})^{-1} A_{n\times n}^*$, where $A^*$ is the conjugate transpose of matrix $A$. For matrices whose elements are real numbers instead of complex numbers, $A^* = A^T$. 
Where, \( C \) is calculated as follows:

\[
C \equiv \nabla^2_{yy} f(z(\mu)) - \lambda(\mu)^T \nabla^2_{yy} g(z(\mu)) = \text{diagonal} \left[ p_i(\phi^{*''}(y_j) - p^{*''}(y_j)) \right]
\]

Here, \( p^*(y) \) is obtained from equation (2.42) and \( C \) can be expressed as

\[
\begin{bmatrix}
  p_1[\phi^{*''}_{yy}(y_1) - p^{*''}_{yy}(y_1)] & 0 \\
  0 & p_3[\phi^{*''}_{yy}(y_3) - p^{*''}_{yy}(y_3)]
\end{bmatrix}
\]

After the gradient and Hessian are obtained, we can use the modified Newton’s method to complete the calculation. Since the first-order condition for optimizing (2.37) requires \( \lambda_5 = 0 \) (as we discussed in Section 2.2.2), the algorithm can be described as follows:

1. Calculate \( \lambda = \nabla_{\mu} f(z(\mu(k))^T \) from equation (2.40) and \( H = \nabla^2_{\mu\mu} f(z(\mu(k))) \) from equation (2.43). If \( \lambda_5 = 0 \), then stop, otherwise let

\[
\mu(k + 1) = \mu(k) - (0, 0, 0, 0, \frac{\lambda_5}{H_{6,6}})
\]

2. If \( f(z(\mu(k + 1))) \leq f(z(\mu(k))) \), let

\[
\mu(k + 1) = 0.5[\mu(k + 1) + \mu(k)]
\]

and repeat this step, otherwise set \( k = k + 1 \) and return to step (1).

### 2.5 The Maximum-Entropy Method

There are many methods available that can be used to construct representative distributions to match moments, such as Gram-Charlier method, nonparametric kernel method, Edgeworth approximation, etc. The maximum-entropy method is a general and powerful one. It guarantees to produce a valid probability distribution. Unlike Edgeworth approximation, the maximum-entropy method does not rely on a normal distribution assumption. Smith (1990, page 84) shows that for the discrete distribution, the maximum-entropy distribution is the distribution that can be realized in the greatest number of ways. This characteristic asymptotically holds for continuous distributions.

The maximum-entropy method has its theoretical basis in the work of Shannon in information
theory and Jaynes in statistical physics. More than fifty years ago, they proposed the measure\(^7\)

\[
H(f(x)) = -\int_a^b f(x) \log f(x) \, dx
\]

as a basis for assigning entropy to a distribution.

Consider the following maximization problem:

\[
\max_{f(x)} \quad -\int_a^b f(x) \log f(x) \, dx
\]

subject to

\[
\int_a^b x^i f(x) \, dx = \mu_i \quad \text{for all } i = 0, 1, \ldots, n
\]

(2.44)

where \(\mu_0, \mu_1, \ldots, \mu_n\) is the given sequence of moments. The solution to the above problem, \(f^*(x)\), is called the maximum-entropy distribution function. The support is in \(I = [a, b]\), which could be \(\mathbb{R}\). The maximal-entropy distribution is sensitive to the support interval, which has to be specified in advance. Trials and errors lead to an appropriate support in the numerical works.

By maximizing this measure of uncertainty subject to moment constraints, the distribution \(f^*(x)\) is said to be the “maximally non-committal” or the “least informative” distribution that is consistent with the given moments.

Denote the Lagrange multipliers associated with the moment constraints by \(\lambda_0, \lambda_1, \ldots, \lambda_n\). The Lagrangian is written as

\[
L(f(x), \lambda_0, \ldots, \lambda_n) = -\int_a^b f(x) \log f(x) \, dx + \sum_{i=0}^n \lambda_i \left( \mu_i - \int_a^b x^i f(x) \, dx \right)
\]

(2.45)

Differentiating (2.45) with respect to \(f(x)\), we get:

\[
\frac{dL}{df(x)} = -\log f(x) - 1 - \sum_{i=0}^n \lambda_i x^i.
\]

Setting the derivative equal to zero, we get the maximum-entropy distribution:

\[
f^*(x) = \exp \left( -1 - \sum_{i=0}^n \lambda_i x^i \right).
\]

\(^7\)The entropy measure is uniquely defined for discrete distributions only. For continuous distributions, the appropriate generalization is \(\int_a^b f(x) \log(f(x)/m(x)) \, dx\) where \(m(x)\) is “an invariant measure proportional to the limiting density of discrete points” (Jaynes, 1968). Given no moment constraints, the measure \(m(x)\) is the maximum-entropy distribution and thus can be interpreted as a “prior distribution representing a state of complete ignorance” (Jaynes, 1968). We assume that \(m(x) = 1\).
$f^*(x)$ can be computed using dual methods of non-linear programming (Luenberger, 1984). We seek Lagrange multipliers that minimize a dual objective function $\phi(\lambda_0, \ldots, \lambda_n)$:

$$
\min_{\lambda_0, \ldots, \lambda_n} \phi(\lambda_0, \ldots, \lambda_n) = \min_{\lambda_0, \ldots, \lambda_n} \left[ \max_{f(x)} L(f(x), \lambda_0, \ldots, \lambda_n) \right]
$$

$$
= \min_{\lambda_0, \ldots, \lambda_n} L(f^*(x, \lambda_0, \ldots, \lambda_n), \lambda_0, \ldots, \lambda_n)
$$

$$
= \min_{\lambda_0, \ldots, \lambda_n} \left[ \int_{x=a}^{b} f^*(x, \lambda_0, \ldots, \lambda_n) \, dx + \sum_{i=0}^{n} \lambda_i \mu_i \right]
$$

(2.46)

The gradient and Hessian of the dual objective function are easily computed and interpreted. If we denote the moments of $f^*(x, \lambda_0, \ldots, \lambda_n)$ by

$$
\langle x^i \rangle = \int_{x=a}^{b} x^i f^*(x, \lambda_0, \ldots, \lambda_n) \, dx
$$

(2.47)

The gradient and Hessian matrix can be written as:

$$
G = \frac{\partial \phi(\lambda_0, \ldots, \lambda_n)}{\partial \lambda} = \begin{bmatrix} (\mu_0 - \langle x^0 \rangle) & \cdots & (\mu_n - \langle x^n \rangle) \end{bmatrix}
$$

(2.48)

$$
H = \frac{\partial^2 \phi(\lambda_0, \ldots, \lambda_n)}{\partial \lambda^2} = \begin{bmatrix} \langle x^0 \rangle & \langle x^1 \rangle & \cdots & \langle x^n \rangle \\
\langle x^1 \rangle & \langle x^2 \rangle & \cdots & \langle x^{n+1} \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle x^n \rangle & \langle x^{n+1} \rangle & \cdots & \langle x^{2n} \rangle \end{bmatrix}
$$

(2.49)

The dual minimization problem can be solved using the modified Newton method (Luenberger, 1984) which we briefly describe below. We denote the vector of Lagrange multiplier after $k$-th iterations of the algorithm by $\lambda(k) = [\lambda_0(k), \ldots, \lambda_n(k)]$ and assume that we begin with $k = 1$ and $\lambda(0) = 0$.

1. If $G \neq 0$, let $\lambda(k+1) = \lambda(k) - [H]^{-1}G$; otherwise stop. $\lambda(k)$ is the optimal solution.

2. If $\phi(\lambda(k+1)) > \phi(\lambda(k))$, let $\lambda(k+1) = \lambda(k) + 1/2[\lambda(k+1) - \lambda(k)]$ and repeat this step; otherwise let $k = k + 1$ and return to (1).

Figure 2.5 shows how the maximum-entropy distribution matches the first four moments of a gamma distribution with $k = 9$ and $\theta = 0.5$. The support interval is chosen at $[0, 12]$. The first 4 moments are $[\mu_1, \mu_2, \mu_3, \mu_4] = [4.5, 22.5, 123.75, 742.5]$. It can be seen that when 4 moments are considered, the maximum-entropy distribution matches the underline distribution very well.
2.6 Numerical Analysis

In this section, we perform numerical analysis to illustrate bounds calculation given moments with or without unimodal assumption. We first analyze some special examples whose underline distributions are already known. By changing the number of given moments, we test the relationship between the number of moment constraints and the tightness of the bounds. We find bounds get narrower as more moments are given, but the improvement wears out after four or five moments. On the other hand, this does not mean the more moment constraints specified, the tighter the estimated bounds. We find when more than 10 moments are given, the bounds become non-smooth and erratic in some support values. Then, we find bounds on empirical insurance industry data provided by the reinsurance company General Re-New England Asset Management, Inc. (GR-NEAM). At the end of this section, robustness tests are performed to examine the sensitivity of the bound estimations with respect to the change of moments. Instead of using the theoretical moments of the underlying distributions as in Section 2.6.1, we estimate the empirical moments of the random samples. Moreover, we analyze the reliability of the moment method when the underlying distribution does not have finite higher moments.
2.6.1 Special Distribution Analysis

Beta Distribution

In this example, the moments are from a beta distribution with \( a = 2, b = 3 \), and the scale parameter \( \theta = 5 \), using the notation of Loss Model (Klugman et al., 2004), which also provides formulae for moments and the mode. The support is \( I = [0, 5] \). The \( k \)-th moment is calculated as

\[
E(X^k) = \frac{\theta^k \Gamma(a + b) \Gamma(a + k)}{\Gamma(a) \Gamma(a + b + k)} = \frac{5^k \Gamma(5) \Gamma(k + 2)}{\Gamma(2) \Gamma(k + 5)}.
\]

The mode is

\[
\frac{(a - 1)\theta}{a + b - 2} = \frac{5}{3}.
\]

The cumulative density function in terms of the incomplete beta function is \( F(x) = \beta(x/\theta; a, b) \).

In Figure 2.7, we show the upper and lower bounds on \( \Pr(X \leq t) \) subject to two to ten raw moments constraints, comparing the bounds obtained by the moment constraints with unimodal assumption and the true beta cumulative distribution with the same moments and mode. We also show the arbitrary distribution bounds which satisfies the moment constraints without unimodal assumption. The first ten raw moments of the beta distribution with parameters \( a = 2, b = 3 \) and \( \theta = 5 \) are \( \mu = (2, 5, 14.29, 44.64, 148.81, 520.83, 1893.94, 7102.27, 27316.43, 107314.56) \). For each set of moments, the bounds derived with the unimodal assumption are much narrower than the bounds subject only to the moment constraints. In addition, except for the two-moment case, the semiparametric unimodal bounds are very close to the true beta distribution. There is little improvement on bounds beyond four or six moments, especially when the unimodal assumption is added. Each graph is a piecewise linear interpolation based on calculation of 11 points, for each arbitrary and unimodal distributions.

The left and right plots in Figure 2.8 show bounds given 14 and 16 moments, respectively. When we consider 16 moments, the bounds become non-smooth and erratic in some points. The failure might result from the high-dimensional numerical errors of the SOS program for the high-moment problems. Actually, in the real world, it’s hard to estimate such high moments accurately even one has large samples. So to consider bounds problem with more than 10 moments is, to some extent, meaningless.

Normal Distribution

We replicate an example from Popescu (2005) by considering a random variable \( X \) with support \( (-\infty, +\infty) \) and moments the same as a standard normal distribution. The moments are \( \mu = (0, 1, 0, 3, 0, 15, 0, 105, 0, 945) \). We calculate both upper and lower bounds on the value at risk.
Figure 2.7. The moments given are the same as the beta distribution with $a = 2$, $b = 3$ and $\theta = 5$. The support is $[0, 5]$ and the mode is 1.67. The graphs show the cases where the specified numbers of moments are $k = 2, 4, 6, 8,$ and $10$, with $k = 2$ on the upper left and running to the right then down. The last figure shows all of the bounds. In each graph, the highest and lowest lines with $−o−$ are, respectively, the upper and lower bounds on $Pr(X \leq t)$ for any distribution on the same interval with the same set of moments. The solid lines within the arbitrary bounds represent the upper and lower bounds for any distribution under unimodal assumption with the same mode and the same moments as the beta distribution. The middle line with $−∗−$ is the true beta distribution with given parameters.

$Pr(X \leq t)$ over a range of values of $t$, for both arbitrary and unimodal distributions. These are shown in Figure 2.9 for variables with $k$ moments given, for $k = 2, 4, 6, 8,$ and $10$. Again, the unimodal bounds are narrower than the corresponding arbitrary bounds with the same set of moments. The bounds get narrower as $k$ increases, but there is little improvement after $k = 4$. Each bound is drawn based on the piecewise linear interpolation of 11 points.

We do one more experiment to test the performance of bound estimation with more than 10 moment constraints. Figure 2.10 shows that if we specify the first 14 moments (right graph), the estimation fails for $t$ is less than or equal to 3. The bounds on cumulative distribution function should be monotonically increasing throughout the support. This confirms our conclusion for beta distribution that bounds problem given more than 10 moments might obtain unreliable estimation due to high-dimensional numerical errors.
Lognormal Distribution

Consider the gross return variable \( R^G = \frac{S_t}{S_0} \) on \( T = 0 \leq X < +\infty \), where \( S_t = S_0 \exp(\mu + \sigma W) \) and \( W \) is a Wiener process. So \( R^G = \exp(\mu + \sigma W) \) has a lognormal distribution. Figure 2.11 shows the semiparametric arbitrary and unimodal bounds on the lognormal distribution with parameters \( \mu = 0.05 \) and \( \sigma = 0.1 \). Each bound is drawn based on the linear interpolation of 21 points.

For the lognormal distribution, the improvement from adding more moment constraints diminishes even earlier than that of either the beta or normal distribution. There is no dramatic improvement in bounds when more than 2 moments are considered. Therefore, we believe that when the tradeoff between bounds accuracy and the cost of estimating higher moments is concerned, considering only the first four moments of the distribution guarantees to give us a reliable bounds estimation.

2.6.2 Empirical analysis

The empirical analysis is based on the insurance industry data\(^8\) We use annual margin data ranging from 1980 to 2005.

The margin on the insurance business is defined as

\[
M = 1 - CR = 1 - LR - ER,
\]

\(^8\)The data is provided by the reinsurance company General Re-New England Asset Management (GR-NEAM), Inc. Thanks for Dr. Jim Backman’s kindness to offer the data.
Figure 2.9. The moments are $\mu = (0, 1, 0, 3, 0, 15, 0, 105, 0, 945)$ – the same moments as the standard normal distribution. The support is $(-\infty, +\infty) \text{ and the mode is } 0$. The graphs show the cases where the specified numbers of moments are $k = 2, 4, 6, 8, \text{ and } 10$, with $k = 2$ on the upper left and running to the right then down. The last figure shows all of the bounds. In each graph, the highest and lowest lines with $-\circ-$ are, respectively, the upper and lower bounds on $Pr(X \leq t)$ for any distribution on the same interval with the same set of moments. The solid lines within the arbitrary bounds represent the upper and lower bounds for any distribution under unimodal assumption with the same mode and the same moments as the standard normal distribution. The middle line with $-\ast-$ is the true standard normal distribution.

where CR is the combined ratio, LR is the loss ratio with $LR = \frac{\text{Losses Incurred}}{\text{Earned premiums}}$, and ER is the expense ratio with $ER = \frac{\text{Expenses}}{\text{Written premiums}}$. It can be considered as the profit of insurance business per dollar premium earned (or written).

Below is a data summary of the three lines of business, Allied, PPauto, and Comp. There are $n = 26$ observations for each line.

Figure 2.12 draws the histograms of the business lines Allied, PPauto and Comp. For each lines, 6 bins are used. We used the histogram to estimate the mode of each line. The estimated modes are -5.0, -5.1, and -21.4, respectively. From the histograms, we can see that PPauto and Comp do not look unimodal.

First, we analysis arbitrary distribution without the unimodal assumption. For each line of
Figure 2.10. Arbitrary and unimodal bounds on standard normal distribution. The left and right graphs show bounds given 12 and 14 moments, respectively. Each bound is drawn based on the piecewise linear interpolation of 21 points.

Figure 2.11. The moments are the same as the moments of the gross return on an asset with a lognormal price with parameters $\mu = 0.05$ and $\sigma = 0.1$. The moments are $\mu = (1.06, 1.13, 1.22, 1.32)$. The support is $[0, +\infty]$ and the mode is 1.0408. The first graph shows bounds on distribution given the first two moments. The second one shows bounds with four moments considered. They are shown together in the third graph. In each graph, the lines with $-o-$ are bounds on arbitrary distribution. The solid lines within the arbitrary bounds are bounds for any distribution under unimodal assumption with the same mode and the same moments. The middle line with $-*-$ is the true lognormal distribution with parameters $\mu = 0.05$ and $\sigma = 0.1$.

As shown in Figure 2.13, the solid curves obtained by the SOS programming solver are bounds on the distributions given two or four moments. For any random variable $X$ with the same moments, its distribution $\Pr(X \leq t)$ must fall within the interval formed by the bounds. The solutions of Smith’s method are plotted by the lines with $-o-$ if only the first two moments are given and by
Figure 2.12. Histograms (left to right, top to bottom) of Allied, PPauto, and Comp.

Figure 2.13. Bounds on $F(t) = \Pr(X \leq t)$ for three lines of business, Allied, PPauto and Comp (left to right, then down). The solid lines are solutions obtained by the SOS method. The lines with $-o-$ represent the bounds of Smith’s approach, given only the first two moments. Smith’s bounds given 4 moments are denoted by the lines with $-\ast-$. 
Table 2.1. Descriptive Statistics of three lines of business from 1980 to 2005

<table>
<thead>
<tr>
<th></th>
<th>Allied</th>
<th>PPAuto</th>
<th>Comp</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E(X))</td>
<td>0.30344</td>
<td>-7.3231</td>
<td>-9.7802</td>
</tr>
<tr>
<td>(E(X^2))</td>
<td>97.008</td>
<td>360.61</td>
<td>233.59</td>
</tr>
<tr>
<td>(E(X^3))</td>
<td>334.61</td>
<td>-10,011</td>
<td>-8,433.3</td>
</tr>
<tr>
<td>(E(X^4))</td>
<td>23196</td>
<td>496,000</td>
<td>402,000</td>
</tr>
<tr>
<td>Mode</td>
<td>-7.5</td>
<td>-0.5</td>
<td>-12.5</td>
</tr>
<tr>
<td>Maximum</td>
<td>20.146</td>
<td>20.3</td>
<td>4.9</td>
</tr>
<tr>
<td>Minimum</td>
<td>-19.203</td>
<td>-52</td>
<td>-55.717</td>
</tr>
<tr>
<td>Range</td>
<td>39.349</td>
<td>72.3</td>
<td>60.617</td>
</tr>
</tbody>
</table>

Figure 2.14. Comparison of arbitrary and unimodal bounds on \(F(t) = \Pr(X \leq t)\) for the line Allied given 4 moments.

The Smith’s bounds fall exactly on the SOS solutions, confirming each other. Adding more moment constraints tighten bounds on \(\Pr(X \leq t)\), but not uniformly with \(t\).

When the unimodal assumption is added, the bounds are greatly improved. Again, for each line of business, we calculate upper and lower bounds on \(F(t) = \Pr(X \leq t)\) for a range of values of \(t\) running over the support of \(X\) given 2 (for Comp) or 4 moments (for Allied and PPAuto) and the estimated mode \(m\). In Figures 2.14 and 2.15, we show the upper and lower bounds on distributions of Allied and PPAuto respectively, comparing the bounds obtained by the first four raw moments with unimodal assumption and the arbitrary distribution bounds given only the same set of moments. The red lines with --o-- are the bounds of arbitrary distributions and the solid blue lines represent the bounds of the unimodal distribution.
Figure 2.15. Comparison of arbitrary and unimodal bounds on $F(t) = \Pr(X \leq t)$ for the line PPAuto given 4 moments.

Figure 2.16. The upper left plot shows the comparison of arbitrary and unimodal bounds on $F(t) = \Pr(X \leq t)$ for Comp given 2 moments and the estimated empirical mode $m = -5$. The upper right plot shows unimodal bounds for Comp with 2 (in red) or 4 (in blue) moments. Maximum-entropy distributions for Comp given 2 (in red) or 4 (in black) moments are drawn in the bottom plot.
Figure 2.16 compares the arbitrary and unimodal bounds for Comp. Specifying the first four moments makes the unimodal bounds collapse. As a representative distribution with given moments, the maximum-entropy distribution is analyzed to check the unimodality. When 4 moments are considered, the distribution is bimodal. Constraints with only the first 2 moments fail to capture this bimodal characteristic. Therefore, if the data with a given number of moments is not unimodally distributed, adding the unimodal assumption to "force" the distribution to be unimodal makes the bounds estimation fail.

### 2.7 Stability Experiments

In the real world, people may have some difficulty in obtaining big samples. For example, in the insurance market, margin data are generally available only annually. Small samples make the estimation of moments inaccurate. In this section, we first analyze the sensitivity of the bounds estimation with respect to the changes in the moments by altering sample sizes, given fixed number of moments. Then we test the ability of moment method to capture the information of the underlying distribution about the existence of higher moments.

We use Pareto distribution to test the stability of the bounds with respect to the sample size. We choose the Pareto distribution because it has a long tail and the empirical financial data generally exhibit long tails. Loss Model (Klugman et al., 2004) gives a formal definition of long tail. Intuitively it means a distribution assigns relatively high probabilities to regions far from the mean or median.

Here, we test how the estimation of bounds changes with sample size and how accurately it matches the underlying true distribution. We simulate a set of Pareto random variables with sample size $n = 26, 100, 500$ and $1000$. The smallest sample is set at 26 because the empirical margin data we analyzed in Section 2.6.2 has only 26 observations. As the sample size increases, we expect the estimation of bounds will become more and more accurate and closer to the true distribution because the estimates of empirical moments will be closer to the theoretical moments. Since we are more interested in the stability of bound estimation on the tail of the distribution, we show

$$p(t) \leq \Pr(X \leq t) \leq \overline{p}(t), \quad \text{for } t \geq \mu(X) + 2\sqrt{\text{Var}(X)}.$$

For each experiment, the estimation is iterated 10 times under the same conditions.

**Example 1.** We first examine the stability of bounds on a Pareto distribution with $\alpha = 5$ and $\theta = 10$. The underlying distribution has the first four theoretical moments. According to the experiments in Section 2.6.1, 4-moment bounds give relatively reliable estimate of 100% confidence interval of the distribution. Therefore, in the following experiments, we estimate only the first
Figure 2.17. Bounds of $\Pr(X \leq t)$ for $t \geq \mathbb{E}(X) + 2\sqrt{\text{Var}(X)}$ on Pareto distribution with $\alpha = 5, \theta = 10$, given 4 moments, and no other constraint. Each upper and lower bound is calculated 10 times to test stability. The upper left and right graphs are drawn based on samples of 26 and 100 observations, respectively. The lower left and right graphs show bounds on the distribution from samples of 500 and 1000 observations, respectively. In each graph, the uppermost curves represent upper bounds and the lowermost curves denote lower bounds. The middle line with $-\Star -$ is the true Pareto distribution with parameters $\alpha = 5, \theta = 10$.

The upper left and right graphs are drawn based on samples of 26 and 100 observations, respectively. The lower left and right graphs show bounds on the distribution from samples of 500 and 1000 observations, respectively. In each graph, the uppermost curves represent upper bounds and the lowermost curves denote lower bounds. The middle line with $-\Star -$ is the true Pareto distribution with parameters $\alpha = 5, \theta = 10$.

four empirical moments of the random samples and focus on analyzing the stability of 4-moment bounds.

Figure 2.17 shows that larger sample does improve the stability of bounds estimations. When the sample size increases, the bounds get more and more stable. However, we do not observe a significant improvement when the sample size increases far beyond 100. Notice that the curve with $-\Star -$ is the true Pareto distribution with given parameters $\alpha$ and $\theta$. The bounds, as expected, capture the true distribution on the right tail.

For the upper left plot in Figure 2.17, since the sample has only 26 observations, the last data point contributes to the probability higher than 0.9615 (the solid horizontal line in the graph.). We find crossovers of the upper and lower bounds when the probability is between 0.9615 and 1. So a 26 observation sample cannot guarantee to obtain reliable bounds, especially in the tail.

As shown in Figure 2.18, when the unimodal assumption is added, the bounds become narrower and more stable. Notice that the lower bound for each sample is higher than the corresponding one without unimodal assumption. This time, sample with no less than 500 observations gives us relatively stable unimodal bounds.

Example 2. In this example, we choose a Pareto distribution with $\alpha = 1$ and $\theta = 10$ as
the underlying distribution. It has only one finite raw moment, i.e., the mean. We want to know whether the moment method can capture this information and convey it in the bounds estimation.

Figure 2.19 shows that although the underlying distribution does not have finite variance, we still can get 2-moment arbitrary bounds. When four moments are used to estimate the bounds, the bounds calculation fails. Since theoretically, only the mean exists, it does not matter whether we choose small or large samples. Both will give us inaccurate and meaningless estimates of higher moments. Experiments with larger samples (e.g., 1000 or 10000 observations) designate the similar pattern as what we show in Figure 2.19. When the unimodal assumption is added, we get similar results. We still can barely estimate the 2-moment unimodal bounds, but the bounds estimation given 4 sample moments will fail.

Furthermore, we investigate the performance of bounds estimation on a Pareto distribution with $\alpha = 0.5$ and $\theta = 10$. In this case, all calculations fail. This means that the moment method can capture the information of the underlying distribution about whether finite moments exist. Therefore, one can use the moment method to test the existence of higher moments of the empirical data.

In sum, larger samples make the bounds estimation more reliable and accurate. When the trade-
Figure 2.19. Bounds on a Pareto with $\alpha = 1, \theta = 10$, given two or four sample moments, and no other constraint. Each upper and lower bound is calculated 10 times to test stability. The upper left and right graphs are drawn based on samples of 26 and 100 observations, respectively, given the first two sample moments. The lower left and right graphs shows bounds given four moments, from samples of 26 and 100 observations, respectively. The curve with $-\ast-\ast$ in the middle is the true Pareto distribution with parameters $\alpha = 1, \theta = 10$.

If one can figure out that the data are unimodal, adding unimodal assumption greatly improves the estimated bounds. In addition, although the bounds estimation is sensitive to the moment estimates, it can capture the information about the existence of higher moments of the underlying distribution.

2.8 Conclusion

In this chapter, we calculate the semiparametric upper and lower bonds on the probability $\Pr(X \leq t)$ given $X$ has specified moments for a range of value of $t$. We use two different methods, the SOS approach and Smith’s approach, to calculate the bounds. Both methods give us the exactly same solutions, confirming each other. In addition, we computed improved bounds when the unimodal assumption is added. Using the maximum-entropy method, we calculated representative distributions based only on moment information. Finally, we first test the stability of bounds estimates with respect to the sample size. Then we examine the ability of moment method to capture the information about the existence of higher moments.
Chapter 3

Bounds for Extreme Probabilities and Value-at-Risk

In this chapter, we study the moment problems with two correlated random variables. We derive semiparametric upper and lower bounds on value-at-risk (VaR) to estimate the risk of joint extreme events. The bounds depend not only on the means and variance, but also on the covariance of the random variables. We compute these bounds numerically by reformulating the corresponding semiparametric bound problems as sum of squares (SOS) programs. Then the SOS programs are solved via SOS programming solvers. We demonstrate the methodology using three specific applications. The first finds bounds on the probability of the joint event $X_1 \leq t_1$ and $X_2 \leq t_2$ for low values of $t_1$ and $t_2$, given up to second order moment information. As the second application, we analyze bounds on the tail probability of a portfolio consisting of two components, $\Pr(w_1X_1 + w_2X_2 \leq a)$, given second order moments. We then add the information about the expected payoff of an exchange option on portfolio components to obtain tighter bounds. This shows how additional information tightens bounds. In the third problem, we apply the moment approach to the stop-loss payment on $Y = X_1 + X_2$ given moments of $X_1$ and $X_2$. The payoff of a call or put option can be considered as a special case of the stop-loss payment. In the last example, Cox (1991)’s method is also investigated to confirm our SOS program solutions.

3.1 Introduction

In the real world, the phenomena involving two or more correlated factors are everywhere. For example, suppose that in a model, $X_1$ and $X_2$ denote random variables such as a random discount factor and a random future insurance payment. If the insurance payment is subject to economic inflation, then it is correlated with the interest rate which determines the discount factor. As another
example, the variables $X_1$ and $X_2$ can be the returns of two stocks, both of which respond to security market forces. Usually, models of risk-based capital management and enterprise risk management involve several random variables, such as losses, stock prices, interest rates, currency exchange rates and so on, many of which are correlated. A novel aspect of this chapter is how we take this correlation between random variables into account.

Risk managers may be interested in measuring the joint distribution of $X_1$ and $X_2$, especially in the tail when $X_1$ and $X_2$ simultaneously take extreme values. There is an active interest in obtaining information on distributions of joint extreme events. For example, the insurer would like to know the probability of having loss payments exceeding a given threshold and a loss in their asset investment below a certain level at the same time. One way to estimate these probabilities is to derive parameters of an assumed distribution (typically joint normal) and then measure joint extreme events. In many instances, the lack of data makes it impossible to reach sound conclusions with the parametric approach. Even in some cases, where plenty of observations are available (e.g., daily price observations), assuming a particular distribution may be perilous when we lack observations of the extreme events.

The aim of this chapter is to solve for the semiparametric upper and lower bounds on the probability of such extreme events, given the first two sets of moments of the joint distribution. We first show how to numerically compute the upper and lower bounds on joint “extreme” events, when two variables simultaneously have extremely low values. We compute bounds on $Pr(X_1 \leq t_1 \text{ and } X_2 \leq t_2)$ for some appropriate values of $t_1, t_2 \in \mathbb{R}^+$, given the first two moments. For this problem, we consider the random variables to be non-negative, like loss random variables. Second, we consider the probabilities of the “value at risk (VaR)” event, which occurs when the sum of two financial variables takes a very low value. That is, we compute $Pr(w_1X_1 + w_2X_2 \leq a)$ for some appropriate values of $w_1, w_2, a \in \mathbb{R}$, when assuming up to the second order moment information (means, variances, and covariance) and the support of $X_1$ and $X_2$. In the end, bounds on the stop-loss payment are computed given the support and moments.

In all these applications, we use a sum of squares optimization program to solve for the semiparametric bounds. As we noted in Chapter 2, these semiparametric bounds are robust bounds that any reasonable model must satisfy. Throughout this chapter, we focus on showing how the semiparametric bounds considered here can be computed numerically via readily available optimization software, instead of focusing on the mathematics behind the bound problems. Moreover, they provide not only a mechanism for checking the consistency of models but also an initial estimate for cumulative probabilities regardless of any model specifications.

The remainder of the chapter is organized as follows. In Section 3.2, we formally state the semiparametric bound problems considered here. Furthermore, we outline the key well-known results that will be used in Section 3.3, showing how the desired semiparametric bounds can be
numerically computed with readily available optimization solvers. In Section 3.4, we present relevant numerical examples to illustrate the application of our methods. Section 3.5 is for our conclusions.

### 3.2 Preliminaries and Notation

The upper bound problem is to maximize

$$\bar{p} = \max \mathbb{E}_F[\phi(X_1, X_2)]$$

subject to

$$\mathbb{E}_F(X_i) = \mu_i, \quad i = 1, 2, \tag{3.1}$$
$$\mathbb{E}_F(X_i^2) = \mu_i^{(2)}, \quad i = 1, 2,$$
$$\mathbb{E}_F(X_1X_2) = \mu_{12},$$

where $F(x_1, x_2)$ is a probability distribution on $D$, for relevant choice of the given function $\phi(x_1, x_2)$.

The lower bound problem is an analog, except for that the objective function is

$$p = \min \mathbb{E}_F[\phi(X_1, X_2)], \tag{3.2}$$

with the same constraints as (3.1). The given information is the support of $(X_1, X_2)$, $D \subseteq \mathbb{R}^2$ and values of $\mu_i, \mu_i^{(2)}, i = 1, 2$, and $\mu_{12}$, the given first and second order non-central moments of the random variables $X_1, X_2$. Thus, problem (3.1) or problem (3.2) finds the best upper bound or the best lower bound of the expected value $\mathbb{E}_F[\phi(X_1, X_2)]$ over all joint probability distributions $F$ with the given moments and support in $D \subseteq \mathbb{R}^2$.

Notice that from the definition of $\bar{p}$ and $p$ in problems (3.1) and (3.2), the interval $[p, \bar{p}]$ is a sharp “100% confidence interval” on the expected value of $\phi(X_1, X_2)$ for all models of the joint distribution of $(X_1, X_2)$ with the given moments and support. It follows that for any $p' \geq p$ and $p' \leq \bar{p}$, the interval $[p', \bar{p}]$ is also a “100% confidence interval”, although not necessarily sharp. Our aim is to compute numerically “useful” 100% confidence intervals for relevant choices of the function $\phi(x_1, x_2)$, balancing computational effort and tightness of the confidence interval.

In particular, given $t_1, t_2 \in \mathbb{R}^+$ and non-negative random variables $X_1$ and $X_2$, we compute 100% confidence intervals on the probability of the extreme events $X_1 \leq t_1$ and $X_2 \leq t_2$, by setting $\phi(x_1, x_2) = \mathbb{I}_{\{x_1 \leq t_1 \text{ and } x_2 \leq t_2\}}$ and $D = \mathbb{R}^{+2}$. Similarly, given $w_1, w_2, a \in \mathbb{R}$, we compute 100% confidence intervals on the VaR probability $\Pr(w_1X_1 + w_2X_2 \leq a)$, for random variables $X_1$ and $X_2$, by setting $\phi(x_1, x_2) = \mathbb{I}_{\{w_1x_1 + w_2x_2 \leq a\}}$ and $D = \mathbb{R}^2$. In the second case, we strengthen the bounds in problems (3.1) and (3.2) by adding an additional moment constraint, $\mathbb{E}_F((X_1 - X_2)^+) =$
\[ \gamma \] where \( x^+ = \max\{x, 0\} \). That is, we strengthen the bounds of problems (3.1) and (3.2) by only considering distributions of \( X_1, X_2 \) that can replicate the expected payoff \( \gamma \) of an exchange option on \( X_1 \) and \( X_2 \). This illustrates how additional information can be included in the problem. Finally, given \( a, b \in \mathbb{R}^+ \), we compute semiparametric bounds on a stop-loss payment \( \phi(x_1, x_2) \) which is defined as

\[
\phi(x_1, x_2) = \begin{cases} 
  b & \text{if } x_1 + x_2 \geq a + b \\
  x_1 + x_2 - a & \text{if } a \leq x_1 + x_2 \leq a + b \\
  0 & \text{if } x_1 + x_2 \leq a,
\end{cases}
\]

for non-negative random variables \( X_1 \) and \( X_2 \).

The following dual of the upper bound problem (3.1) (see, e.g., Karlin and Studden (1966), Bertsimas and Popescu (2002), and Zuluaga and Peña (2005)):

\[
\overline{d} = \min \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{20}\mu_{1(2)} + y_{02}\mu_{2(2)} + y_{11}\mu_{12}
\]

subject to \( p(x_1, x_2) \geq \phi(x_1, x_2) \), for all \( (x_1, x_2) \in \mathcal{D} \),

(3.3)

and the dual of the lower problem (3.2),

\[
d = \max \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{20}\mu_{1(2)} + y_{02}\mu_{2(2)} + y_{11}\mu_{12}
\]

subject to \( p(x_1, x_2) \leq \phi(x_1, x_2) \), for all \( (x_1, x_2) \in \mathcal{D} \),

(3.4)

will be used throughout this chapter, where the quadratic polynomial

\[
p(x_1, x_2) = y_{00} + y_{10}x_1 + y_{01}x_2 + y_{20}x_1^2 + y_{02}x_2^2 + y_{11}x_1x_2.
\]

It is not difficult to see that weak duality holds between (3.1) and (3.3) (or between (3.2) and (3.4)) (Chvatal, 1983, p.139); that is, \( p \leq \overline{d} \) (or \( p \geq d \)). More importantly, for the specific problems considered here, one can show that strong duality holds between (3.1) and (3.3) (or between (3.2) and (3.4)); that is, \( p = \overline{d} \) (or \( p = d \)), as long as problem (3.1) (or problem (3.2)) has solution (feasible).

Similar as we discussed in Chapter 2 Section 2.2.2 (page 11), if problem (3.1) is feasible and there exist \( y_{00}, y_{01}, y_{10}, y_{20}, y_{02}, y_{11} \) such that

\[
p(x_1, x_2) > \phi(x_1, x_2), \text{ for all } (x_1, x_2) \in \mathcal{D},
\]
then $\bar{p} = \bar{d}$. In the analog, for problem (3.2), we reverse the inequality and replace $\bar{p} = \bar{d}$ with $p = d$. Zuluaga and Peña (2005, Proposition 4.1(ii)) established this also.

For our first two examples with $\phi(x_1, x_2)$ a indicator function bounded on $[0, 1]$, the inequality $p(x_1, x_2) > \phi(x_1, x_2)$ of problem (3.1) holds by setting $y_{00} > 1$ and $y_{ij} = 0$ for all $(i, j) \neq (0, 0)$; and by setting $y_{00} < 0$ and $y_{ij} = 0$ for all $(i, j) \neq (0, 0)$, the inequality of the lower bound problem (3.2), $p(x_1, x_2) < \phi(x_1, x_2)$ holds. When the bounds on stop-loss payment are considered, $\phi(x_1, x_2)$ is bounded on $[0, b]$. So one can set $y_{00} > b$ for the upper bound problem (3.1) or $y_{00} < 0$ for the lower bound problem (3.2), and set $y_{ij} = 0$ for all $(i, j) \neq (0, 0)$ to satisfy the strict inequality requirement of strong duality. Thus, as long as problem (3.1) (or problem (3.2)) is feasible, $\bar{p} = \bar{d}$ (or $p = d$) and one can solve (3.3) and (3.4) to obtain the desired semiparametric bounds.

Before explaining the methodology to solve (3.3) and (3.4), recall Hilbert’s Theorem (Theorem 1) and Diananda’s Theorem (Theorem 2) discussed in Chapter 2 (see page 16). In this chapter, for our special application to the bivariate moment problems, the case (2) of Hilbert’s Theorem will be applied when the support $D = \mathbb{R}^2$. If the support is $D \subseteq \mathbb{R}^2$, we apply Diananda’s theorem to a quadratic polynomial $p(x_1, x_2)$ as follows: To check if

$$p(x_1, x_2) = y_{00} + y_{10}x_1 + y_{01}x_2 + y_{20}x_1^2 + y_{02}x_2^2 + y_{11}x_1x_2$$

is positive for all $x_1, x_2 \geq 0$, one can check whether

$$p(x_1^2, x_2^2) = y_{00} + y_{10}x_1^2 + y_{01}x_2^2 + y_{20}x_1^4 + y_{02}x_2^4 + y_{11}x_1^2x_2^2$$

is a SOS polynomial.

Loosely speaking, in order to solve (3.3) (or (3.4)), we will break the constraint

$$p(x_1, x_2) \geq \phi(x_1, x_2), \text{ for all } (x_1, x_2) \in D$$

into a number of constraints of the form

$$p_i(x_1, x_2) \geq 0, \text{ for all } (x_1, x_2) \in \mathbb{R}^2, i = 1, \ldots, m,$$  \hspace{1cm} (3.5)

where $p_i, i = 1, \ldots, m$ are suitable quadratic polynomials whose coefficients are linear functions of the coefficients of $p(x_1, x_2)$. Notice that from Diananda’s theorems it follows that (3.5) is equivalent to requiring that

$$p_i(x_1^2, x_2^2) \text{ is a SOS polynomial}, i = 1, \ldots, m.$$
As we already discussed in Section 2.2.4 for univariate problems, and will show in detail in Section 3.3 for bivariate problems, this allows us to reformulate problems (3.3) and (3.4) as SOS programs.

### 3.3 SOS Programming Formulations

In this section we formally present the SOS formulations that will be used to compute 100% confidence intervals for bivariate extreme events, the VaR probability of portfolio returns and the tail probability of stop-loss payments.

#### 3.3.1 Extreme probability bounds

Here, we consider the problem of finding upper and lower bounds on the probability $\Pr(X_1 \leq t_1 \text{ and } X_2 \leq t_2)$ of two non-negative random variables $X_1, X_2$, without making any additional assumption on the distribution of the random variables $X_1, X_2$, other than the knowledge of the first and second order moments of their joint distribution (means, variances, and covariance). The upper semiparametric bounds for this problem can be obtained by setting problem (3.1) with $\phi(x_1, x_2) = 1\{x_1 \leq t_1 \text{ and } x_2 \leq t_2\}$ and $D = \mathbb{R}^+$. (cf. Section 3.2):

$$p_{\text{Extreme}} = \max \mathbb{E}_F(1\{X_1 \leq t_1 \text{ and } X_2 \leq t_2\})$$

subject to $\mathbb{E}_F(X_i) = \mu_i$, $\mathbb{E}_F(X_i^2) = \mu_i^{(2)}$, $i = 1, 2$, $\mathbb{E}_F(X_1X_2) = \mu_{12}$, $F(x_1, x_2)$ a probability distribution on $\mathbb{R}^+$.  

(3.6)

Similarly, the lower semiparametric bounds for this problem can be obtained by setting the objective function of problem (3.2) as follows:

$$p_{\text{Extreme}} = \min \mathbb{E}_F(1\{X_1 \leq t_1 \text{ and } X_2 \leq t_2\}),$$

(3.7)

with the same constraints as (3.6).

Before obtaining the SOS programming formulation of these problems, we discuss their feasibility in terms of the moment information.

Problems (3.6) and (3.7) are feasible, which means they have solutions, provided the moment matrix $\Sigma$ is a positive definite matrix (i.e., all eigenvalues are greater than zero) and all elements
of \( \Sigma \) are greater than zero, where \( \Sigma \) is the moment matrix:
\[
\Sigma = \begin{bmatrix}
1 & \mu_1 & \mu_2 \\
\mu_1 & \mu_1^{(2)} & \mu_{12} \\
\mu_2 & \mu_{12} & \mu_2^{(2)}
\end{bmatrix}.
\]

Zuluaga (2004b) has shown that this follows from Diananda’s Theorem (Theorem 2) and convex duality (Rockafellar, 1970). This means that given moment information \( \Sigma \), we can test for the feasibility of a solution before we begin to solve it.

Next we derive SOS programs to numerically approximate \( \overline{p}_{\text{Extreme}} \) and \( \underline{p}_{\text{Extreme}} \) using SOS programming solvers.

**Upper bound**

We begin by stating the dual of the upper bound problem (3.6):

\[
\overline{d}_{\text{Extreme}} = \min \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{20}\mu_1^{(2)} + y_{02}\mu_2^{(2)} + y_{11}\mu_{12}
\]

subject to \( p(x_1, x_2) \geq 1 \), for all \( x_1 \leq t_1 \) and \( x_2 \leq t_2 \), \( p(x_1, x_2) \geq 0 \), for all \( x_1, x_2 \geq 0 \). \( (3.8) \)

To formulate problem (3.8) as a SOS program, we proceed as follows. First notice that the constraint in (3.8) is equivalent to

\[
p(x_1, x_2) \geq 1, \text{ for all } 0 \leq x_1 \leq t_1, 0 \leq x_2 \leq t_2
\]

\[
p(x_1, x_2) \geq 0, \text{ for all } x_1, x_2 \geq 0. \tag{3.9}
\]

While the second constraint of (3.9) can be directly reformulated as a SOS constraint using Theorem 2, the first constraint is difficult to reformulate as a SOS constraint. That is, there is no linear transformation from \( 0 \leq x_1 \leq t_1, 0 \leq x_2 \leq t_2 \) to \( \mathbb{R}^2 + \) (that would allow us to use Theorem 2). Thus, we change the problem to obtain a SOS program that either exactly or approximately solves problem (3.9). Specifically, consider the following problem related to (3.9):

\[
\overline{d}'_{\text{Extreme}} = \min \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{20}\mu_1^{(2)} + y_{02}\mu_2^{(2)} + y_{11}\mu_{12}
\]

subject to \( p(x_1, x_2) \geq 1 \), for all \( x_1 \leq t_1, x_2 \leq t_2 \)

\[
p(x_1, x_2) \geq 0, \text{ for all } x_1, x_2 \geq 0. \tag{3.10}
\]

Notice that the constraints in (3.10) are stricter than those in (3.9) since the first constraint of (3.10) includes more values of \( x_1 \) and \( x_2 \). Thus, \( \overline{d}'_{\text{Extreme}} \) is a (not necessarily sharp) upper bound on \( \overline{d}_{\text{Extreme}} \); that is, \( \overline{d}'_{\text{Extreme}} \geq \overline{d}_{\text{Extreme}} \).
After we apply the substitution $x_1 \rightarrow t_1 - x_1, x_2 \rightarrow t_2 - x_2$ to the first constraint of (3.10), the constraints of (3.10) can be rewritten as
\begin{align*}
p(t_1 - x_1, t_2 - x_2) - 1 & \geq 0, \quad \text{for all } x_1, x_2 \geq 0 \\
p(x_1, x_2) & \geq 0, \quad \text{for all } x_1, x_2 \geq 0.
\end{align*}

(3.11)

To finish, we apply Theorem 2 to the constraints (3.11) and conclude that (3.10) is equivalent to the following SOS program:

\[
d'_{\text{Extreme}} = \min \quad y_{00} + y_{10}t_1 + y_{01}t_2 + y_{20}t_1^2 + y_{02}t_2^2 + y_{11}t_1t_2
\]

subject to
\begin{align*}
p(t_1 - x_1^2, t_2 - x_2^2) - 1 & \text{ is a SOS polynomial} \\
p(x_1, x_2) & \text{ is a SOS polynomial}.
\end{align*}

(3.12)

The SOS program (3.12) can be readily solved with a SOS programming solver. Thus, if problem (3.6) is feasible (page 55), then we can numerically obtain a (not necessarily sharp) semiparametric upper bound on the extreme probability, $\Pr(X_1 \leq t_1, X_2 \leq t_2) \leq d'_{\text{Extreme}}$, by solving problem (3.12) with a SOS solver.

Lower bound

We begin by stating the dual of the lower bound problem (3.7):

\[
d_{\text{Extreme}} = \max \quad y_{00} + y_{10}t_1 + y_{01}t_2 + y_{20}t_1^2 + y_{02}t_2^2 + y_{11}t_1t_2
\]

subject to
\begin{align*}
p(x_1, x_2) & \leq I\{x_1 \leq t_1 \text{ and } x_2 \leq t_2\}, \quad \text{for all } x_1, x_2 \geq 0.
\end{align*}

(3.13)

The constraint in problem (3.13) is equivalent to:
\begin{align*}
p(x_1, x_2) & \leq 1, \quad \text{for all } 0 \leq x_1 \leq t_1, 0 \leq x_2 \leq t_2 \\
p(x_1, x_2) & \leq 0, \quad \text{for all } x_1 \geq t_1, x_2 \geq 0, \\
p(x_1, x_2) & \leq 0, \quad \text{for all } x_1 \geq 0, x_2 \geq t_2.
\end{align*}

\footnote{SOS program uses a new polynomial $q(x_1, x_2) = p(t_1 - x_1, t_2 - x_2) - 1$, where
\[
q(x_1, x_2) = \left(y_{00} + y_{10}t_1 + y_{01}t_2 + y_{20}t_1^2 + y_{02}t_2^2 + y_{11}t_1t_2 - 1\right) \\
+(-y_{10} - 2y_{20}t_1 - y_{11}t_2)x_1 \\
+(-y_{01} - 2y_{02}t_2 - y_{11}t_1)x_2 \\
+y_{20}x_1^2 + y_{02}x_2^2 + y_{11}x_1x_2.
\]}

The first constraint of (3.11) can be replaced by $q(x_1, x_2) \geq 0$, for all $x_1, x_2 \geq 0$. But with current SOS solvers, it is unnecessary to provide the expanded algebraic expression of the polynomials on the left hand side of the inequality constraints.
Proceeding as in Section 3.3.1, we now change the problem to obtain a SOS program that either exactly or approximately solves problem (3.13). Specifically, consider the following problem related to (3.13):

\[
\begin{align*}
\tilde{d}'_{\text{Extreme}} &= \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12} \\
\text{subject to} \quad & p(x_1, x_2) \leq 1, \quad \text{for all } x_1 \leq t_1, x_2 \leq t_2 \\
& p(x_1, x_2) \leq 0, \quad \text{for all } x_1 \geq t_1, x_2 \geq 0, \\
& p(x_1, x_2) \leq 0, \quad \text{for all } x_1 \geq 0, x_2 \geq t_2.
\end{align*}
\]  

(3.14)

Notice that the constraints in (3.14) are stricter than those in (3.13). Thus, \(\tilde{d}'_{\text{Extreme}}\) is a (not necessarily sharp) lower bound on \(d_{\text{Extreme}}\); that is, \(\tilde{d}'_{\text{Extreme}} \leq d_{\text{Extreme}}\).

Applying the substitutions \(x_1 \rightarrow t_1 - x_1, x_2 \rightarrow t_2 - x_2\) to the first constraint of (3.14) and \(x_1 \rightarrow x_1 + t_1, x_2 \rightarrow x_2 + t_2\) to the second and third constraints respectively, it follows that problem (3.14) is equivalent to the following SOS program when Diananda’s Theorem is applied:

\[
\begin{align*}
\tilde{d}'_{\text{Extreme}} &= \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12} \\
& \quad 1 - p(t_1 - x_1^2, t_2 - x_2^2) \quad \text{is a SOS polynomial} \\
& \quad -p(t_1 + x_1^2, t_2^2) \quad \text{is a SOS polynomial} \\
& \quad -p(x_1^2, t_2 + x_2^2) \quad \text{is a SOS polynomial}
\end{align*}
\]  

(3.15)

The SOS program (3.15) can be readily solved with a SOS programming solver. Thus, if problem (3.7) is feasible, then we can numerically obtain a (not necessarily sharp) semiparametric lower bound on \(d_{\text{Extreme}}\), that is, \(\tilde{d}'_{\text{Extreme}} \leq d_{\text{Extreme}}\). Applying the substitutions \(x_1 \rightarrow t_1 - x_1, x_2 \rightarrow t_2 - x_2\) to the first constraint of (3.14) and \(x_1 \rightarrow x_1 + t_1, x_2 \rightarrow x_2 + t_2\) to the second and third constraints respectively, it follows that problem (3.14) is equivalent to the following SOS program when Diananda’s Theorem is applied:

Following the same technique outlined in this section, one can also derive the upper and lower bounds on the joint survival probability \(\Pr(X_1 \geq t_1, X_2 \geq t_2)\) of two non-negative random variables \(X_1, X_2\). The detailed derivation is attached in Appendix A.

### 3.3.2 VaR probability bounds

In this section, we find upper and lower bounds on the probability that a portfolio \(w_1 X_1 + w_2 X_2\) \((w_1, w_2 \in \mathbb{R}^+)\) attains values lower than or equal to \(a \in \mathbb{R}\), given up to the second order moment information (means, variances, and covariance) on the random variables \(X_1, X_2\). Finding the sharp upper and lower semiparametric bounds for this problem can be formulated by setting \(\phi(x_1, x_2) =\)
\[ \mathbb{I}_{\{w_1 x_1 + w_2 x_2 \leq a\}}, \text{ and } D = \mathbb{R}^2 \text{ in problems (3.1) and (3.2) (cf. Section 3.2).} \]

Specifically, the upper bound is
\[
P_{\text{VaR}} = \max \quad \mathbb{E}_F(\mathbb{I}_{\{w_1 x_1 + w_2 x_2 \leq a\}}) \]
subject to \[
\mathbb{E}_F(X_i) = \mu_i, \quad i = 1, 2, \\
\mathbb{E}_F(X_i^2) = \mu_i^{(2)}, \quad i = 1, 2, \\
\mathbb{E}_F(X_1 X_2) = \mu_{12},
\]
with the same constraints as (3.16)\(^2\).

And the lower bound has the objective function:
\[
P_{\text{VaR}} = \min \quad \mathbb{E}_F(\mathbb{I}_{\{w_1 x_1 + w_2 x_2 \leq a\}}), \tag{3.17}
\]
with the same constraints as (3.16)\(^2\).

Notice that unlike that in Section 3.3.1, the support of the random variables \(X_1, X_2\) considered here is unrestricted. However, if the interest is on non-negative random variables, problems (3.16) and (3.17) still give valid bounds for the corresponding problems with non-negative variables. Generally, the unrestricted bounds and the non-negative bounds are very close in problems such as (3.16) and (3.17) (see e.g., Zuluaga and Peña (2005), Boyle and Lin (1997)).

Before obtaining the SOS programming formulation of these problems, let us state the well-known feasibility condition in terms of the moment parameters (Bertsimas and Sethuraman, 2000, Theorem 16.1.2).

The feasibility of problems (3.16) and (3.17) depends on the moment matrix \(\Sigma\). There are solutions if \(\Sigma\) is a positive semidefinite matrix (i.e., all eigenvalues are greater than or equal to zero), where \(\Sigma\) is the moment matrix:
\[
\Sigma = \begin{bmatrix}
1 & \mu_1 & \mu_2 \\
\mu_1 & \mu_1^{(2)} & \mu_{12} \\
\mu_2 & \mu_{12} & \mu_2^{(2)}
\end{bmatrix}.
\]

\(^2\)To solve the problems (3.29) and (3.30), one can assume \(w_1 = w_2 = 1\) without loss of generality. In this case, we find bounds on \(\Pr(X_1 + X_2 \leq a)\). We can easily convert the problem of \(\Pr(w_1 X_1 + w_2 X_2 \leq a)\) to that of \(\Pr(X_1 + X_2 \leq a)\) by adjusting the moments of \(X_1\) and \(X_2\). Let \(X'_1 = w_1 X_1\) and \(X'_2 = w_2 X_2\). Then we have the following relationships:
\[
\begin{align*}
\mathbb{E}(X'_i) &= \mathbb{E}(w_i X_i) = w_i \mu_i, \quad i = 1, 2, \\
\mathbb{E}(X'_i^2) &= \mathbb{E}(w_i^2 X_i^2) = w_i^2 \mu_i^{(2)}, \quad i = 1, 2, \\
\mathbb{E}(X'_1 X'_2) &= \mathbb{E}(w_1 X_1 w_2 X_2) = w_1 w_2 \mu_{12}.
\end{align*} \tag{3.18}
\]
That is, we can rescale a problem in the form \(w_1 X_1 + w_2 X_2 \leq a\) to the form \(X_1 + X_2 \leq a\).
Note that this is analogous to the feasibility requirements of problems (3.6) and (3.7), except we need only positive semidefinite rather than positive definite (page 55). Zuluaga (2004b) established this also.

Next we derive SOS programs to numerically compute $\overline{\text{VaR}}$, and $\underline{\text{VaR}}$ by using SOS programming solvers.

**Upper bound**

We begin by stating the dual problem of (3.16):

$$\overline{d}_{\text{VaR}} = \min \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}$$

subject to $\quad p(x_1, x_2) \geq 1, \text{ for all } x_1, x_2 \in \mathbb{R}$. \hspace{1cm} (3.19)

To formulate problem (3.19) as a SOS program, we proceed as follows. First notice that the constraint in (3.19) is equivalent to

$$p(x_1, x_2) \geq 1, \text{ for all } x_1, x_2 \text{ with } w_1 x_1 + w_2 x_2 \leq a$$

$$p(x_1, x_2) \geq 0, \text{ for all } x_1, x_2 \in \mathbb{R}. \hspace{1cm} (3.20)$$

Notice that we can directly express the second constraint in (3.20) as a SOS constraint by using Hilbert’s Theorem. For the first constraint, however, we need more work. Specifically, consider the transformation of the axes below:

\begin{align*}
x'_1 &= x_1 \cos \alpha + x_2 \sin \alpha - \frac{a}{w_1} \cos \alpha \\
x'_2 &= -x_1 \sin \alpha + x_2 \cos \alpha \\
\text{and} \\
x_1 &= x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha \\
x_2 &= x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha
\end{align*}

(3.21)
Applying the substitution
\[ x_1 \rightarrow (x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha) \]
\[ x_2 \rightarrow (x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha) \]
and to the first constraint of (3.20) becomes
\[ p \left[ (x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \ (x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha) \right] \geq 1, \text{ for all } x'_1 \leq 0, x'_2 \in \mathbb{R}. \]

This is equivalent to
\[ p[(x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \ (x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)] \geq 1, \text{ for all } x'_1 \leq 0, x'_2 \geq 0 \]
\[ p[(x'_1 \cos \alpha + x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \ (x'_1 \sin \alpha - x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)] \geq 1, \text{ for all } x'_1 \leq 0, x'_2 \leq 0. \] (3.22)

Applying the substitutions \( x'_1 \rightarrow -x'_1 \) to the first constraint of (3.22) and \( x'_1 \rightarrow -x'_1, x'_2 \rightarrow -x'_2 \) to the second constraint, it follows that problem (3.20) is equivalent to
\[ \overline{d}_{\text{VaR}} = \min \ y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_1 \]
subject to
\[ p[(-x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \ (-x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)] - 1 \geq 0, \ \forall \ x'_1 \geq 0, x'_2 \geq 0 \]
\[ p[(-x'_1 \cos \alpha + x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \ (-x'_1 \sin \alpha - x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)] - 1 \geq 0, \ \forall \ x'_1 \geq 0, x'_2 \geq 0. \]
\[ p(x_1, x_2) \geq 0, \text{ for all } x_1, x_2 \in \mathbb{R}. \] (3.23)

From Theorem 2 (applied to the first two constraints of (3.23)) and Theorem ?? (applied to the last constraint of (3.23)), it follows that (3.23) is equivalent to the following SOS program:
\[ \overline{d}_{\text{VaR}} = \min \ y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_1 \]
where the following three polynomials are SOS polynomials:
\[ p[(-x'_2 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \ (-x'_2 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)] - 1 \] (3.24)
\[ p[(-x'_1 \cos \alpha + x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \ (-x'_1 \sin \alpha - x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)] - 1 \]
\[ p(x'_1, x'_2) \]

We dropped the primes in the variable labels. The SOS program (3.24) can be solved with a SOS programming solver. Thus, if problem (3.16) is feasible (page 59), then we can obtain the semiparametric upper bound \( \overline{p}_{\text{VaR}} \) on VaR probability, by solving problem (3.24) with a SOS
Lower bound

We begin by stating the dual of the lower bound problem (3.17):

\[
\begin{align*}
\frac{d}{VaR} &= \max \quad y_0 + y_0\mu_1 + y_0\mu_2 + y_{02}\mu_1^{(2)} + y_{02}\mu_2^{(2)} + y_{11}\mu_1 \\
\text{subject to} \quad p(x_1, x_2) &\leq \mathbb{I}_{\{w_1x_1+w_2x_2 \leq a\}}, \text{ for all } x_1, x_2 \in \mathbb{R}.
\end{align*}
\]  

(3.25)

To formulate problem (3.25) as a SOS program, we proceed as follows. First notice that the constraint of (3.25) is equivalent to

\[
p(x_1, x_2) \leq 1, \text{ for all } x_1, x_2 \in \mathbb{R},
\]

(3.26)

and

\[
p(x_1, x_2) \leq 0, \text{ for all } x_1, x_2 \text{ with } x_1 + x_2 \leq a.
\]

Notice that we can directly express the first constraint in (3.26) as a SOS constraint by using Hilbert’s Theorem. For the second constraint, however, we need to employ the coordinate transformation (3.21).

Taking the same substitutions as for problem (3.20) and following the analogous steps, we obtain that the second constraint of (3.26) can be expressed as:

\[
\begin{align*}
p&\left((-x_1'\cos \alpha - x_2'\sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \quad (-x_1'\sin \alpha + x_2'\cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)\right) \leq 0, \text{ for all } x_1' \geq 0, x_2' \geq 0 \\
p&\left((-x_1'\cos \alpha + x_2'\sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \quad (-x_1'\sin \alpha - x_2'\cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)\right) \leq 0, \text{ for all } x_1' \geq 0, x_2' \geq 0.
\end{align*}
\]  

(3.27)

Applying Theorem 2 and Theorem 1, (3.25) is equivalent to the following SOS program:

\[
\frac{d}{VaR} = \max \quad y_0 + y_0\mu_1 + y_0\mu_2 + y_{02}\mu_1^{(2)} + y_{02}\mu_2^{(2)} + y_{11}\mu_1
\]

where the following three polynomials are SOS polynomials:

\[
\begin{align*}
-p&\left((-x_1^2\cos \alpha - x_2^2\sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \quad (-x_1^2\sin \alpha + x_2^2\cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)\right) \quad (3.28) \\
-p&\left((-x_1^2\cos \alpha + x_2^2\sin \alpha + \frac{a}{w_1} \sin^2 \alpha), \quad (-x_1^2\sin \alpha - x_2^2\cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha)\right) \\
1 - p&\left(x_1^2, x_2^2\right)
\end{align*}
\]

The SOS program (3.28) can be solved with a SOS programming solver. Thus, if problem (3.17) is feasible, it follows that we can obtain the VaR probability semiparametric lower bound \(\frac{d}{VaR}\) by solving problem (3.28) with a SOS solver.

When the complement any problem of finding bounds on \(\Pr(w_1X_1 + w_2X_2 \geq a)\) is considered, one can use the relationship \(\Pr(w_1X_1 + w_2X_2 \geq a) = 1 - \Pr(w_1X_1 + w_2X_2 \leq a)\) to solve it.
As long as we know the upper and lower bounds on $\Pr(X_1 + X_2 \leq a)$, the bounds on $\Pr(w_1 X_1 + w_2 X_2 \geq a)$ can be obtained as follows:

$$
\bar{p}_{\text{VaR}} = \max \ E_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \geq a\}}) = 1 - \min \ E_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \leq a\}}) = 1 - \bar{p}_{\text{VaR}}
$$

$$
\bar{p}_{c\text{VaR}} = \min \ E_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \geq a\}}) = 1 - \max \ E_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \leq a\}}) = 1 - \bar{p}_{c\text{VaR}}.
$$

In addition, this problem can be directly solved by deriving its own SOS programs. The details are discussed in Appendix C.

**VaR probability bounds with information of an exchange option**

To obtain tighter bounds (see numerical results in Section 3.4.2), we include the information of the expected payoff $\gamma$ of an exchange option on the assets; that is, we add the moment constraint $E_F((X_1 - X_2)^+) = \gamma$ (where $x^+ = \max\{0, x\}$) to obtain the following upper bound problem:

$$
\bar{p}_{\text{VaR}} = \max \ E_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \leq a\}})
$$

subject to

$$
E_F(X_i) = \mu_i, \quad i = 1, 2, \quad E_F(X_1^2) = \mu_i^{(2)}, \quad i = 1, 2, \\
E_F(X_1 X_2) = \mu_{12},
$$

$$
E_F((X_1 - X_2)^+) = \gamma,
$$

$F(x_1, x_2)$ a probability distribution on $\mathbb{R}^2$.

Similarly, the lower bound problem has the objective function as follows:

$$
\bar{p}_{\text{VaR}} = \min \ E_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \leq a\}}),
$$

with the same constraints as (3.29).

The duality results discussed in Section 3.2 for problems (3.1) and (3.2) extend to these two problems. Before obtaining the SOS programming formulation of problems (3.29) and (3.30), we discuss its feasibility condition in terms of the moment parameters, which readily follows from classical moment theory (see, e.g., Bertsimas and Sethuraman (2000, Theorem 16.1.2)).

First, consider the feasibility of problems (3.29) and (3.30). They have solutions if and only if $\Sigma$ is a positive semidefinite matrix and $\underline{p}_{\text{Exch}} < \gamma < \bar{p}_{\text{Exch}}$, where $\Sigma$ is the moment matrix

$$
\Sigma = \begin{bmatrix}
1 & \mu_1 & \mu_2 \\
\mu_1 & \mu_1^{(2)} & \mu_{12} \\
\mu_2 & \mu_{12} & \mu_2^{(2)}
\end{bmatrix},
$$

where $\mu_i$, $\mu_i^{(2)}$, and $\mu_{12}$ are the first and second moments of $X_i$ and $X_1 X_2$, respectively.
and $p_{\text{Exch}}$, $\bar{p}_{\text{Exch}}$ are the upper and lower bounds of problems (3.1) and (3.2) when $\phi(x_1, x_2) = (x_1 - x_2)^+$ (which can be readily computed using SOS techniques (see Zuluaga and Peña (2005)).

Next we derive SOS programs to compute $p_{\text{VaR}}$, and $\bar{p}_{\text{VaR}}$ numerically using SOS programming solvers.

As in Sections 3.3.1 and 3.3.2, the key is to solve the dual of problems (3.29) and (3.30) using SOS programming solvers. From a straightforward generalization of the discussion in Section 3.2, it follows that the dual of problem (3.29) is

$$d_{\text{VaR}} = \min \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{02}\mu_2^{(2)} + y_{11}\mu_{12} + y_0\gamma$$  
$$\text{subject to} \quad p(x_1, x_2) + y_0(x_1 - x_2)^+ \geq \mathbb{I}_{w_1 x_1 + w_2 x_2 \leq a}, \quad \text{for all } x_1, x_2 \in \mathbb{R}. \quad (3.31)$$

Similarly, the dual of problem (3.30) is:

$$d_{\text{VaR}} = \max \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{02}\mu_2^{(2)} + y_{11}\mu_{12} + y_0\gamma$$  
$$\text{subject to} \quad p(x_1, x_2) + y_0(x_1 - x_2)^+ \leq \mathbb{I}_{w_1 x_1 + w_2 x_2 \leq a}, \quad \text{for all } x_1, x_2 \in \mathbb{R}. \quad (3.32)$$

Also, as a straightforward generalization of Proposition 3.2 (using Zuluaga and Peña (2005, Proposition 4.1(ii))), if problem (3.29) (or problem (3.30)) is feasible and strong duality holds between problems (3.29) and (3.31) (or between problems (3.30) and (3.32)), we have $p_{\text{VaR}} = d_{\text{VaR}}$ (or $\bar{p}_{\text{VaR}} = d_{\text{VaR}}$). Also notice that if we set $y_0 = 0$, the exchange option constraint is dropped and we go back to problems (3.16) and (3.17).

**Upper bound with exchange option**

To formulate the upper bound problem (3.31) as a SOS program, we proceed as follows. First notice that the upper bound version of (3.31) is equivalent to

$$d_{\text{VaR}} = \min \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{02}\sigma_1^2 + y_{11}\sigma_{12} + y_0\gamma$$  
$$\text{subject to} \quad p(x_1, x_2) + y_0(x_1 - x_2)^+ \geq 1, \quad \text{for all } x_1, x_2 \text{ with } w_1 x_1 + w_2 x_2 \leq a, x_1 \geq x_2$$  
$$p(x_1, x_2) \geq 1, \quad \text{for all } x_1, x_2 \text{ with } w_1 x_1 + w_2 x_2 \leq a, x_1 \leq x_2$$  
$$p(x_1, x_2) + y_0(x_1 - x_2)^+ \geq 0, \quad \text{for all } x_1, x_2 \text{ with } x_1 \geq x_2$$  
$$p(x_1, x_2) \geq 0, \quad \text{for all } x_1, x_2 \text{ with } x_1 \leq x_2. \quad (3.33)$$
In order to use Theorem 2, we will use the following transformations:

\[
\begin{align*}
    x_1 &= z_1 + z_2 \\
    x_2 &= z_2 \\
    x_1 &= z_1 \\
    x_2 &= z_1 + z_2
\end{align*}
\tag{3.34}
\]

Applying the upper left transformation in (3.34) to the first and third constraints of problem (3.33) and applying the upper right transformation in (3.34) to the second and fourth constraints of problem (3.33), the constraints in (3.33) are equivalent to

\[
\begin{align*}
p(z_1 + z_2, z_2) + y_0 z_1 &\geq 1, \quad \text{for all } z_1, z_2 \text{ with } w_1(z_1 + z_2) + w_2 z_2 \leq a, z_1 \geq 0 \\
p(z_1, z_1 + z_2) &\geq 1, \quad \text{for all } z_1, z_2 \text{ with } w_1 z_1 + w_2(z_1 + z_2) \leq a, z_2 \geq 0 \\
p(z_1 + z_2, z_2) + y_0 z_1 &\geq 0, \quad \text{for all } z_1, z_2 \text{ with } z_1 \geq 0 \\
p(z_1, z_1 + z_2) &\geq 0, \quad \text{for all } z_1, z_2 \text{ with } z_2 \geq 0.
\end{align*}
\tag{3.35}
\]

Now applying the lower left and right transformations in (3.34) to the first two constraints of (3.35) respectively, these two constraints are equivalent to

\[
\begin{align*}
p(t_1 + \frac{a - w_1 t_1}{w_1 + w_2} - t_2, \frac{a - w_1 t_1}{w_1 + w_2} - t_2) + y_0 t_1 &\geq 1, \quad \text{for all } t_1 \geq 0, t_2 \geq 0 \\
p(\frac{a - w_2 t_2}{w_1 + w_2} - t_1, \frac{a - w_2 t_2}{w_1 + w_2} - t_1 + t_2) &\geq 1, \quad \text{for all } t_1 \geq 0, t_2 \geq 0.
\end{align*}
\tag{3.36}
\]
Finally, the last two constraints in (3.35) are equivalent to
\begin{align*}
p(z_1 + z_2, z_2) + y_0 z_1 & \geq 0, \quad \text{for all } z_1 \geq 0, z_2 \geq 0 \\
p(z_1 - z_2, -z_2) + y_0 z_1 & \geq 0, \quad \text{for all } z_1 \geq 0, z_2 \geq 0 \\
p(z_1, z_1 + z_2) & \geq 0, \quad \text{for all } z_1 \geq 0, z_2 \geq 0 \\
p(-z_1, -z_1 + z_2) & \geq 0, \quad \text{for all } z_1 \geq 0, z_2 \geq 0.
\end{align*}

After applying Diananda’s Theorem, we obtain the SOS formulation for the upper bound of problem (3.31):
\begin{equation}
\overline{d}_{\text{VaR}} = \min \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \sigma_1^2 + y_{02} \sigma_2^2 + y_{11} \sigma_{12} + y_0 \gamma
\end{equation}
for which the following are SOS polynomials:
\begin{align*}
p(t_1^2 + \frac{a-w_1 t_1^2}{w_1+w_2} - t_2^2, \frac{a-w_1 t_1^2}{w_1+w_2} - t_2^2) + y_0 t_1^2 - 1 \\
p(\frac{a-w_2 t_1^2}{w_1+w_2} - t_1^2, \frac{a-w_2 t_1^2}{w_1+w_2} - t_2^2 + t_1^2) - 1 \\
p(z_1^2 + z_2^2, z_2) + y_0 z_1^2 \\
p(z_1^2 - z_2^2, -z_2^2) + y_0 z_1^2 \\
p(z_1^2, z_1^2 + z_2^2) \\
p(-z_1^2, -z_1^2 + z_2^2)
\end{align*}

Thus, if problem (3.29) is feasible, we can obtain the sharp semiparametric upper bound \(\overline{d}_{\text{VaR}}\) by solving problem (3.37) with a SOS solver.

**Lower bound with exchange option**

To solve the lower bound problem, first expand the constraint in problem (3.32) and obtain the following equivalent problem:
\begin{equation}
d_{\text{VaR}} = \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \sigma_1^2 + y_{02} \sigma_2^2 + y_{11} \sigma_{12} + y_0 \gamma
\end{equation}
subject to
\begin{align*}
p(x_1, x_2) + y_0 (x_1 - x_2) & \leq 1, \quad \text{for all } x_1, x_2 \text{ with } x_1 \geq x_2 \\
p(x_1, x_2) & \leq 1, \quad \text{for all } x_1, x_2 \text{ with } x_1 \leq x_2 \\
p(x_1, x_2) + y_0 (x_1 - x_2) & \leq 0, \quad \text{for all } x_1, x_2 \text{ with } w_1 x_1 + w_2 x_2 \leq a, x_1 \geq x_2 \\
p(x_1, x_2) & \leq 0, \quad \text{for all } x_1, x_2 \text{ with } w_1 x_1 + w_2 x_2 \leq a, x_1 \leq x_2.
\end{align*}
(3.38)
Applying the upper left transformation in (3.34) to the first and third constraints of problem (3.38) and applying the upper right transformation in (3.34) to the second and fourth constraints of problem (3.38), the constraints in (3.38) are equivalent to

\[
\begin{align*}
    p(z_1 + z_2, z_2) + y_0 z_1 &\leq 1, \text{ for all } z_1, z_2 \text{ with } z_1 \geq 0 \\
    p(z_1, z_1 + z_2) &\leq 1, \text{ for all } z_1, z_2 \text{ with } z_2 \geq 0 \\
    p(z_1 + z_2, z_2) + y_0 z_1 &\leq 0, \text{ for all } z_1, z_2 \text{ with } w_1 x_1 + w_2 x_2 \leq a, z_1 \geq 0 \\
    p(z_1, z_1 + z_2) &\leq 0, \text{ for all } z_1, z_2 \text{ with } w_1 x_1 + w_2 x_2 \leq a, z_2 \geq 0.
\end{align*}
\]

(3.40)

Applying the left and right transformations in (3.39) to the third and fourth constraints of (3.40) respectively, these two constraints are equivalent to

\[
\begin{align*}
    p(t_1 + \frac{a-w_1 t_1}{w_1+w_2} + t_2, \frac{a-w_1 t_1}{w_1+w_2} + t_2) + y_0 t_1 &\leq 0, \text{ for all } t_1, t_2 \geq 0 \\
    p(\frac{a-w_2 t_2}{w_1+w_2} + t_1, \frac{a-w_2 t_2}{w_1+w_2} + t_1 + t_2) &\leq 0, \text{ for all } t_1, t_2 \geq 0
\end{align*}
\]

(3.41)

Finally, the first two constraints in (3.40) are equivalent to

\[
\begin{align*}
    p(z_1 + z_2, z_2) + y_0 z_1 &\leq 1, \text{ for all } z_1 \geq 0, z_2 \geq 0 \\
    p(z_1 - z_2, -z_2) + y_0 z_1 &\leq 1, \text{ for all } z_1 \geq 0, z_2 \geq 0 \\
    p(z_1, z_1 + z_2) &\leq 1, \text{ for all } z_1 \geq 0, z_2 \geq 0 \\
    p(-z_1, -z_1 + z_2) &\leq 1, \text{ for all } z_1 \geq 0, z_2 \geq 0.
\end{align*}
\]

After applying Diananda’s Theorem, we obtain the SOS formulation for the lower bound of
problem (3.32):

\[ d_{\text{VaR}} = \max y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \sigma_1^2 + y_{02} \sigma_2^2 + y_{11} \sigma_1 \sigma_2 + y_0 \gamma \]  \hspace{1cm} (3.42)

for which the following are SOS polynomials:

\[ -p(t_1^2 + \frac{a-w_1 t_1^2}{w_1+w_2} + t_2^2 + \frac{a-w_2 t_2^2}{w_1+w_2} + t_1^2 + t_2^2) - y_0 t_1^2 \]
\[ -p(\frac{a-w_2 t_1^2}{w_1+w_2} + t_1^2, \frac{a-w_2 t_2^2}{w_1+w_2} + t_1^2 + t_2^2) \]
\[ 1 - p(z_1^2 + z_2^2) - y_0 z_1 \]
\[ 1 - p(z_1^2 - z_2^2, -z_2^2) - y_0 z_1 \]
\[ 1 - p(z_1^2 - z_2^2, z_2^2) \]
\[ 1 - p(-z_1^2, -z_1^2 + z_2^2) \]

If problem (3.30) is feasible, then we can obtain the sharp VaR probability semiparametric lower bound \( d_{\text{VaR}} \) by solving problem (3.42) with a SOS solver. Furthermore, we obtain a 100% confidence interval \([d_{\text{VaR}}, \overline{d}_{\text{VaR}}]\) on the VaR probability \( \Pr(w_1 X_1 + w_2 X_2 \leq a) \) subject to the given moment and exchange option information.

As an extension, given the fact that the probability \( \Pr(w_1 X_1 + w_2 X_2 \leq a) = 1 - \Pr(w_1 X_1 + w_2 X_2 > a) \), we can calculate the bounds on the \( \Pr(w_1 X_1 + w_2 X_2 > a) \) from the upper and lower bounds on \( \Pr(w_1 X_1 + w_2 X_2 \leq a) \).

### 3.3.3 Bounds on Stop-Loss payments

Stop-loss payments we consider here have two loss components \( X_1 \) and \( X_2 \). For example, a homeowner’s policy covers both property losses \( X_1 \) and liability losses \( X_2 \). Similarly, \( X_1 \) could be hospital room and board costs and \( X_2 \) could be surgical expenses in health insurance. We find the upper and lower bounds on the aggregate loss \( Z = X_1 + X_2 \), given the mean, variance and covariance of \( X_1 \) and \( X_2 \). Consider a stop-loss contract which pays nothing below a retained level \( a \), pays \( X_1 + X_2 - a \) when \( X_1 + X_2 \) exceeds \( a \) and has a maximum payment of \( b \), then our function \( \phi(x_1, x_2) \) in problem (3.1) is defined as follows:

\[
\phi(x_1, x_2) = \begin{cases} 
    b & \text{if } x_1 + x_2 \geq a + b \\
    x_1 + x_2 - a & \text{if } a \leq x_1 + x_2 \leq a + b \\
    0 & \text{if } x_1 + x_2 \leq a.
\end{cases} 
\hspace{1cm} (3.43)
\]

Specifically, the value \( \phi(x_1, x_2) \) represents the benefits a direct insurer pays to a reinsurer, given losses of \( X_1 \) and \( X_2 \). Under this contract, when the total losses are less than \( a \), the direct insurer...
retains all losses. When the sum exceeds the threshold \( a \), the reinsurer pays the excess up to a maximum of \( b \). If the total losses exceed \( a + b \), the part higher than \( b \) will be retained or ceded to other reinsurers by the direct insurer. Here, instead of calculating bounds on probabilities, we calculate bounds on payments.

Given the objective function (3.43) and \( D = \mathbb{R}^+ \), the upper semiparametric bounds problem is formulated as follows:

\[
\overline{p}_{\text{StopLoss}} = \max \quad \mathbb{E}_F(\phi(X_1 + X_2))
\]

subject to \( \mathbb{E}_F(X_i) = \mu_i, \quad i = 1, 2, \)

\( \mathbb{E}_F(X_i^2) = \mu_i^{(2)}, \quad i = 1, 2, \)

\( \mathbb{E}_F(X_1 X_2) = \mu_{12}, \)

\( F(x_1, x_2) \) a probability distribution in \( \mathbb{R}^+ \).

And the lower bound problem has the objective function

\[
\underline{p}_{\text{StopLoss}} = \min \quad \mathbb{E}_F(\phi(X_1 + X_2)),
\]

with the same constraints as (3.44).

The feasibility of problems (3.44) and (3.45) in terms of their moment parameters follows the same rule as for the extreme probability bounds (page 55). That is, the problems (3.44) and (3.45) are feasible if and only if \( \Sigma \) is a positive definite matrix and all elements of \( \Sigma \) are greater than zero, where \( \Sigma \) is, as usual:

\[
\Sigma = \begin{bmatrix}
1 & \mu_1 & \mu_2 \\
\mu_1 & \mu_1^{(2)} & \mu_{12} \\
\mu_2 & \mu_{12} & \mu_2^{(2)}
\end{bmatrix}.
\]

Compared with the previous problems, bounds on stop-loss coverage is relatively easy to compute since \( X_1 \) and \( X_2 \) always appear in the form of \( X_1 + X_2 \) in the objective function (3.43). Therefore, this problem can be considered as a single variable problem by setting \( Z = X_1 + X_2 \). With this transformation, the objective function (3.43) can be written as:

\[
\phi(z) = \begin{cases} 
  b & \text{if } z \geq a + b \\
  z - a & \text{if } a \leq z \leq a + b \\
  0 & \text{if } z \leq a.
\end{cases}
\]
The moments of $Z$ are calculated as follows:

$$
\mu_z = \mu_1 + \mu_2 \quad \text{and} \quad \mu_z^{(2)} = \mu_1^{(2)} + \mu_2^{(2)} + 2\mu_{12}.
$$

We have discussed how to calculate univariate bounds in Chapter 2. As for this specific problem, Cox (1991) provides an explicit solution to the transformed problem (3.46). In this section, we first solve this problem numerically with a SOS solver, and then compare its results with those obtained from Cox’s method to test the robustness of the SOS approach.

By setting $Z = X_1 + X_2$, problem (3.44) is transferred to the univariate bounds problem as

$$
p_{\text{StopLoss}} = \max \ E_F(\phi(Z))
$$

subject to

$$
E_F(Z) = \mu_z \\
E_F(Z^2) = \mu_z^{(2)} \\
F(z) \text{ a probability distribution in } \mathbb{R}^+,
$$

and the lower bound (3.45) is converted to

$$
p_{\text{StopLoss}} = \min \ E_F(\phi(Z)),
$$

with the same constraints as (3.47).

The dual problem of (3.47) is

$$
\bar{d}_{\text{Stoploss}} = \min \quad a_0 + a_1\mu_z + a_2\mu_z^{(2)}
$$

subject to

$$
p(z) \geq \phi(z), \text{ for all } z \geq 0,
$$

and the dual problem of (3.48) is written as

$$
\underline{d}_{\text{Stoploss}} = \max \quad a_0 + a_1\mu_z + a_2\mu_z^{(2)}
$$

subject to

$$
p(z) \leq \phi(z), \text{ for all } z \geq 0.
$$

where $p(z) = a_0 + a_1z + a_2z^2$.

It is easy to see that weak duality holds between (3.47) and (3.49) (or between (3.48) and (3.50)). In Chapter 3.2 (page 53), we discussed the strong duality of bivariate moment problems with bounded objective function, $\phi(x_1, x_2)$. Since the payoff of the stop-loss contract (equation (3.43)) is bounded on $[0, b]$, the strong duality holds. For the transferred problems (3.49) and

---

3Only very few univariate bound problems have explicit solutions, although almost all of them can be solved numerically.
(3.50), the following requirement guarantee $\bar{p} = \bar{d}$ (or $p = d$).

If problem (3.49) is feasible and there exist $a_0, a_1, a_2$ such that

$$p(z) > \phi(z), \text{ for all } z \in \mathbb{R}^+,$$

then $\bar{p} = \bar{d}$. In the analog, for problem (3.50), we reverse the inequality and replace $\bar{p} = \bar{d}$ with $p = d$.

Since $\phi(z)$ is bounded on $[0, b]$, the dual solution $a_0 > b$, and $a_1 = a_2 = 0$ strictly satisfies (i.e., with $>$) the constraint in (3.49) for all $z \in \mathbb{R}^+$. By setting $a_0 < 0$ and $a_1 = a_2 = 0$, the inequality of the lower bound problem (3.50) strictly holds. Thus, as long as problem (3.47) (or problem (3.48)) is feasible, $\bar{p} = \bar{d}$ (or $p = d$) and one can solve (3.49) and (3.50) to obtain the desired semiparametric bounds.

Now, we derive SOS programs to numerically compute $\bar{p}_{\text{StopLoss}}$ and $\underline{p}_{\text{StopLoss}}$ by using SOS programming solvers.

To formulate problem (3.49) as a SOS program, we rewrite the inequality constraint in (3.49) as the following three simultaneous inequalities:

$$p(z) - b \geq 0, \text{ for all } z \in [a + b, \infty)$$
$$p(z) - z + a \geq 0, \text{ for all } z \in [a, a + b]$$
$$p(z) \geq 0, \text{ for all } z \in [0, a].$$

Applying Diananda’s Theorem, problem (3.49) is equivalent to the following SOS program:

$$\bar{d}_{\text{StopLoss}} = \min \quad a_0 + a_1 \mu_z + a_2 \mu_z^{(2)}$$

for which the following are SOS polynomials:

$$p(a + b + z^2) - b$$
$$p(a + b - z^2) - b + z^2$$
$$p(a + z^2) - z^2$$
$$p(a - z^2)$$
$$p(z^2)$$

Thus, if problem (3.44) is feasible, then we can obtain the semiparametric upper bound $\bar{p}_{\text{StopLoss}}$ by solving problem (3.52) with a SOS solver.

The same applies to the dual of the lower bound problem. Problem (3.50) is equivalent to the
following SOS program when Diananda’s Theorem is applied:

\[
d_{\text{StopLoss}} = \max a_0 + a_1 \mu_z + a_2 \mu_z^{(2)}
\]

(3.53)

for which the following are SOS polynomials:

\[
\begin{align*}
& b - p(a + b + z^2) \\
& b - z^2 - p(a + b - z^2) \\
& z^2 - p(a + z^2) \\
& -p(a - z^2) \\
& -p(z^2)
\end{align*}
\]

If problem (3.48) is feasible, then we can numerically obtain the semiparametric lower bound \( \underline{p}_{\text{StopLoss}} \) by solving problem (3.53) with a SOS solver.

In addition, by defining \( \psi(z) = z - \phi(z) \) as follows,

\[
\psi(z) = \begin{cases} 
  z - b & \text{if } z \geq a + b \\
  a & \text{if } a \leq z \leq a + b \\
  z & \text{if } z \leq a
\end{cases}
\]

(3.54)

the lower bound of stop-loss payment \( \underline{p}(\phi) \) can be obtained by solving the upper bound of a transformed problem with objective function \( \psi(z) \); that is, \( \underline{p}(\phi) = \min \{ E_F[\phi(Z)] \} \), equals \( \mu_z \) minus the upper bound of \( \psi(Z) \).

\[
\underline{p}(\phi) = \mu_z - \overline{p}(\psi).
\]

The semiparametric upper bound on \( \overline{p}(\psi) \) can be obtained by solving the following dual problem (3.55) with a SOS solver:

\[
\overline{d}(\psi) = \min y_0 + y_1 \mu_z + y_2 \mu_z^{(2)}
\]

subject to \( p(z) - (z - b) \geq 0, \) for all \( z \in [a + b, \infty) \)

(3.55)

\[
\begin{align*}
& p(z) - a \geq 0, & \text{for all } z \in [a, a + b] \\
& p(z) - z \geq 0, & \text{for all } z \in [0, a].
\end{align*}
\]

Cox (1991)’s Method

Cox (1991) develops an explicit solution to the bounds of the expected claim payment \( E[\psi(Z)] \) of the direct insurer, given mean and variance of \( Z \). \( \overline{p}(\psi) \), the upper bound on \( E[\psi(Z)] \), is described as follows:
(1) If $0 \leq a < \mu_z$, 
\[
\overline{p}(\psi) = \begin{cases} 
\frac{(\mu_z - b)(\mu_z - a)^2 + \mu_z \sigma_z^2}{(\mu_z - a)^2 + \sigma_z^2} & \text{if } a \leq a + b \leq \frac{\sigma_z^2 + \mu_z^2 - a^2}{2(\mu_z - a)}, \\
\frac{1}{2} \left[ \mu_z - a - b + \sqrt{(a + b - \mu_z)^2 + \sigma_z^2} \right] & \text{if } a + b > \frac{\sigma_z^2 + \mu_z^2 - a^2}{2(\mu_z - a)}. 
\end{cases}
\]

where $\sigma_z = \sqrt{\mu_z^2 - \mu_z^2}$.

(2) If $a \geq \mu_z$, the upper bound $\overline{p}(\psi) = \mu_z$.

The lower bound on $E[\psi(Z)]$, $\underline{p}(\psi)$, is described as follows:

(1) If $0 \leq a + b \leq \mu_z$,
\[
\underline{p}(\psi) = \mu_z - b.
\]

(2) If $\mu_z \leq a + b \leq \mu_z + \frac{\sigma_z^2}{\mu_z}$,
\[
\underline{p}(\psi) = \frac{a \mu_z}{a + b}.
\]

(3) If $a + b \geq \mu_z + \frac{\sigma_z^2}{\mu_z}$,
\[
\underline{p}(\psi) = \begin{cases} 
\frac{a \mu_z^2}{\sigma_z^2 + \mu_z^2} & \text{if } 0 \leq a \leq \frac{\mu_z}{2} + \frac{\sigma_z^2}{2\mu_z}, \\
\frac{1}{2} \left[ \mu_z + a - \sqrt{(\mu_z - a)^2 + \sigma_z^2} \right] & \text{if } \frac{\mu_z}{2} + \frac{\sigma_z^2}{2\mu_z} < a \leq \frac{(a + b)^2 - \mu_z^2 - \sigma_z^2}{2(a + b - \mu_z)}, \\
\frac{\mu_z(a + b - \mu_z)^2 + (\mu_z - b)\sigma_z^2}{(a + b - \mu_z)^2 + \sigma_z^2} & \text{if } \frac{(a + b)^2 - \mu_z^2 - \sigma_z^2}{2(a + b - \mu_z)} \leq a 
\end{cases}
\]

After the upper and lower bounds $\overline{p}(\psi)$ and $\underline{p}(\psi)$ are calculated, the bounds on the stop-loss payment $\phi(Z) = Z - \psi(Z)$ can be found by the relations $\overline{p}(\phi) = \mu_z - \overline{p}(\psi)$ and $\underline{p}(\phi) = \mu_z - \underline{p}(\psi)$.

In Section 3.4.3, we use both methods to calculate bounds on stop-loss payments. These methods obtain exactly the same solutions and confirm each other.

3.4 Numerical Examples

In this section, we illustrate the results in Section 3.3 with some relevant numerical examples.

3.4.1 Example of Extreme Probability Bounds

What makes the moment methods valuable for our analysis is that, they do not depend on restrictive assumption to analyze default risk, ruin probability and so on. We show how to find bounds
on the joint probability of extreme events, regardless of the distribution, subject only to moment information. We detail an example to compute the bounds of a joint probability event involving asset returns and insurance margins.

For this example we consider a property/casualty insurance company that faces the problem of managing the risk of unexpectedly high claims and simultaneously suffering unanticipated poor asset returns. This leads us to calculate the bounds on $\Pr(R \leq t_1, M \leq t_2)$ given moment information, where $R$ is the company’s return on its invested assets and $M$ is the margin on its insurance business.

The return $R_i$ of asset $i$ in the insurer’s portfolio is equal to $P_{i,t}/P_{i,t-1} - 1$ where $P_{i,t-1}$ and $P_{i,t}$ denote the prices of asset $i$ at the beginning and the end of the period. We illustrate this with publicly available data on American International Group (AIG). AIG’s asset portfolio return $R$ is the weighted average return of six asset classes: stocks, government bonds, corporate bonds, real estates, mortgages and short-term investments; that is

\[
R = \sum_{i=1}^{6} w_i R_i
\]

\[
= \sum_{i=1}^{6} w_i \left( \frac{P_{i,t}}{P_{i,t-1}} - 1 \right)
\]

\[
= \sum_{i=1}^{6} w_i \frac{P_{i,t}}{P_{i,t-1}} - 1
\]

\[
= X_1 - 1,
\]

where $w_i$ is the weight of asset class $i (i = 1, 2, \ldots, 6)$ in the portfolio and $X_1 = \sum_{i=1}^{6} w_i \frac{P_{i,t}}{P_{i,t-1}}$.

Notice that the following inequalities are equivalent:

\[
R \leq t_1 \iff X_1 \leq t_1 + 1 \quad (3.56)
\]

We make this shift from asset returns to price ratios to apply our SOS results because we need non-negative random variables.

As defined in Chapter 2 Section 2.6.2 (page 40), the margin equals 1 minus the sum of loss ratio and expense ratio, designating the profit of insurance business line.

\[
M = 1 - LR - ER.
\]

Following a standard measure in the insurance literature (Cummins, 1990; Phillips et al., 1998;
Yu and Lin, 2007), we calculate the economic loss ratio as

\[ LR = \frac{\sum_{k=1}^{12} PVF_k \times NLI_k}{\sum_{k=1}^{12} NPE_k}, \]

where \( PVF_k \) is the present value factor for future losses for loss category \( k \), \( NLI_k \) is the net loss incurred for category \( k \), and \( NPE_k \) is the net premium earned for category \( k \) \((k = 1, 2, \ldots, 12)\).\(^4\) The present value factor \( PVF_k \) is calculated from the industry liability payout factor for loss category \( k \) and the term structure of interest rates. The interest rates are the risk-free rates estimated from the U.S. Treasury spot-rate yield curves.\(^5\)

We follow Cummins (1990) in calculating the values of PVF, NLI and NPE. Using the actual earned premium in the denominator and the riskless present value of losses in the numerator allows us to capture changes in loss ratios due to insurance shocks. The insurance company’s liability for future loss payments with respect to its current book of business is included in the product

\[ PVF_k \times NLI_k \]

even for very long term lines of business. This is because the net loss incurred includes not only the observed incurred losses but also a statistical estimate of the incurred but not reported losses. The present value factor is based on the industry payment history rather than the AIG’s own expenses, which is not available. Thus, \( M \) represents the company’s estimates of its net return, although the actual return on the current book may not be realized for many years. That being said, when catastrophic events to occur during the year, they have an immediate impact on reported incurred losses.

Similarly the expense ratio is calculated as follows:

\[ ER = \frac{\sum_{k=1}^{12} NE_k}{\sum_{k=1}^{12} NPW_k}, \]

where \( NE_k \) and \( NPW_k \) are the net expenses and net premium written for the line of business \( k \), respectively.

In order to reformulate the condition \( M \leq t_2 \) so that the condition fits our SOS results, similar to the asset return case, we replace \( M \leq t_2 \) with \( X_2 \leq t_2 + 1 \) where \( X_2 = M + 1 \). Using this with

---

\(^4\)Following the NAIC classifications, we classify AIG’s business into twelve categories. The twelve insurance business categories include farmowners and homeowners multiple peril; private passenger auto liability; workers’ compensation; commercial multiple peril; medical malpractice; special liability; special property; automobile physical damage; fidelity and surety; other; financial guarantee and mortgage guarantee; and other liability and product liability.

\(^5\)Data source: the Federal Reserve Bank of St. Louis’ Federal Reserve Economic Data (FRED).
(3.56) we get the following:

$$\Pr(R \leq t_1, M \leq t_2) = \Pr(X_1 \leq t_1 + 1, X_2 \leq t_2 + 1).$$

The weights $w_i$ of different asset categories are calculated from the quarterly data of the National Association of Insurance Commissioners (NAIC). The quarterly AIG losses, expenses and premiums are also obtained from the NAIC. We use the annualized quarterly returns of the Standard & Poor’s 500 (S&P500), the LB IT government bond index, the domestic high-yield corporate bond index, the NAREIT-All index, the ML mortgage index and the U.S. 30 Day T-Bill as proxies for AIG’s stock returns, government bond returns, corporate bond returns, real estate returns, mortgage returns and short-term investment returns, respectively. In sum, we have 52 quarterly observations from 1991 to 2003. Here are their moments:

| $E(X_1)$ | $E(X_2)$ | $E(X_1^2)$ | $E(X_2^2)$ | $E(X_1X_2)$ | $\text{Cov}(X_1, X_2)$ | $\text{Var}(X_1)$ | $\text{Var}(X_2)$ | $\rho$ |
| 1.0442  | 1.1555  | 1.0967     | 1.3715     | 1.2086      | 0.0021           | 0.0063         | 0.0364         | 0.1387 |

AIG’s average margin on its insurance business ($E(M) = 0.1555$) is higher than its average asset return ($E(R) = 0.0442$), while the margin is more volatile ($\text{Var}(M) > \text{Var}(R)$). Moreover, the asset return and insurance margin are somewhat positively correlated (0.1387). This implies that occasionally AIG’s insurance business and investment performances move in the same direction.

Now we compute bounds on the tail probability $\Pr(R \leq t_1, M \leq t_2)$ using SOS programming. Then we compare it to the bivariate normal cumulative joint probability with the same moments. The upper left plot in Figure 3.1 shows the upper bounds of the joint probability $\Pr(R \leq t_1, M \leq t_2)$ for different values of $t_1$ and $t_2$, and the upper right one is the corresponding bivariate normal cumulative joint probabilities. Since we are looking at low values of $t_1$ and $t_2$ corresponding to joint “extreme” events, it is not surprising that lower bound is zero over this range of their values. The ratios of the upper bounds to the bivariate normal cumulative joint probabilities are shown in the lower graphs.

The ratio is very large when $t_1$ and $t_2$ are very low. For example, consider the event that AIG has no investment earnings and simultaneously it has an aggregate loss on its insurance business. In the model notation this is stated as $R \leq 0, M \leq 0$. From the lower right graph of Figure 3.1, we see that for $t_1 = 0$ and $t_2 = 0$, the upper bound is about 7.2 times higher than the cumulative joint normal probability. That is, the upper bound has a much longer tail than the bivariate normal distribution, so it is possible that the actual underlying joint distribution has a much fatter tail than
the normal. Tail event probabilities can be much larger than the estimates based on the normal distribution with the same moments.

Next, we explore the upper bound implication for the joint probabilities across different values of $t_i$, given $t_j$ is fixed ($i = 1$ or 2, and $i \neq j$). Specifically, we are interested in how the asset return ($t_1$) changes the joint tail probability when the insurance margin ($t_2$) is fixed. That is, we fix the insurance margin at $t_2$ and then solve the upper bound of joint probability $Pr(R \leq t_1, M \leq t_2)$ by changing $t_1$. In Figure 3.2, we set the variable $t_2$ (insurance margin) at six different levels based
Figure 3.2. Each plot shows the upper bound on the joint probability $\Pr(R \leq t_1, M \leq t_2)$ (the upper curve with $-\ast-$) and the bivariate normal cumulative probability with the same moments (the lower curve with $-o-$) for AIG. They are a function of asset return $R$ given an insurance margin $(M)$ level $t_2$. Six graphs fix $t_2$ at $E(M) - k \sqrt{\text{Var}(M)}$ with $k = 0.25, 0.50, 0.75, 1, 1.25$ and $1.5$, $k$ increasing to the right then down. The vertical axis is the probability and the horizontal axis is the return on asset $R$.

on $0.25, 0.5, 0.75, 1, 1.25$ and $1.50$ standard deviations lower than the mean; that is

\[
\begin{align*}
t_2 &= E(M) - 0.25 \sqrt{\text{Var}(M)} = 1.1077, \\
t_2 &= E(M) - 0.50 \sqrt{\text{Var}(M)} = 1.06, \\
t_2 &= E(M) - 0.75 \sqrt{\text{Var}(M)} = 1.0123, \\
t_2 &= E(M) - \sqrt{\text{Var}(M)} = 0.9646, \\
t_2 &= E(M) - 1.25 \sqrt{\text{Var}(M)} = 0.9168, \\
t_2 &= E(M) - 1.50 \sqrt{\text{Var}(M)} = 0.8691. 
\end{align*}
\]
The trend of these graphs are consistent with our expectation. As $t_2$ decreases, the cumulative joint probability levels out at a lower value. For example, when $t_2 = 1.1077$, the upper bound of cumulative joint probability stays at 0.95 after it reaches this level. However, the stable level is only about 0.3 when $t_2 = 0.8691$.

We also compare the upper bounds to the bivariate normal distribution with the same moments. As we expect, the bivariate normal curve is below the upper bound in all graphs. That is, the upper bound has a fatter tail, which suggests a higher ruin probability. In addition, a lower upper bound is associated with a lower normal cumulative probability.

### 3.4.2 Example of VaR Probability Bounds

#### Bounds without Exchange Option Information

Given a specified tail probability $\beta$, the weights $w_1$ and $w_2$, as well as the moment information on $X_1$ and $X_2$, the VaR bound problem finds the upper and lower bounds on $a$ where $\Pr(w_1X_1 + w_2X_2 \leq a) = \beta$. To solve this problem, we first find bounds on $\Pr(w_1X_1 + w_2X_2 \leq a)$ for different values of $a$ and then solve the inverse bounds problem for $a$ given $\beta$.

To show how to solve the bound on VaR, we study a possible extreme scenario in the international stock markets. That is, what may happen if the stock indices of two countries both reach some very low levels. Specifically, we analyze the joint tail probability of total return of a portfolio investing in the S&P500 and Nikkei indices.

First, we calculate the moments of the S&P500 monthly log-return in percentage (denoted $R_{sp}$) and that of the Nikkei (denoted $R_{nk}$) based on the monthly historical data from 1984 to 2006. There are 276 observations in our sample. Their moments are as follows:

\[
\begin{align*}
E(X_1) & = 0.7858 & E(X_1^2) & = 19.2800 \\
E(X_2) & = 0.1907 & E(X_2^2) & = 36.5350 \\
E(X_1X_2) & = 11.4434 & \text{Cov}(X_1, X_2) & = 11.2935 \\
\text{Var}(X_1) & = 18.6641 & \text{Var}(X_2) & = 36.4986 \\
\rho & = 0.432700186.
\end{align*}
\]

On average, the S&P500 monthly log-return (0.79%) is higher than that of Nikkei (0.19%) and the S&P500 is less volatile ($\text{Var}(R_{sp}) < \text{Var}(R_{nk})$). Moreover, they have a positive correlation 0.432700186. This relatively high correlation reflects the impact of economic globalization, which weakens the diversification effect.

Suppose we invest 50% of our assets in the S&P500 and 50% in Nikkei, i.e. $0.5X_1 + 0.5X_2 = 0.5R_{sp} + 0.5R_{nk}$. We calculate the upper and lower bounds on the probability when the portfolio return falls below the level $a$, i.e. $\Pr(0.5R_{sp} + 0.5R_{nk} \leq a)$. The upper and lower lines in Figure 3.3
respectively represent the upper and lower bounds of joint probabilities for different values of \( a \). The upper and lower bounds include all possible joint probabilities, including the bivariate normal joint probability shown as the middle line. This means that although we know only the first two moments, we are sure the probability of the rare event \( 0.5R_{sp} + 0.5R_{nk} \leq a \) is between the upper and lower bounds.

Figure 3.3. Upper and lower bounds for the probability \( \Pr(0.5R_{sp} + 0.5R_{nk} \leq a) \) where \( R_{sp} \) is the monthly log-return on the S&P500 index and \( R_{nk} \) is that of the Nikkei index. The vertical axis is the probability and the horizontal axis stands for different values of \( a \).

Figure 3.3 gives us an idea how likely the return of this portfolio will be lower than \( a \) over a year under different conditions. The inverse problem helps us to obtain the upper and lower bounds of the value-at-risk, \( a \), given a tail probability. Popular left tail levels usually are 1% and 5%. For example, if we focus on the 5%-VaR, the upper bound \( a_L \) tells us that there is a 5% chance the portfolio return would fall below \( a_L = -20\% \) and the lower bound suggests that the VaR reaches a value higher than \( a_U = 1.8\% \) with the same probability. Also notice that the 5% \( \text{VaR}_{0.05} \) of the normal distribution equals to \( a = -7\% \). It falls between the semiparametric lower bound \( a_L \) and upper bound \( a_U \). That is, \( a_L < \text{VaR}_{0.05} < a_U \). For all joint distribution of \( R_{sp} \) and \( R_{nk} \) with the given moments, the normal estimates \( \text{VaR}_{0.05} = -7\% \), which could very much underestimate the danger of a large loss.
Bounds with Exchange Option Information

When we add more constraints to the optimization, the bounds will become tighter. VaR bounds with exchange option information will not wider than those without exchange option information.

However, according to our simulations, not all exchange option constraints can actually improve bounds. Only in the case when one component of the portfolio has relatively high volatility and the other has low volatility, will adding exchange option information improve the bounds. Below is the summary of our simulation results.

<table>
<thead>
<tr>
<th>X₁ volatility</th>
<th>X₂ volatility</th>
<th>Correlation</th>
<th>Improved?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>5</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>6</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

Below we analyze an empirical example when the volatilities of both components of the portfolio are high. The portfolio has the S&P500 Index and the Dow Jones U.S. Small-Cap Index. The Jones U.S. Small-Cap Index is much more volatile than the Nikkei.

The moments are based on the daily historical log-returns (log-price ratios) from February 24, 2000 to October 24, 2007. There are 1,923 observations in our sample. Let \( X₁ \) and \( X₂ \) be the log-return of the S&P500 Index and Dow Jones U.S. Small-Cap Index in percentage per day \( (X_i(t) = 100 \log(S_i(t+1)/S_i(t)) \) for day \( t \). Their moments are as follows:

\[
\begin{align*}
E(X₁) &= 0.0059 & E(X₁^2) &= 1.2158 \\
E(X₂) &= -0.2117 & E(X₂^2) &= 112.8609 \\
E(X₁X₂) &= 1.4161 & \text{Cov}(X₁, X₂) &= 1.41736 \\
\text{Var}(X₁) &= 1.2158 & \text{Var}(X₂) &= 112.81602 \\
\rho &= 0.121021815 & E((X₁ - X₂)^+) &= 0.4464. \\
\end{align*}
\]

(3.57)

Suppose we invest 1/3 of our assets in the S&P500 Index, 1/3 in the Dow Jones U.S. Small-Cap Index and 1/3 in a fixed fund paying a flat 0.01 percent per day. Thus, our portfolio return is \((1/3)X₁ + (1/3)X₂ + (1/3)0.01\). We now calculate the upper and lower bounds for the probability when the portfolio return falls below a given level \( a \), i.e.

\[ \Pr((1/3)X₁ + (1/3)X₂ + (1/3)0.01 \leq a). \]
The corresponding bounds are shown in Figure 3.4. The lines with $-o-$ represent the upper and lower bounds on the VaR probability, without using the exchange option information. These bounds are obtained by setting $y_0 = 0$ in equations (3.37) and (3.42). The lines with $-*-$ represent the upper and lower bounds on the VaR probability, using the exchange option information. Obviously the exchange option information tightens the VaR probability bounds significantly. These semiparametric upper and lower bounds apply to all possible joint probabilities, including the bivariate normal joint probability. The VaR probability corresponding to a normal distribution with the same first and second order moments, using the broken line in the middle. Interestingly, the normal VaR probability lies outside the tighter bounds using the exchange option information. This means that the normal model is not consistent with the sample data which satisfies the constraint $E_F[(X_1 - X_2)^+] = \gamma$.

Figure 3.4. Comparison of VaR probability bounds with and without exchange option information.

Now we use the VaR probability bounds in Figure 3.4 to obtain the upper and lower bounds of the VaR itself, with and without the exchange option price. Figure 3.4 gives us an idea how likely the return of this portfolio will be lower than $a$ in 1 day under different conditions. Consider a 5% VaR. We look at the horizontal line through the 0.05 level on the vertical axis, we see that it
intersects the \( -o- \) curves at \( a \) values of -16 and 1. The best we can say is that

\[-16 < \text{VaR}_{0.05} < 1\]

percent per day. Now using the curves reflecting the exchange option information \( (−∗−) \) and find that

\[-6.2 < \text{VaR}_{0.05} < 0\]

percent per day. Clearly the additional information greatly improves our knowledge of future possible outcomes.

### 3.4.3 Example of Stop-loss Payments

In this section, we find the upper and lower bounds on the expected payment of a stop-loss contract written by a reinsurance company. Suppose AIG sells $1 million new homeowners insurance and $1 million new private passenger auto liability insurance this year. It reinsures claim costs in excess of \( a \) million arising from these two businesses to Swiss Re. Swiss Re pays part of AIG’s claims only if the threshold or deductible \( a \) is reached, subject to a policy limit \( b \) million. The upper and lower bounds on the expected payment of Swiss Re is examined here following Section 3.3.3.

The quarterly data of AIG from 1991 to 2004 are obtained from the NAIC. There are 56 observations from which we calculate the moments of AIG loss payments per $1 million premium earned, respectively, for its homeowners insurance \( (L_{HO}) \) and its private passenger auto liability insurance \( (L_{PPA}) \). Their moments of loss amounts in million dollars are summarized as follows:

\[
\begin{align*}
E(X_1) &= 0.6370 & E(X_2^2) &= 0.9364 \\
E(X_2) &= 0.6844 & E(X_2^2) &= 0.5073 \\
E(X_1X_2) &= 0.4596 & \text{Cov}(X_1, X_2) &= 0.0237 \\
\text{Var}(X_1) &= 0.5306 & \text{Var}(X_2) &= 0.0390 \\
\rho &= 0.1647.
\end{align*}
\]

On average, the expected claim payments of these two lines of business are similar although the homeowners insurance is much more volatile since the homeowners business is more vulnerable to catastrophes and other weather-related claims.

Since the stop-loss payment depends on given levels of the deductible \( a \) and the policy limit \( b \), we only change one parameter (e.g. \( b \)) and fix the other (e.g. \( a \)) to show our upper and lower bounds. Figure 3.5 illustrates the upper and lower bounds with different policy limits \( b \) given a certain deductible level of \( a \). The upper and lower solid lines in each graph stand for the upper and lower bounds numerically solved by the SOS program and the lines with \( -o- \) are the upper and
Figure 3.5. Each plot shows the upper (the top curve in each graph) and lower bounds (the curve in the bottom) on the expected stop-loss payment. They are a function of the policy limit $b$ given a level of the deductible $a$. The solid lines are the upper and lower bounds obtained from the SOS programs. The lines with $-$ show the upper and lower bound solutions based on the Cox (1991)'s explicit formula. Six graphs fix $a$ at 0, 0.25, 0.5, 0.75, 1 and 1.5 million dollars respectively, with $a = 0$ on the upper left and running to the right then down. The vertical axis is the expected payment and the horizontal axis is the policy limit $b$, both in million dollars.

lower bounds computed from Cox (1991)'s method. The upper left graph in Figure 3.5 shows the expected stop-loss payment of Swiss Re to AIG with no deductible ($a = 0$). In the rest five graphs, $a$ increases to the right then down. For the cases that the deductible $a$ is fixed at the level 0.5, 0.75, 1 or 1.5, the solutions of the SOS approach and Cox (1991)'s method are almost identical. When $a = 0$ or 0.25, the lower bound of these two methods matches pretty well while the upper bound
from a SOS solver levels out at a relatively higher value than that of Cox (1991)’s method. It is due to the numerical error of the SOS program.

![Graph showing upper and lower bounds on expected stop-loss payment](image)

**Figure 3.6.** Each plot shows the upper (the top curve in each graph) and lower bounds (the curve in the bottom) on the expected stop-loss payment. They are a function of the deductible $a$ given a level of the policy limit $b$. The solid lines are the upper and lower bounds obtained from the SOS programs. The bubble lines show the upper and lower bound solutions based on the Cox (1991)’s explicit formula. Six graphs fix $b$ at 0, 0.25, 0.5, 0.75, 1 and 1.5 million dollars respectively, with $b = 0$ on the upper left and running to the right then down. The vertical axis is the expected payment and the horizontal axis is the deductible $a$, both in million dollars.

To further show the robustness of SOS program solutions, we consider the upper and lower bounds of a Swiss Re stop-loss policy paying up to a fixed level $b$ while AIG could select different
deductibles $a$. Each graph in Figure 3.6 shows the upper and lower bounds given a certain policy limit $b$ with different deductibles $a$. As we expect, the bounds on Swiss Re’s expected payments increase as the fixed value $b$ increases (i.e. the stop-loss policy covers more losses). Again, bounds calculated from the SOS program and the Cox (1991)’s method remain quantitatively equal. This suggests that the SOS program works pretty well for the stop-loss payment problem. In addition, we should note that SOS program can be flexibly applied to a general class of moment problems, most of which cannot be explicitly solved.

### 3.5 Conclusions

In this chapter, we use an optimization technique known as sum of squares (SOS) programming, to find optimal bounds for the probability of extreme events involving two random variables, given only first and second order moment information. An interesting aspect is that we work solely under the physical measure. This avoids the difficulty of estimating moments of the risk neutral distribution. We extend the application of classical moment problems (or semiparametric methods) to finance, insurance and actuarial science to three extreme probability problems, all taking into account the correlation between different random variables. The first allows us to put “100% confidence intervals” on the probability of joint extreme events. The second finds VaR probability bounds on the sum of two variables. The third computes bounds on the expected payment of a stop-loss payments consisting of two components.

In each case the moment information may be based on historical observations or judgements from scenario analysis. We provide examples to illustrate the potential usefulness of moment methods in assessing probability of rare events. There are other applications where our approach could be useful. For example, this approach can be used to estimate the default probability of fixed-income securities, under incomplete knowledge of the enterprise and economic factors driving the credit risk. In other areas such as inventory and supply chain management, this approach can be applied to find inventory policies that will be applicable to different (unknown) demand distributions in the future. Even when the distributions of the random variables are assumed to be known, this approach can be implemented to measure sensitivity of the given joint probabilities, VaR and expected benefits to model misspecification (Lo, 1987; Hobson et al., 2005).
Chapter 4

Portfolio Optimization with CVaR-like Constraints

In chapters 2 and 3, we have discussed how to calculate bounds on the tail risk measure, value-at-risk (VaR) via moment methods. In this chapter, we will propose a approach to improve portfolio selection in terms of mean-variance-skewness frontier using a coherent risk measure, the conditional VaR (CVaR).

In his original monograph on portfolio selection, Markowitz (1952) discusses the tradeoff between the mean and variance of a portfolio. Since then, especially recently, much attention has been focused on managing the asymmetric distributions to minimize risks with given return goal for investors who have special skewness preferences. To address this issue, Krokhmal et al. (2002) suggest to adding CVaR constraints. Dr. Zuluaga proposes that we extend their approach by imposing CVaR-related functions to the portfolio selection problem. The CVaR approach controls tail risk by maximizing the conditional VaR of the return. The CVaR-like constraints approach reshapes the distribution by adding one or more CVaR-like constraints to the mean-variance optimization problem. This CVaR optimization technique manipulates the portfolio distribution with quantile constraints.

In addition to analyzing the classical asset portfolio, we extend our portfolio risk management to the asset-liability portfolio which considers both the asset return of investments and the liability of the financial institutions. We compare the CVaR-like constraints approach with the traditional Markowitz method, the CVaR approach, the Boyle-Ding approach and the mean-absolute deviation (MAD) approach. Our numerical analysis indicates that the CVaR-like constraints approach is superior in terms of skewness improvement of mean-variance portfolios.
4.1 Introduction

One of the fundamental roles of banks, insurance companies and other financial institutions is to invest in various financial assets. When the portfolio is determined, the investors are interested in its future returns and at the same time keep an eye on the tail probability of its extreme losses. We have discussed how to use moment method to compute 100% confidence intervals on tail probabilities in the previous two chapters. In this chapter, we investigate how to improve portfolio selection, the starting step of portfolio management, considering not only the tail probabilities (VaR) but also the potential of getting higher returns with higher probabilities, i.e., the skewness.

A half century ago, Markowitz (1952) quantified the trade-off between the risk and expected return of a portfolio within a static context. This theoretical framework of portfolio performance assessment has profound impact on portfolio risk management. During the recent two decades, there are two trends of portfolio risk management. One is the broad use of VaR as the tail risk measure. Since the tail risk management becomes more and more crucial for financial institutions (Wright, 2007), people choose VaR, the threshold not exceeded with a given probability defined as the confidence level, to measure the market risk of portfolios. However, the VaR-based risk management, do not capture all aspects of risk. VaR is a quantile measure that disregards information conveyed by the sizes of tail losses. Basak and Shapiro (2001) exhibit that, when a large loss occurs, the loss under VaR-based risk management (VaR-RM) is larger than that when not engaging in the VaR-RM. Moreover, Artzner et al. (1999) show that VaR has undesirable properties; e.g., it is not sub-additive. The VaR of a portfolio may be greater than the sum of portfolio component VaRs.

To overcome the limitations of the VaR-RM, Basak and Shapiro (2001) propose an alternative form of risk management that maintains a given level of CVaR when losses occur. CVaR is also called mean excess loss, mean shortfall, or tail VaR. It is the conditional expected loss (or return) exceeding (or below) the given VaR. CVaR is a more consistent measure of risk than VaR because it is sub-additive and concave (Acerbi and Tasche (2002), Artzner et al. (1999)). It can be optimized using linear programming (LP) and nonsmooth optimization algorithms. Therefore, it can be applied to portfolio selection with very large numbers of instruments and scenarios (Uryasev, 2000) and dramatically reduces computational costs. Moreover, empirical applications show that in contrast to the VaR-RM, losses in the CVaR-based risk management (CVaR-RM) are lower than those without CVaR consideration.

The second trend of portfolio management is to consider higher moments of the portfolio, not only the mean and variance. This argument can traced back to some literature a half century ago. Markowitz (1952), Borch (1969) and Feldstein (1969) argue that introducing skewness of returns adds the dimension needed to improve the approximation provided by the mean and variance. This
suggests that high returns must be sacrificed to gain access to higher positive skewness, that is, greater potential for upside moves (Mitton and Vorkink, 2007; Wright, 2007). Jean (1971), Arditti and Levy (1975), Ingersoll (1975), Kraus and Litzenberger (1976), Simkowitz and Beedles (1978) and Conine and Tamarkin (1981) contribute to the early-stage portfolio analysis including the third moments.

Although the theories on portfolio two-moment or three-moment problems and tail risk management are rich, there are few studies explicitly examining the link between them. To fill this gap, this chapter sheds light on the theoretical and empirical impact of tail risk management on the portfolio efficient frontier. We propose models to construct the portfolio efficient frontier using CVaR-related functions, either by setting a CVaR objective function or by adding CVaR-like constraints to the traditional mean-variance portfolio optimization problem. If portfolio managers disclose and monitor CVaR, their optimal behavior will not only reduce losses in the most adverse states (Basak and Shapiro, 2001), but also maximize the skewness given that, portfolios are not extremely positively skewed (Kane, 1982). Our approaches extend the results of Rockafellar and Uryasev (2000).

In addition to these two approaches, we discuss another two newly raised methods in the area of portfolio risk management. They are the Boyle-Ding (BD) approach and the mean absolute deviation (MAD) approach. The BD approach increases skewness by linearizing skewness within a small interval of the initial portfolio. The rationale behind is that it increases the skewness at the expense of a small increase of the variance. We show how to apply these methods to the asset-liability portfolio management of an insurance company.

By setting the traditional Markowitz approach as the benchmark, we compare these four methods in terms of mean-variance tradeoff as well as the tradeoff between the skewness and variance. Our numerical analysis shows that among the five methods analyzed, the CVaR-like constraint approach is a more effective way to improve the skewness. It does not deviate too much from the traditional MV frontier in terms of mean and variance. The empirical experiments also demonstrate that the CVaR-like constraints approach can be used to successfully manage the asset-liability portfolios of insurance companies.

This chapter is organized as follows: Section 4.2 lays the foundation of the analysis. We introduce the asset-liability portfolio and derive the terms of the optimization problems. Section 4.3 develops the CVaR approach and the CVaR-like constraints approach. Section 4.4 theoretically compares these two CVaR-related approaches with the Boyle-Ding and the mean-absolute-deviation approaches. Section 4.5 presents numerical illustrations with empirical data. Section 4.6 concludes the chapter.
4.2 Portfolio and Efficient Frontier: Descriptions

In this section, we describe the notation we will use throughout the whole chapter. Instead of considering the general asset portfolio optimization, we discuss the asset-liability portfolio problem of insurance companies. The asset-liability portfolio takes both the asset investments and insurance business into account. We can see that the general asset portfolio is a special case of the asset-liability portfolio by setting the parameters corresponding to the insurance business equal to zero.

4.2.1 Definition and Notation

Consider a portfolio with \( n \) components. Each component of the portfolio has \( m \) observations. For simplicity, we assume each period is a year. Let \( R_i \) denote the annual return of component \( i \) \((i = 1, \ldots, n)\). The first three moments of \( R_i \) are as follows:

\[
\mu_i = \mathbb{E}[R_i], \quad \text{for all } i = 1, \ldots, n;
\]
\[
\sigma_{ij} = \mathbb{E}[(R_i - \mu_i)(R_j - \mu_j)], \quad \text{for all } i, j = 1, \ldots, n; \tag{4.1}
\]
\[
\gamma_{ijk} = \mathbb{E}[(R_i - \mu_i)(R_j - \mu_j)(R_k - \mu_k)], \quad \text{for all } i, j, k = 1, \ldots, n.
\]

Let \( r_{il} \) denote the observed value of \( R_i \) in year \( l \) \((l = 1, \ldots, m)\). Given the sample returns \( \{r_{il}\} \), we write the empirical distribution moments (mean, covariance and co-skewness) of \( R_i \) as follows:

\[
\hat{\mu}_i = \frac{1}{m} \sum_{l=1}^{m} r_{il}, \quad \text{for all } i = 1, \ldots, n;
\]
\[
\hat{\sigma}_{ij} = \frac{1}{m} \sum_{l=1}^{m} (r_{il} - \hat{\mu}_i)(r_{jl} - \hat{\mu}_j), \quad \text{for all } i, j = 1, \ldots, n; \tag{4.2}
\]
\[
\hat{\gamma}_{ijk} = \frac{1}{m} \sum_{l=1}^{m} (r_{il} - \hat{\mu}_i)(r_{jl} - \hat{\mu}_j)(r_{kl} - \hat{\mu}_k), \quad \text{for all } i, j, k = 1, \ldots, n.
\]

After obtaining the moments of each component from (4.2), we calculate the portfolio empirical moments as follows. First, let \( x_i \) denote the proportion invested in component \( i \). The vector \( x = [x_1, x_2, \ldots, x_n] \) determines the portfolio. The portfolio model return is \( R(x) = \sum_{i=1}^{n} \mu_i x_i \). Denote the portfolio empirical return in year \( l \) by

\[
\hat{\mu}(x)_l = \sum_{i=1}^{n} r_{il} x_i \quad \text{for all } l = 1, \ldots, m.
\]
The first three empirical moments of the portfolio are equal to

\[ \hat{\mu}(x) = \frac{1}{m} \sum_{l=1}^{m} \hat{\mu}(x)_l = \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{n} r_{il}x_i = \sum_{i=1}^{n} \hat{\mu}_i x_i, \]

\[ \hat{\sigma}^2(x) = \frac{1}{m} \sum_{l=1}^{m} [\hat{\mu}(x)_l - \hat{\mu}(x)]^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\sigma}_{ij} x_i x_j, \]

(4.3)

\[ \frac{1}{m} \sum_{l=1}^{m} [\hat{\mu}(x)_l - \hat{\mu}(x)]^3 = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \hat{\gamma}_{ijk} x_i x_j x_k. \]

4.2.2 Asset-Liability Portfolio

Portfolio theory can be applied to asset-liability management (ALM). An insurance company’s ALM emphasizes the aggregate return earned on the asset side as well as the liability side. Premiums collected at the beginning of the year from several lines of business are invested with additional capital reserve in assets such as stocks, bonds, real estate, etc. We assume the company pays losses and expenses at the end of the year.

At the beginning of a year, the company writes a line of business \( i \) with premium \( \Pi_i \) for \( i = 1, 2, \ldots, k_1 \). The total premium is \( \Pi = \Pi_1 + \cdots + \Pi_{k_1} \). For each line of business, \( (1 + \lambda_i)\Pi_i \) is invested. The total investment is

\[ \sum_{i=1}^{k_1} (1 + \lambda_i)\Pi_i = \sum_{i=1}^{k_1} \Pi_i + \sum_{i=1}^{k_1} \lambda_i \Pi_i = \Pi(1 + \lambda), \]

(4.4)

where \( \lambda = \frac{1}{\Pi} \sum_{i=1}^{k_1} \lambda_i \Pi_i \). The value of \( \Pi_i, \Pi, \lambda_i \) and \( \lambda \) are known at the beginning of the period.

Let \( L_i \) be the loss for line of business \( i \), including claim payments and administrative expenses, so the margin for the line is

\[ M_i = 1 - \frac{L_i}{\Pi_i}. \]

The amount in the company’s favor at the end of the year is \( \Pi_i M_i \) for line \( i \) and the total for all lines is

\[ \sum_{i=1}^{k_1} \Pi_i M_i = \Pi \sum_{i=1}^{k_1} a_i M_i, \]

where the weight of line \( i \) is \( a_i = \Pi_i / \Pi \).

Let \( Y_j \) be the return on asset \( j \) over the year. Denote \( b_j \) the proportion of the total investment \( \Pi(1 + \lambda) \) to be invested in asset \( j \) for \( j = 1, \ldots, k_2 \). The aggregate portfolio for the company consists of returns on investments in assets and returns from its insurance lines of business.
net profit for the year can be written in terms of these returns as follows:

\[
\text{Net Profit} = (1 + \lambda)\Pi \sum_{j=1}^{k_2} b_j (1 + Y_j) - \sum_{i=1}^{k_1} L_i - \lambda\Pi
\]

\[
= (1 + \lambda)\Pi \left[ \sum_{j=1}^{k_2} b_j Y_j + \frac{1}{1 + \lambda} \sum_{i=1}^{k_1} a_i M_i \right],
\]

where \( \sum_{i=1}^{k_1} a_i = 1 \) and \( \sum_{j=1}^{k_2} b_j = 1 \). In order to treat this as a portfolio theory problem, we rename the variables. We have \( n = k_1 + k_2 \) “returns”, including returns of assets and margins of lines of business. We relabel the returns \( R_1, R_2, \ldots, R_n \) with the understanding that the first \( k_1 \) are margins; that is:

\[
R = \begin{bmatrix}
R_1 \\
R_2 \\
\vdots \\
R_{k_1} \\
R_{k_1+1} \\
\vdots \\
R_n
\end{bmatrix} = \begin{bmatrix}
M_1 \\
\vdots \\
M_{k_1} \\
Y_1 \\
\vdots \\
Y_{k_2}
\end{bmatrix}.
\]

The weights are \( x = [x_1, x_2, \ldots, x_n] \), renamed as

\[
x_i = \begin{cases} 
\frac{a_i}{1 + \lambda} & \text{for } i = 1, \ldots, k_1, \\
b_i & \text{for } i = k_1 + 1, \ldots, n.
\end{cases}
\]

Generally, \( a_i \) is given since it’s hard for an insurance company to adjust or even close some lines of its business. So that the first \( k_1 \) weights are the known factors \( \frac{a_i}{1 + \lambda} \) and the next \( k_2 \) are the decision variables. In this notation, the company’s net profit is

\[
\text{Net Profit} = (1 + \lambda)\Pi (x^T R).
\]

Because \( \Pi \) and \( \lambda \) are known at the beginning of the year, as constants, they have no effect on the portfolio optimization process; that is, they have no influence on return maximization; nor do they contribute anything to the variance or higher-order moment constraints of portfolio. Therefore we may focus on the return variable \( x^T R \) in applying portfolio optimization. With this notation, the portfolio optimization is to determine the weights \( x = [x_i]_{i=1}^n \) to minimize the variance \( \text{Var}(x^T R) \) subject to the return and higher-moment constraints.
It is clear that in the short run, the net profit’s uncertainty depends only on the return variable. But in the long run, we allow the $a_i$ to be changed to determine the optimal business strategy of the insurance company in terms of the proportions invested in different lines of business. In Section 4.5, we analyze a numerical example to illustrate this case when the weights of the both assets and lines of business are decision variables.

### 4.2.3 Optimization Problem Description

The classical Markowitz mean-variance (MV) portfolio problem is expressed as follows: Given $\mu$,

$$
\min_{x_1, \ldots, x_n} \quad \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \\
\sum_{i=1}^{n} \mu_i x_i = \mu \\
x_i \geq 0, \quad \text{for all } i = 1, 2, \ldots, n.
$$

The mean-variance frontier consists of the points $(\sigma^2(x), \mu)$, where $\mu$ is the target return varying over a range of feasible values and $\sigma^2(x)$ is the variance corresponding to the solution $x$ of the optimization problem. The constraint $x_i \geq 0$ can be adjusted to allow short selling of the $i$-th asset. Other weight constraints can be added to reflect special proportion requirements on investment.

Now we consider a problem of selecting a portfolio with $k_1$ lines of business and $k_2$ assets ($k_1 + k_2 = n$). If $k_1 = 0$, we solve the general asset portfolio problem. On the other hand, if $k_2 = 0$, we are considering a special case in which the insurance company runs only the business of underwriting insurance policies. As we mentioned, in the short run, the company cannot easily change its business, so the weights of the margins, $x_i$ with $i \leq k_1$ are given constants. When the long-run business strategy is concerned, one can revise the optimization problem by adding the first $k_1$ weights as decision variables and putting one more constraint, $\sum_{i=1}^{k_1} x_i = \frac{1}{1+\lambda}$ into the problem. When the asset-liability portfolio is considered, problem (4.6) is generalized as follows:
Given $\mu$,

$$\min_{x_{k_1+1}, \ldots, x_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j$$

subject to

$$\sum_{i=k_1+1}^{n} x_i = 1$$

$$\sum_{i=1}^{n} \mu_i x_i = \mu,$$

$$x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n,$$

where $k_1$ is the number of lines of business in the portfolio and $k_1 + k_2 = n$. Given a certain level of overall return $\mu$, we can minimize the overall variance $\sigma^2(x)$ to obtain the optimal weights of assets. Obviously, if the portfolio includes only assets ($k_1 = 0$ and $k_2 = n$), we go back to the classical portfolio problem (4.6). Both the assets and lines of business contribute to the variance of asset-liability portfolio.

4.3 Improving Skewness of Mean-Variance Portfolio with CVaR

In MV analysis, variance captures a portfolio’s risk. As a newly introduced risk measure, VaR has been widely used for measuring downside risk and has become a part of the financial regulations in many countries (Jorion, 1997; Dowd, 1998; Saunders, 1999). It measures how the return of an asset or a portfolio is likely to decrease over a certain time period. The $\beta$-level VaR is defined as follows:

$$\alpha_{\beta} = \alpha(x, \beta) = \min\{\alpha \in \mathbb{R} : \Pr(R(x) \leq \alpha) \geq \beta\}.$$ 

The variable $\alpha(x, \beta)$ is the $\beta$-lower quantile of the distribution of the portfolio return $R(x)$. Typically, the quantile $\beta$ is set around 5%. Unfortunately, VaR is not the panacea of risk measurement methodologies. A major technical problem is that VaR is not sub-additive. For example, the variance of the sum of two variables $\text{Var}(A + B)$ could be larger than the sum of these two variables’ variances $\text{Var}(A) + \text{Var}(B)$. This imposes a problem for portfolio risk management because we hope portfolio diversification would reduce risk.

As an improved risk measure, the $\beta$-level CVaR, is the expected portfolio return, conditioned on the portfolio returns being lower than the $\beta$-level VaR over a given period. It is defined as

$$\text{CVaR}_{\beta} = \text{CVaR}(x, \beta) = \mathbb{E}(R(x) | R(x) \leq \alpha(x, \beta)).$$

CVaR has some superior characteristics over variance and VaR (Rockafellar and Uryasev, 2000; Uryasev, 2000; D. Bertsimas and Samarovc, 2004; Wu et al., 2005). Variance is a symmetric
measure and it does not differentiate between the desirable upside and the undesirable downside risks (Wu et al., 2005). For example, $R(x)$ and $-R(x)$ have the same variance.

In contrast, CVaR relies on the left (or right) tail of the distribution, so we can use it to improve a portfolio’s skewness. Compared with VaR, CVaR takes into account not only the probability of a loss but also the size of a loss. CVaR is a coherent risk measure \(^1\) that satisfies properties of monotonicity, sub-additivity, homogeneity, and translational invariance. Some of those desirable properties (e.g. sub-additivity) do not hold for VaR. Other merits of using CVaR include:

- Using CVaR is efficient in computation because it involves a linear program. Therefore, we are able to solve problems with a large number of assets (and/or lines of business). It also provides a closed-form solution to the portfolio problem, which has good computational characteristics versus using variance (Wu et al., 2005).

- The CVaR methods require less data. A long-run window for the historical data is generally not required because we do not need to compute the variance-covariance matrix. Large amounts of longitudinal historical data are typically required to obtain a positive semidefinite variance-covariance matrix, which incurs a higher likelihood of numerical errors. The CVaR methods reduce this risk.

- Using CVaR may have lower transaction costs. The MV models generally suggest a lot of quite small asset (and/or line of business) positions be held in the optimal portfolio. Such a portfolio leads to higher transaction costs because more time and resources must be devoted to managing the larger number of assets (and/or lines of business) in it. This typically will not happen with a linear program as in the CVaR-related approaches. As will be shown in Section 4.5, the number of assets (and/or lines of business) in a CVaR efficient portfolio is usually fewer than those of a MV efficient portfolio which achieves about the same mean and variance objectives.

Krokhmal et al. (2002) suggest adding CVaR constraints to improve the skewness of MV portfolio. Dr. Zuluaga proposes that we extend their approach to other classes of problems with CVaR functions, either as the objective functions or as the additional constraints of the portfolio optimization problems. We show that this approach can be used to maximize reward functions (e.g., expected returns) under CVaR-like constraints as well as maximizing the CVaR directly.

Moreover, it is possible to impose many CVaR constraints with different confidence levels $\beta$’s and shape the return distribution according to the preferences of the decision makers. These preferences are specified directly in percentile terms, compared to the traditional approach, which

\(^1\)A coherent risk measure is a risk measure that satisfies properties of monotonicity, sub-additivity, homogeneity, and translational invariance.
specifies risk preferences in terms of utility functions. For instance, we may require that the mean of the lowest 1%, 5% and 10% returns are limited by some values. In addition, we show how to apply this approach to the asset-liability portfolio problems of insurance companies. This approach provides a new efficient and flexible risk management tool and adds to the ALM literature.

We will first show how to optimize portfolios with CVaR as the objective function, and then extend Krokhmal et al. (2002)'s suggestion to a method which increases the skewness of Markowitz MV portfolios by adding one or more CVaR-like constraints.

### 4.3.1 Optimization with CVaR Objective Function

Krokhmal et al. (2002) introduced the CVaR optimization approach in which CVaR is the objective function. They minimize the CVaR of loss portfolios; the analog for return portfolios is to maximize CVaR. This controls risks and increases the likelihood of getting higher returns. That is, we maximize the worst (left tail) portfolio returns to minimize the expected impact, given an adverse return occurs. This is consistent with the goal of investors or risk managers who are keenly interested in avoiding catastrophic losses. The portfolio allocation model using CVaR as its objective function is: Given $\mu$,

$$
\max_{x_{k_1+1}, \ldots, x_n} \text{CVaR}(x, \beta)
$$

subject to

$$
\sum_{i=k_1+1}^{n} x_i = 1
$$

$$
\sum_{i=1}^{n} \mu_i x_i = \mu
$$

$$
x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n
$$

(4.8)

where $k_1$ is the number of lines of business in the portfolio.

Rockafellar and Uryasev (2000) show that the left-tail $\beta$-level CVaR is the solution to an optimization problem:

$$
\text{CVaR}(x, \beta) = \max_{\alpha} \alpha - \frac{1}{\beta} \mathbb{E} \left[ (\alpha - R(x))^+ \right],
$$

(4.9)

where $(a)^+$ is defined as $\max(a, 0)$. 
With this transformation, we obtain the CVaR portfolio allocation model as follows: Given $\mu$,

$$
\max_{x_{k_1+1}, \ldots, x_n} \quad \alpha - \frac{1}{\beta} \mathbb{E}[(\alpha - R(x))^+]
$$

subject to

$$
\sum_{i=k_1+1}^{n} x_i = 1
$$

$$
\sum_{i=1}^{n} \mu_i x_i = \mu
$$

$$
\alpha \in \mathbb{R}
$$

$$
x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n,
$$

where $k_1$ is the number of lines of business in the portfolio.

Using the historical data $r_{il}$, the empirical distribution version of the portfolio optimization problem is written as: Given $\mu$,

$$
\max_{x_{k_1+1}, \ldots, x_n} \quad \alpha - \frac{1}{\beta} \frac{1}{m} \sum_{l=1}^{m} \left( \alpha - \sum_{i=1}^{n} r_{il} x_i \right)^+
$$

subject to

$$
\sum_{i=k_1+1}^{n} x_i = 1
$$

$$
\sum_{i=1}^{n} \mu_i x_i = \mu
$$

$$
\alpha \in \mathbb{R}
$$

$$
x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n,
$$

where $k_1$ is the number of lines of business in the portfolio.

To deal with “$(\cdot)^+$” in the objective function, we apply the technique proposed by Konno and Yamazaki (1991). This adds $m$ new variables $y_1, \ldots, y_m$ to remove the “$(\cdot)^+$” and it has a linear
objective function. Namely, the problem above is equivalent to the following: Given $\mu$,

$$\max_{\alpha; x_{k_1+1}, \ldots, x_n; y_1, \ldots, y_m} \quad \alpha - \frac{1}{\beta} \frac{1}{m} \sum_{l=1}^{m} y_l$$

subject to

$$y_l \geq \alpha - \sum_{i=1}^{n} r_{il} x_i, \quad \text{for all } l = 1, \ldots, m$$

$$\sum_{i=k_1+1}^{n} x_i = 1$$

$$\sum_{i=1}^{n} \mu_i x_i = \mu$$

$$\alpha \in \mathbb{R}$$

$$x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n$$

$$y_l \geq 0, \quad \text{for all } l = 1, \ldots, m,$$

where $k_1$ is the number of lines of business in the portfolio. For each year $l (l = 1, \ldots, m)$, if $\alpha - \sum_{i=1}^{n} r_{il} x_i \leq 0$, the constraint $y_l \geq \alpha - \sum_{i=1}^{n} r_{il} x_i$ is redundant as $y_l \geq 0$ dominates it. Since $y_l$ is required to be as small as possible to maximize the objective function, we have $y_l = 0$. If $\alpha - \sum_{i=1}^{n} r_{il} x_i \geq 0$, the constraint $y_l \geq \alpha - \sum_{i=1}^{n} r_{il} x_i$ dominates the constraint $y_l \geq 0$. Again we want $y_l$ to be as small as possible, therefore $y_l = \alpha - \sum_{i=1}^{n} r_{il} x_i$. In both cases, the variable $y_l = (\alpha - \sum_{i=1}^{n} r_{il} x_i)^+$. If the portfolio returns are normally distributed, the traditional MV approach and the CVaR approach generate the same efficient frontiers. However, when the return has an asymmetric distribution, the CVaR approach may generate a significantly different frontier from that of Markowitz method (Rockafellar and Uryasev, 2000). If $\beta$ is small (less than 50%), the CVaR optimization technique reshapes the left tail of the distribution and does not significantly affect the right tail. On the contrary, the Markowitz approach uses variance to measure risks, and thus does not differentiate between the left tail and right tails of the distribution. Therefore, when the distributions are skewed (e.g. portfolios with skewed option returns or with skewed margins), these two methods find quite different optimal solutions (Mausser and Rosen, 1999; Larsen et al., 2002). We compare the numerical results of the CVaR approach and traditional Markowitz MV approach in Section 4.5.

### 4.3.2 Optimization with CVaR-like Constraints

Some investors, especially institutional investors, prefer using CVaR to control downward risk and increase skewness, but they are reluctant to deviate far away from the Markowitz frontier. To
achieve this goal, Krokhmal et al. (2002) suggest using CVaR constraints to improve the skewness of MV portfolio. We extend it to a method for increasing the skewness of Markowitz MV portfolios by adding CVaR-like constraints.

Specifically, given \( \beta, w \in \mathbb{R} \) and a sample of asset returns (and/or margins of business lines), we write the sample version of the traditional Markowitz MV model (4.6) with a CVaR constraint as follows: Given \( \mu \),

\[
\min_{x_{k_1+1}, \ldots, x_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{subject to} \quad \text{CVaR}(x, \beta) \geq w \\
\sum_{i=k_1+1}^{n} x_i = 1 \\
\sum_{i=1}^{n} \mu_i x_i = \mu \\
x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n,
\]

where \( k_1 \) is the number of lines of business in the portfolio. The above CVaR constraint ensures the left tail expectation in an amount at least equal to \( w \). Based on the Equation (4.9), the model (4.12) can be written as: Given \( \mu \),

\[
\min_{x_{k_1+1}, \ldots, x_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{subject to} \quad \max_{\alpha} \alpha - \frac{1}{\beta} \frac{1}{m} \sum_{i=1}^{m} \left( \alpha - \sum_{i=1}^{n} r_{ij} x_i \right)^+ \geq w \\
\sum_{i=k_1+1}^{n} x_i = 1 \\
\sum_{i=1}^{n} \mu_i x_i = \mu \\
\alpha \in \mathbb{R} \\
x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n,
\]

where \( k_1 \) is the number of lines of business in the portfolio. Because obtaining a tractable formulation for (4.13) is difficult, Krokhmal et al. (2002) suggest dropping its maximization over \( \alpha \) (see details in Krokhmal et al. (2002)[Theorem 2]) and get an approximation to the \( \beta - \text{CVaR} \) target,
CVaR \((x, \beta) \geq w\). So the optimization problem becomes: Given \(\mu\),

\[
\begin{align*}
\min_{\alpha; x_{k_1+1}, \ldots, x_n} & \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{subject to} & \quad \alpha - \frac{1}{\beta} \frac{1}{m} \sum_{l=1}^{m} \left( \alpha - \sum_{i=1}^{n} r_{il} x_i \right) + \geq w \\
& \quad \sum_{i=k_1+1}^{n} x_i = 1 \\
& \quad \sum_{i=1}^{n} \mu_i x_i = \mu \\
& \quad \alpha \in \mathbb{R} \\
& \quad x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n,
\end{align*}
\]  

where \(k_1\) is the number of lines of business in the portfolio. Notice that the first constraint in the above model is not exactly a CVaR constraint, but a CVaR-like constraint. We call this method the “CVaR-like constraint approach” or “MV + CVaR approach”. Applying the same linearization technique as in (4.11), problem (4.14) is equivalent to this: Given \(\mu\),

\[
\begin{align*}
\min_{\alpha; x_{k_1+1}, \ldots, x_n; y_1, \ldots, y_m} & \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{subject to} & \quad \alpha - \frac{1}{\beta} \frac{1}{m} \sum_{l=1}^{m} y_l \geq w \\
& \quad y_l \geq \alpha - \sum_{i=1}^{n} r_{il} x_i, \quad \text{for all } l = 1, \ldots, m \\
& \quad \sum_{i=k_1+1}^{n} x_i = 1 \\
& \quad \sum_{i=1}^{n} \mu_i x_i = \mu \\
& \quad \alpha \in \mathbb{R} \\
& \quad x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n \\
& \quad y_l \geq 0, \quad \text{for all } l = 1, 2, \ldots, m,
\end{align*}
\]  

where \(k_1\) is the number of lines of business in the portfolio. Notice that model (4.15) is a tractable problem. It has a quadratic convex objective and linear constraints, and thus can be solved as easy as the Markowitz MV problem, so long as \(\Sigma = \{\sigma_{ij}\}\) is positive semidefinite.
As mentioned in Section 4.3, we can add two or more CVaR-like constraints by selecting several different $\beta$-levels and reshape the return distribution according to the investors’ preferences. Specifically, we add $p$ CVaR-like constraints by choosing $p$ quantiles $\beta_1, \beta_2, \ldots, \beta_p \in (0, 1)$ and $p$ corresponding thresholds $w_1, w_2, \ldots, w_p \in \mathbb{R}$. The optimization problem is: Given $\mu$,

$$
\min_{\{\alpha_q, x_{k_1+1}, \ldots, x_n, \{y_q\}\}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j
$$

subject to

$$
\alpha_q - \frac{1}{\beta_q} \frac{1}{m} \sum_{l=1}^{m} y_{ql} \geq w_q, \quad \text{for all } q = 1, 2, \ldots, p
$$

$$
y_{ql} \geq \alpha_q - \sum_{i=1}^{n} r_{il} x_i, \quad \text{for all } l = 1, 2, \ldots, m; q = 1, 2, \ldots, p
$$

$$
\sum_{i=k_1+1}^{n} x_i = 1
$$

$$
\sum_{i=1}^{n} \mu_i x_i = \mu
$$

$$
\alpha_q \in \mathbb{R}, \quad \text{for all } q = 1, 2, \ldots, p
$$

$$
x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n
$$

$$
y_{ql} \geq 0, \quad \text{for all } l = 1, 2, \ldots, m; q = 1, 2, \ldots, p,
$$

(4.16)

where $k_1$ is the number of lines of business in the portfolio. That is, we require that the conditional mean values of the worst $\beta_1, \beta_2, \ldots, \beta_p \in (0, 1)$ returns are limited below by different values of $w_1, w_2, \ldots, w_p \in \mathbb{R}$ based on the investors’ risk tolerance. In addition, our proposed model (4.16) has an additional desirable feature: adding two or more CVaR-like constraints does not significantly increase computational costs while we can achieve portfolio optimization and increase skewness at the same time.

### 4.4 Other Portfolio Optimization Approaches

In this section, we discuss another two portfolio optimization approaches, the Boyle-Ding (BD) approach and the mean-absolute deviation (MAD) approach. The first one also considers portfolio selection in a three-moment world. It replaces the skewness constraint by a set of linear inequalities within a small interval around the original portfolio. This method requires a number of try-and-error experiments to set effective parameters and it performs the optimization iteratively. The MAD approach, instead, linearizes the variance of return to increase optimization efficiency and greatly decrease computational costs. Although the second method does not take investors’ skewness
preferences into account, it is very attractive to those large-scale optimization problems consisting of more than 1000 stocks. In addition to introducing their optimization processes, we theoretically compare these two methods with the CVaR portfolio selection techniques discussed in Section 4.3. The empirical comparison among these four methods are in Section 4.5.

4.4.1 Boyle-Ding Approach

The MV frontier, as it is usually determined, has no explicit reference to skewness. Boyle and Ding (2006) give a method to increase the skewness of a given portfolio $x$, obtaining a new portfolio $x^*$ for which the means are equal, the variances are almost equal, and the skewness of the new portfolio $x^*$ is greater than the skewness of the original $x$. Investors prefer $x^*$ to $x$ because a small increase in variance allows for a relatively large increase in skewness (and greater likelihood of a large return). These conditions can be written as follows:

$$
\begin{align*}
\mu(x^*) &= \mu(x) \\
\sigma^2(x^*) &\geq \sigma^2(x) + f(\epsilon) \\
\frac{1}{m} \sum_{l=1}^{m} [\mu(x^*)_l - \mu(x^*)]^3 &\geq \frac{1}{m} \sum_{l=1}^{m} [\mu(x)_l - \mu(x)]^3 + \delta,
\end{align*}
$$

(4.17)

where both $\epsilon$ and $\delta$ are small positive constant parameters, and $f(.)$ is a monotonic function increasing with $\epsilon$.

The third moment (or skewness) condition in (4.17) is difficult to handle because it is highly non-linear. The common portfolio optimization techniques do not handle such a non-linear constraint. Boyle and Ding (2006) replace this constraint with a set of $m$ linear inequalities. This linear transformation is based on the approximation to $t^3$ obtained by joining the points $(a, a^3)$ and $(b, b^3)$ with a line. In terms of Boyle and Ding (2006)’s notation, $a = t_0 - \epsilon$ and $b = t_0 + \epsilon$, where $\epsilon$ is a small positive number, and

$$
\begin{align*}
t_0 &= \mu(x)_l - \mu(x) = \sum_{i=1}^{n} (r_{il} - \mu_i)x_i = \alpha_l, \\
t &= \mu(x^*)_l - \mu(x^*) = \sum_{i=1}^{n} (r_{il} - \mu_i)x_i^*.
\end{align*}
$$
This gives us

\[ t^3 \approx a^3 + \frac{b^3 - a^3}{b - a} (t - a) = a^3 + [a^2 + ab + b^2] (t - a) \]
\[ = (t_0 - \epsilon)^3 + [(t_0 - \epsilon)^2 + (t_0 - \epsilon)(t_0 + \epsilon) + (t_0 + \epsilon)^2] (t - t_0 + \epsilon) \quad (4.18) \]
\[ = (t_0 - \epsilon)^3 + g(t_0)(t - t_0 + \epsilon). \]

where \( g(t_0) = (t_0 - \epsilon)^2 + (t_0 - \epsilon)(t_0 + \epsilon) + (t_0 + \epsilon)^2. \)

Therefore,

\[
\left( \sum_{i=1}^{n} (r_{il} - \mu_i)x_i^* \right)^3 \approx (\alpha_l - \epsilon)^3 + g(\alpha_l) \left( \sum_{i=1}^{n} (r_{il} - \mu_i)x_i^* - \alpha_l + \epsilon \right) \]
\[ = (\alpha_l - \epsilon)^3 - (\alpha_l - \epsilon)g(\alpha_l) + g(\alpha_l) \sum_{i=1}^{n} (r_{il} - \mu_i)x_i^*, \quad (4.19) \]

This is a good approximation when \( a < t < b \) and \( |b - a| \) is small, i.e., when \( x^* \) satisfies the following inequalities:

\[ \alpha_l - \epsilon < \sum_{i=1}^{n} (r_{il} - \mu_i)x_i^* < \alpha_l + \epsilon, \quad l = 1, \ldots, m. \quad (4.20) \]

The constraints (4.20) are the same as these in Boyle and Ding (2006). This implies that for each observation period \( l \), the mean of the new portfolio cannot deviate from the initial mean by more than \( \pm \epsilon \). And these deviations cancel each other out, so the total mean over the time horizon (\( m \) periods) is unchanged.

Provided the inequalities (4.20) hold for each period \( l \) \( (l = 1, \ldots, m) \), then from (4.18) we have

\[
\sum_{l=1}^{m} \left( \sum_{i=1}^{n} (r_{il} - \mu_i)x_i^* \right)^3 \approx \sum_{l=1}^{m} \left[ (\alpha_l - \epsilon)^3 - (\alpha_l - \epsilon)g(\alpha_l) \right] + \sum_{l=1}^{m} g(\alpha_l) \sum_{i=1}^{n} (r_{il} - \mu_i)x_i^* \]
\[ = C + \sum_{i=1}^{n} c_i x_i^*, \quad (4.21) \]

where

\[
C = \sum_{l=1}^{m} \left[ (\alpha_l - \epsilon)^3 - (\alpha_l - \epsilon)g(\alpha_l) \right] \quad \text{and} \quad c_i = \sum_{l=1}^{m} g(\alpha_l)(r_{il} - \mu_i). \]

Since \( C \) is a constant, it contributes nothing to the portfolio optimization. So maximizing the third moment (or skewness) in (4.17) is equivalent to maximizing \( \sum_{i=1}^{n} c_i x_i^* \).
Applying the same analysis to the original portfolio \( x \), we get:

\[
\sum_{l=1}^{m} \left( \sum_{i=1}^{n} (r_{il} - \mu_{i})x_{i} \right)^{3} \approx C + \sum_{i=1}^{n} c_{i}x_{i} = C + \beta
\]

where

\[
\beta = \sum_{i=1}^{n} c_{i}x_{i}.
\]

When the condition (4.20) holds, the third moment (or skewness) inequality in (4.17) becomes

\[
\sum_{i=1}^{n} c_{i}x_{i}^{*} \geq \beta + \delta,
\]

which is linear.

Boyle and Ding (2006) add a constant \( \delta \geq 0 \) to the right hand side (RHS) of inequality (4.22) to control the increase in skewness in the new portfolio. This is the statement of the new problem: Given \( \mu \),

\[
\begin{align*}
\min_{x_{k_{1}+1}, \ldots, x_{n}} & \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}x_{i}x_{j} \\
\text{subject to} & \quad \sum_{i=1}^{n} x_{i} = 1 \\
& \quad \sum_{i=1}^{n} \mu_{i}x_{i} = \mu \\
& \quad \sum_{i=1}^{n} (r_{il} - \mu_{i})x_{i} \leq \alpha_{l} + \epsilon, \quad \text{for all } l = 1, \ldots, m \\
& \quad \sum_{i=1}^{n} (r_{il} - \mu_{i})x_{i} \geq \alpha_{l} - \epsilon, \quad \text{for all } l = 1, \ldots, m \\
& \quad \sum_{i=1}^{n} c_{i}x_{i} \geq \beta + \delta \\
& \quad x_{i} \geq 0, \quad \text{for all } i = k_{1} + 1, \ldots, n,
\end{align*}
\]

where \( k_{1} \) is the number of lines of business in the portfolio.

The development of imposing the third or higher moments into portfolio selection has mainly been hampered by computational problems (Markowitz, 1991; Mitton and Vorkink, 2007). The computational costs of mean-variance-skewness (MVS) portfolio optimization with the cubic skewness constraint increase exponentially as the number of assets (and/or lines of business) increases.
In contrast, we can solve the MVS portfolio problem with the CVaR-like constraint approach or Boyle-Ding approach more efficiently.

Since the Boyle-Ding approach requires selecting the constant parameters $\epsilon$ and $\delta$ beforehand, one needs several try-and-error experiments to make an effective selection. In addition, Boyle and Ding (2006) suggest the problem be solved iteratively, replacing $x$ by the newly obtained $x^*$ each time, until there is no significant increase in skewness. However, according to our numerical experiments, the iteration process does not significantly improve portfolio selection. In most cases, only the first optimization gives portfolio improvement. In contrast, the CVaR-like constraints approach is easier to implement. In addition, the CVaR-like constraints approach can effectively reshape the portfolio distribution by adding specific quantile constraints with $(\beta, w)$ selected according to the individual’s preferences.

Kane (1982) proved that while a three-moment approximation decision rule improves on the MV solution, it is restricted to portfolios which are not extremely positively skewed. In section 4.5, our numerical analysis shows that as long as the portfolio distribution is not skewed extremely positively, the CVaR-like constraints approach obtains much higher skewness than the Boyle-Ding approach does, with only slight deviation from the Markowitz MV efficient frontier. The experiments also show that the Boyle-Ding approach may only significantly increase skewness of low-variance portfolios. It may not work well when customers prefer relatively high risk. So the MV + CVaR approach is a better choice for the management of high-risk portfolios.

### 4.4.2 Mean-Absolute Deviation Approach

Both MV and MVS approaches require estimation of the variance-covariance matrices, a step that substantially increases computational costs. To handle this difficulty, linear programming, such as mean absolute deviation (MAD) approach, is proposed as a way to improve computational efficiency, especially for the large-scale optimization problems consisting of more than 1000 assets (Konno and Yamazaki, 1991).

Instead of using variance, which is quadratic in the $x$-variables, the mean absolute deviation approach uses the absolute value of deviation to measure the dispersion of the portfolio returns.

$$\text{MAD}(R(x)) = \mathbb{E}(|R(x) - \mathbb{E}(R(x))|) = \mathbb{E}\left(\left|\sum_{i=1}^{n} R_i x_i - \mathbb{E}\left(\sum_{i=1}^{n} R_i x_i\right)\right|\right)$$

The empirical distribution version obtained from the historical data $r_{il}$ is

$$\text{MAD}(R(x)) = \frac{1}{m} \sum_{l=1}^{m} \left|\sum_{i=1}^{n} r_{il} x_i - \sum_{i=1}^{n} \hat{\mu}_i x_i\right| = \frac{1}{m} \sum_{l=1}^{m} \left|\sum_{i=1}^{n} (r_{il} - \hat{\mu}_i) x_i\right|,$$
where

\[ \hat{\mu}_i = \frac{1}{m} \sum_{l=1}^{m} r_{il}. \]

Replacing the variance with MAD in the MV model (4.7), the portfolio optimization problem of the MAD approach is defined as follows: Given \( \mu \),

\[
\begin{align*}
\min_{x_{k_1+1}, \ldots, x_n} & \quad \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{n} (r_{il} - \hat{\mu}_i)x_i \\
\text{subject to} & \quad \sum_{i=k_1+1}^{n} x_i = 1 \\
& \quad \sum_{i=1}^{n} \hat{\mu}_i x_i = \mu \\
& \quad x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n,
\end{align*}
\]

where \( k_1 \) is the number of lines of business in the portfolio. By adding positive variables \( y_l \) (\( l = 1, 2, \ldots, m \)), the above problem is linearized as follows: Given \( \mu \),

\[
\begin{align*}
\min_{x_{k_1+1}, \ldots, x_n; y_1, \ldots, y_m} & \quad \frac{1}{m} \sum_{l=1}^{m} y_l \\
\text{subject to} & \quad y_l \geq \sum_{i=1}^{n} (r_{il} - \hat{\mu}_i)x_i, \quad \text{for all } l = 1, \ldots, m \\
& \quad y_l \geq -\sum_{i=1}^{n} (r_{il} - \hat{\mu}_i)x_i, \quad \text{for all } l = 1, \ldots, m \\
& \quad \sum_{i=k_1+1}^{n} x_i = 1 \\
& \quad \sum_{i=1}^{n} \hat{\mu}_i x_i = \mu \\
& \quad x_i \geq 0, \quad \text{for all } i = k_1 + 1, \ldots, n \\
& \quad y_l \geq 0, \quad \text{for all } l = 1, \ldots, m,
\end{align*}
\]

where \( k_1 \) is the number of lines of business in the portfolio. Notice that for a given year \( l \), if \( \sum_{i=1}^{n} (r_{il} - \hat{\mu}_i)x_i \leq 0 \), the constraint \( y_l \geq \sum_{i=1}^{n} (r_{il} - \hat{\mu}_i)x_i \) is redundant as the constraint \( y_l \geq 0 \) dominates it. At the same time, the other constraint \( y_l \geq -\sum_{i=1}^{n} (r_{il} - \hat{\mu}_i)x_i \) ensures \( y_l \geq |\sum_{i=1}^{n} (r_{il} - \hat{\mu}_i)x_i| \). Since we are minimizing the sum of positive variables \( y_l \)'s, the optimal solution
is \( y_l = |\sum_{i=1}^{n} (r_{il} - \hat{\mu}_i) x_i | \). A similar argument follows when \( \sum_{i=1}^{n} (r_{il} - \hat{\mu}_i) x_i \geq 0 \).

Konno and Yamazaki (1991) prove that if the return is normally distributed, the mean absolute deviation of the portfolio return is proportional to its variance. This means that if the return distribution is close to normal, which is asymptotically true when the number of assets in a portfolio becomes large, the mean absolute deviation approach will obtain portfolios close to those of Markowitz model. However, for small-scale portfolio problems, since the return data are generally not normally distributed, the MAD approach may generate dramatically different MV frontier from the traditional one. Our empirical examples in Section 4.5 shows that the difference between the solutions of MV and MAD models is significant.

Existing theories and empirical evidence do not always support the MAD approach. For example, Lee (1977) shows that the MAD approach may be subject to functional form bias and is not a good risk surrogate. Compared with MAD, CVaR is a coherent risk measure supported by many theoretical and empirical studies (Artzner et al., 1999; Pflug, 2000; Rockafellar and Uryasev, 2000; Acerbi et al., 2001; Acerbi and Tasche, 2002; D. Bertsimas and Samarovc, 2004; Wu et al., 2005). Moreover, unlike variance or MAD, CVaR differentiates the left tail from the right tail of the distribution of portfolio return.

4.5 Empirical Illustration: Multiple Assets and Lines of business

In this section, we first compute the optimal portfolios of five assets \((k_1 = 0 \text{ and } k_2 = 5)\) based on the CVaR, MV + CVaR, Boyle-Ding MVS and MAD approaches, respectively, using yearly data ranging from 1980 to 2005 \((m = 26)\). Then we extend our comparison to 20 assets \((k_1 = 0 \text{ and } k_2 = 20)\). To illustrate asset-liability portfolio selection, we select fourteen lines of business \((k_1 = 14)\) and five assets \((k_2 = 5)\). To evaluate the performance of the CVaR, MV + CVaR, Boyle-Ding MVS and MAD approaches, we plot their efficient frontiers and compare them with the traditional MV frontier. In addition, we compare their skewness-variance graphs and asset mix plots. No borrowing or short selling is allowed in these illustrations.

Table 4.1 summarizes statistics (sample mean, variance and skewness) of the annual returns of 20 assets and the annual margins of 14 lines of business in our examples. These are industry data provided by the reinsurance company General Re-New England Asset Management, Inc. (GR-NEAM). See the Appendix E, page 125, for a longer description of the assets and lines of business.

Of the assets, the S&P 500 has the highest average rate of return and the lowest skewness. It is the only asset that was negatively skewed over this time period. This is consistent with the obser-
Table 4.1. Descriptive Statistics of assets and lines of business from 1980 to 2005

<table>
<thead>
<tr>
<th>Assets</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Lines</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSY: 1-3</td>
<td>0.0794</td>
<td>0.0022</td>
<td>0.6774</td>
<td>Cml Prop</td>
<td>0.0030</td>
<td>0.0101</td>
<td>0.2743</td>
</tr>
<tr>
<td>TSY: 7-10</td>
<td>0.0983</td>
<td>0.0090</td>
<td>0.5775</td>
<td>Allied</td>
<td>-0.0732</td>
<td>0.0319</td>
<td>-0.5677</td>
</tr>
<tr>
<td>MBS</td>
<td>0.0976</td>
<td>0.0075</td>
<td>1.9130</td>
<td>Hm/Fr</td>
<td>-0.0978</td>
<td>0.0143</td>
<td>-2.2626</td>
</tr>
<tr>
<td>Crude</td>
<td>0.0766</td>
<td>0.0757</td>
<td>0.2840</td>
<td>CMP</td>
<td>-0.1146</td>
<td>0.0114</td>
<td>0.0493</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.1431</td>
<td>0.0259</td>
<td>-0.5031</td>
<td>Comp</td>
<td>-0.1225</td>
<td>0.0062</td>
<td>0.1934</td>
</tr>
<tr>
<td>Agcy 1-3</td>
<td>0.0817</td>
<td>0.0023</td>
<td>0.7848</td>
<td>GL</td>
<td>-0.2084</td>
<td>0.0173</td>
<td>-0.8365</td>
</tr>
<tr>
<td>Agcy 7-10</td>
<td>0.0999</td>
<td>0.0080</td>
<td>0.8760</td>
<td>Med/Prof</td>
<td>-0.2607</td>
<td>0.0441</td>
<td>-0.2115</td>
</tr>
<tr>
<td>Corp AAA 3-5</td>
<td>0.0897</td>
<td>0.0035</td>
<td>0.9103</td>
<td>PPAuto</td>
<td>-0.0402</td>
<td>0.0018</td>
<td>0.5091</td>
</tr>
<tr>
<td>Corp AA 3-5</td>
<td>0.0931</td>
<td>0.0037</td>
<td>0.9236</td>
<td>Cauto</td>
<td>-0.0830</td>
<td>0.0086</td>
<td>-0.4159</td>
</tr>
<tr>
<td>Corp A 3-5</td>
<td>0.0950</td>
<td>0.0038</td>
<td>1.0309</td>
<td>FSB</td>
<td>0.1106</td>
<td>0.0168</td>
<td>-0.2171</td>
</tr>
<tr>
<td>Corp BBB 3-5</td>
<td>0.0954</td>
<td>0.0038</td>
<td>1.0475</td>
<td>BC/BS</td>
<td>0.0060</td>
<td>0.0007</td>
<td>-0.6929</td>
</tr>
<tr>
<td>Corp HYld</td>
<td>0.1003</td>
<td>0.0102</td>
<td>0.9256</td>
<td>PCHlth</td>
<td>-0.0551</td>
<td>0.0042</td>
<td>-0.8627</td>
</tr>
<tr>
<td>Sovrgn: Inter</td>
<td>0.1057</td>
<td>0.0042</td>
<td>0.6630</td>
<td>Reins</td>
<td>-0.1607</td>
<td>0.0173</td>
<td>-1.9770</td>
</tr>
<tr>
<td>Yankee</td>
<td>0.0991</td>
<td>0.0068</td>
<td>0.8512</td>
<td>Other</td>
<td>-0.0790</td>
<td>0.0098</td>
<td>0.0760</td>
</tr>
<tr>
<td>ABS</td>
<td>0.0801</td>
<td>0.0018</td>
<td>0.6451</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muni 3-5 Yrs</td>
<td>0.0580</td>
<td>0.0012</td>
<td>0.4472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commodities</td>
<td>0.0269</td>
<td>0.0007</td>
<td>1.7360</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lumber</td>
<td>0.0254</td>
<td>0.0027</td>
<td>1.4153</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currency</td>
<td>0.0114</td>
<td>0.0109</td>
<td>0.4489</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML Convert</td>
<td>0.1228</td>
<td>0.0190</td>
<td>0.0225</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The observations made by David (1997). He concludes that stock market returns exhibit negative skewness and that large negative returns are more common than large positive ones. Moreover, mortgage-backed securities have the highest positive skewness. However, eight of the 14 insurance lines were negatively skewed. The negative skewness suggests that the margins might be pulled down by large losses.

**Example 1.** We first examine a portfolio with five assets ($k_1 = 0$ and $k_2 = 5$). These five assets include short-term US Treasury bills, long-term US Treasury bonds, mortgage-backed securities, crude oil futures and the S&P 500 stock index. Our observation period is from 1980 to 2005 ($m = 26$). The optimization problem is to determine the weights of these five assets, i.e., $x_i$ for $i = 1, \ldots, 5$. The portfolio expected return is

$$\sum_{i=1}^{5} \mu_i x_i = 0.0794 x_1 + 0.0983 x_2 + 0.0976 x_3 + 0.0766 x_4 + 0.1431 x_5.$$ 

We apply the CVaR approach with the model (4.11) at the $\beta = 5\%$ level, and solve the model (4.15) to obtain MV + CVaR optimal portfolios. As for the MV + CVaR approach, the parameter
where $\text{CVaR}^{\text{MV}}_{0.05}(r)$ is the empirical 5%-level CVaR obtained from Markowitz MV optimal portfolio with the same expected return $\mu$. The construction of $w$ is reasonable because the empirical 5%-level CVaR obtained by the MV + CVaR approach should be set close to and, a little larger than its Markowitz MV counterpart (in order to “maximize” CVaR in the model 4.12). Based on a MV portfolio, we apply the Boyle-Ding approach to find a new portfolio that has the same mean, approximately same variance, and higher skewness, by using equation (4.23) with parameters $\epsilon = 0.2$ and $\delta = 0.0001$. We also solve the MAD and traditional Markowitz portfolio optimization problems by using models (4.25) and (4.6), respectively, for the same expected return level $\mu$ and variance-covariance matrix $\Sigma$.

After obtaining the optimal weights $x^*_i$ ($i = 1, 2, \ldots, 5$) for the five methods, respectively, we plotted their efficient frontiers and skewness-variance graphs in Figures 4.1 and 4.2. Figure 4.1 shows the mean-variance frontiers of the Markowitz Mean-Variance approach (“Traditional MV”), the 5%-level CVaR approach (“CVaR”), the Markowitz Mean-Variance approach with 5%-level CVaR-like constraint (“MV + CVaR”), the Boyle-Ding Mean-Variance-Skewness approach (“BD”) and the Mean-Absolute Deviation approach (“MAD”). Except the 5%-level CVaR method, all the other four methods obtain essentially same mean-variance (MV) frontiers. The frontier of the 5%-level CVaR method deviates significantly from the Markowitz one, especially in the low level variance range.

The CVaR approach tends to maximize the expected return below a given level of VaR; that is, it chooses a portfolio which realizes the best outcome of all outcomes below the given VaR level. If the distribution is compact, i.e., low variance, there is more room for the CVaR approach to shift the distribution to the right. Therefore, in this case, it has more flexibility to reshape the tail to increase the skewness of the portfolios, but at some time it sacrifices more portfolio efficiency. On the other hand, given a level of expected return, the CVaR approach minimizes adverse situations. Its goal is allied with the objective of the traditional MV method and keeps the CVaR frontier close to the MV frontier. Because management of adverse tail risk is crucial for all financial institutions, the CVaR portfolios seem to be acceptable, although they are less efficient. Whether the CVaR approach is a commendable technique depends not only on the extent to which the investors are willing to deviate from the traditional MV frontier, but also on the investors’ risk tolerances. Also notice that, as we expected, the frontier of the MV + CVaR approach is almost the same as that of

---

2 The problem is sensitive to the parameters $\epsilon$ and $\delta$. For example, with $\epsilon = 0.3$ and $\delta = 0$, the BD solutions are essentially identical to those obtained by the traditional Markowitz approach.

3 People need to maximizes CVaR when the return portfolios are considered. However, for the loss portfolios, one should minimize the corresponding CVaR and shift the tail to the left.
Figure 4.1. The efficient frontiers of 5-asset portfolios. The vertical axis stands for the expected returns and the horizontal axis is for variances.

the Markowitz MV because it is derived from the traditional MV by adding more constraints to the MV problem.

Figure 4.2 compares the 5-asset skewness-variance graphs of the five approaches. With a reasonable sacrifice of the return variance, the CVaR, MAD and MV + CVaR approaches all have higher skewness than the Markowitz approach. The skewness is increased greatly for the CVaR approach but not much for the Boyle-Ding MVS. Figure 4.2 suggests that those two CVaR-related approaches not only achieve left-tail risk management but also have higher skewness. That is, these two methods allow financial institutions to enjoy more potential for higher returns. Whether to choose CVaR or MV + CVaR depends on how far investors are willing to deviate from the Markowitz MV frontier and their preferences in skewness. In contrast, the skewness of the MAD approach is not always above the Markowitz skewness-variance curve. It implies that the MAD approach may be subject to functional form bias.

We also plot in Figure 4.3 the asset mix for the 20 efficient portfolios for these five methods. The horizontal axis shows only the solution number; the return and variance increase as the solution number increases. We can think of the horizontal axis representing either the return or the variance. The asset “TSY: 1-3” stands for the short-term US Treasury bill; “TSY: 7-10” is the long-term US Treasury bond; “MBS” is the mortgage-backed security; “Crude” is the crude oil future; and “S&P” is the S&P 500 stock index. As the required return increases, the mix shifts from bonds to equity as the weight of MBS first rises and then falls. None of the portfolios in the MV, BD and MAD approaches contain a lot of crude oil futures and those based on CVaR have none.
Interestingly, the CVaR approach requires no more than three assets in all efficient portfolios. In contrast, each of the other four methods requires all five assets to form the efficient frontier. Consistent with the existing literature, our research suggests that the CVaR approach eliminates the non-significant holding of certain small assets and thus reduces transaction and administrative costs. Figure 4.3 also shows the source of skewness. Although the five methods have similar holdings in the S&P 500, the CVaR approach and MV + CVaR approach invest relatively more in the long-term US Treasury bonds and less in crude oil futures. Since the skewness of long-term US Treasury bonds is higher than that of crude oil futures, this confirms our result shown in Figure 4.2.

**Example 2.** More assets are included in the portfolio this time. In addition to considering the 5 assets in Example 1, we include another 15 assets. That is, we expand the sample to 20 assets \(k_1 = 0\) and \(k_2 = 20\). For data statistics and description, see page 107. All parameters are kept same as in Example 1. The mean-variance frontiers and skewness-variance graphs based on the five approaches are shown in Figures 4.4 and 4.5. These graphs are similar to those in Example 1. Specifically, the CVaR approach is the least efficient one among the five in terms of the mean-variance tradeoff, but it offers the highest skewness. Skewness of the MV + CVaR approach is higher than those of the MV and Boyle-Ding approaches although its portfolios are relatively less efficient. Again, the MAD approach is the least desirable one in terms of skewness, especially for the higher variance portfolios.

**Example 3.** In this example, we study the ALM problem by maximizing the overall profits of assets and lines of business. We use 14 lines of business \(k_1 = 14\) discussed earlier and the same
Figure 4.3. The 5-asset mix for the efficient portfolios. The vertical axis stands for the weight of each asset, and the horizontal axis is the solution number for the efficient portfolios in Figure 4.1.

Five assets as in Example 1 ($k_2 = 5$). For data statistics and description, please go back to page 107. Keeping the same parameters as in Example 1 and 2 ($\beta = 5\%$, $\epsilon = 0.2$, and $\delta = 0.0001$) and setting $\lambda = 0$ in equation (4.4), the mean-variance and skewness-variance efficient frontiers are shown in Figure 4.6 and Figure 4.7, respectively.

Figure 4.7 indicates that the CVaR approach still provides the highest skewness. Compared with the traditional Markowitz approach, the skewness improvement of the MV + CVaR approach is less significant than that in Examples 1 and 2. However, in most cases, the skewness of the MV + CVaR portfolios is still higher than that of the Boyle-Ding MVS portfolios. Similar to the
results in Examples 1 and 2, the MAD skewness-variance line jumps up and down relative to the Markowitz curve. These erratic results further confirm that the MAD approach is not as a good method as the other methods.
4.6 Conclusion

In this chapter, we propose two CVaR-related approaches to improve the skewness of the mean-variance portfolio. The CVaR approach sets a \( \beta \)-level CVaR as the objective function and maximize the worst case on the left tail (if \( \beta \) is small) of the return distribution. The CVaR-like constraints approach imposes a set of CVaR-like constraints to the traditional Markowitz problem to reshape
the distribution of return based on the investors’ risk preferences. Both methods take advantage of the CVaR’s ability to manage the asymmetry of the distribution. In addition, we investigate another two newly raised methods in the area of portfolio risk management. They are the Boyle-Ding (BD) approach and the mean absolute deviation (MAD) approach.

We compare the CVaR-related approaches, i.e., the CVaR approach and the CVaR-like constraints approach with the Markowitz (1952) approach, the Boyle and Ding (2006) approach and the mean-variance-skewness (MVS) approach, first theoretically and then numerically. Our numerical experiments provide empirical support to the superiority of CVaR-like constraints approach over its alternatives.
Appendix A

Bounds on $\Pr(X_1 \geq t_1, X_2 \geq t_2)$

We consider the problem of finding sharp upper and lower bounds on the probability $\Pr(X_1 \geq t_1 \text{ and } X_2 \geq t_2)$ of two non-negative random variables $X_1, X_2$, attaining values higher than or equal to $t_1, t_2 \in \mathbb{R}^+$ respectively, given up to second order moment information (means, variances, and covariance) on $X_1, X_2$, without making any other assumption on the distribution of the random variables $X_1, X_2$. Finding the sharp upper and lower semiparametric bounds for this problem can be (respectively) formulated as the following optimization problems, obtained by setting in problem (3.1) (Section 3.2) $\phi(X_1, X_2) = I\{X_1 \geq t_1 \text{ and } X_2 \geq t_2\}$, and $D = \mathbb{R}^{+2}$:

$$
\overline{p}_{\text{Survival}} = \max \quad \mathbb{E}_F(I\{X_1 \geq t_1 \text{ and } X_2 \geq t_2\})
$$

subject to

$$
\mathbb{E}_F(X_i) = \mu_i, \quad \mu_i = \mathbb{E}_F(X_i), \quad i = 1, 2, \quad \mathbb{E}_F(X_1X_2) = \mu_{12},
$$

$F(x_1, x_2)$ a probability distribution on $\mathbb{R}^{+2}$, \hfill (A.1)

and

$$
\underline{p}_{\text{Survival}} = \min \quad \mathbb{E}_F(I\{X_1 \geq t_1 \text{ and } X_2 \geq t_2\})
$$

subject to

$$
\mathbb{E}_F(X_i) = \mu_i, \quad \mu_i = \mathbb{E}_F(X_i), \quad i = 1, 2, \quad \mathbb{E}_F(X_1X_2) = \mu_{12},
$$

$F(x_1, x_2)$ a probability distribution on $\mathbb{R}^{+2}$. \hfill (A.2)

Same as for its complementary problem of bounds on $\Pr(X_1 \leq t_1, X_2 \leq t_2)$ (page 55) in Chapter 3, problems (A.1) and (A.2) are feasible if and only if the moment matrix $\Sigma$ is a positive
definite matrix and all elements of $\Sigma$ are greater then zero, where $\Sigma$ is

$$
\Sigma = \begin{bmatrix}
1 & \mu_1 & \mu_2 \\
\mu_1 & \mu_1^{(2)} & \mu_{12} \\
\mu_2 & \mu_{12} & \mu_2^{(2)}
\end{bmatrix}.
$$

Now we derive SOS programs to numerically compute $\underline{p}$ and $\bar{p}$ by using SOS programming solvers.

**Upper bound**

We begin by stating the dual problem of (A.1):

$$
\bar{d}_{\text{Survival}} = \min \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}
$$

subject to $p(x_1, x_2) \geq \mathbb{1}_{\{x_1 \geq t_1 \text{ and } x_2 \geq t_2\}}$, for all $x_1, x_2 \geq 0$.  \hfill (A.3)

The constraint in (A.3) is equivalent to

$$
p(x_1 + t_1, x_2 + t_2) - 1 \geq 0, \text{ for all } x_1, x_2 \geq 0
$$

$$
p(x_1, x_2) \geq 0, \quad \text{for all } x_1, x_2 \geq 0.
$$

Now we apply Diananda’s Theorem (Theorem 2) to the above inequalities, then the problem (A.3) is equivalent to this;

$$
\bar{d}_{\text{Survival}} = \min \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}
$$

subject to $p(x_1^2 + t_1, x_2^2 + t_2) - 1$ is a SOS polynomial

subject to $p(x_1^2, x_2^2)$ is a SOS polynomial.  \hfill (A.4)

This problem can be solved with SOS solvers as before. Thus if (A.1) is feasible, then we can solve (A.4) numerically to obtain the semiparametric bound on $\Pr(X_1 \geq t_1, X_2 \geq t_2)$.

**Lower bound**

We begin by stating the dual problem of (A.2):

$$
\underline{d}_{\text{Survival}} = \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}
$$

subject to $p(x_1, x_2) \leq \mathbb{1}_{\{x_1 \geq t_1 \text{ and } x_2 \geq t_2\}}$, for all $x_1, x_2 \geq 0$.  \hfill (A.5)
APPENDIX A. BOUNDS ON \( \Pr(X_1 \geq t_1, X_2 \geq t_2) \)

The constraint in problem (A.5) is equivalent to

\[
\begin{align*}
p(x_1, x_2) &\leq 1, \text{ for all } x_1 \geq t_1, x_2 \geq t_2 \\
p(x_1, x_2) &\leq 0, \text{ for all } x_1 \geq 0, 0 \leq x_2 \leq t_2, \\
p(x_1, x_2) &\leq 0, \text{ for all } 0 \leq x_1 \leq t_1, x_2 \geq 0.
\end{align*}
\] (A.6)

Although the first constraint of (A.6) can be easily handled by substituting \( x_i + t_i \) for \( x_i \) (\( i = 1, 2 \)), the last two constraints are difficult to reformulate as SOS constraints. Since there is no linear transformation from \( x_1 \geq 0, 0 \leq x_2 \leq t_2 \) (or \( 0 \leq x_1 \leq t_1, x_2 \geq 0 \)) to \( \mathbb{R}^{+2} \), we change the problem to end up with a SOS program that either exactly or approximately solves problem (A.6). Specifically, consider the following problem related to (A.6):

\[
d_{\text{Survival}}' = \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}
\]

subject to

\[
\begin{align*}
p(x_1, x_2) &\leq 1, \text{ for all } x_1 \geq t_1, x_2 \geq t_2 \\
p(x_1, x_2) &\leq 0, \text{ for all } x_1 \geq 0, x_2 \leq t_2, \\
p(x_1, x_2) &\leq 0, \text{ for all } 0 \leq x_1 \leq t_1, x_2 \geq 0.
\end{align*}
\] (A.7)

The constraints of (A.7) are weaker than those of (A.6) since the last two constraints of (A.7) include more values of \( x_1 \) and \( x_2 \). Thus, \( d_{\text{Survival}}' \) is a lower bound on \( d_{\text{Survival}} \); that is \( d_{\text{Survival}} \leq d_{\text{Survival}}' \). Using substitutions \( x_1 \rightarrow x_1 + t_1, x_2 \rightarrow x_2 + t_2 \) to the first constraint of (A.7) and \( x_2 \rightarrow t_2 - x_2, x_1 \rightarrow t_1 + x_1 \) to the second and third constraints respectively, it follows that problem (A.7) is equivalent to the following SOS program when Diananda’s Theorem is applied:

\[
d_{\text{Survival}}' = \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}
\]

subject to

\[
\begin{align*}
1 - p(x_1^2 + t_1, x_2^2 + t_2) &\text{ is a SOS polynomial} \\
-p(x_1^2 - x_2^2) &\text{ is a SOS polynomial} \\
-p(t_1 - x_1^2, x_2^2) &\text{ is a SOS polynomial.}
\end{align*}
\] (A.8)

As before, we can solve SOS program (A.8) with a SOS programming solver. Thus, if problem (A.2) is feasible, then we can approximate the ruin probability semiparametric lower bound on the survival probability, \( \Pr(X_1 \geq t_1, X_2 \geq t_2) \geq d_{\text{Survival}}' \), by solving problem (A.8) with a SOS solver. Furthermore, notice that by solving (A.4) and (A.8) we obtain a “100\% confidence interval” of the joint survival probability; that is

\[
d_{\text{Survival}}' \leq \Pr(X_1 \geq t_1 \text{ and } X_2 \geq t_2) \leq d_{\text{Survival}}'
\]
given up to the second order moment information on the non-negative random variables \( X_1, X_2 \).
APPENDIX A. BOUNDS ON $\Pr(X_1 \geq t_1, X_2 \geq t_2)$

Figure A.1. Bounds on $\Pr(X_1 \geq t_1, X_2 \geq t_2)$. The left and right graphs show bounds with covariance of $X_1$ and $X_2$ equals 0.5 and -1, respectively. The vertical axis stands for probability, and the horizontal axis is the number of standard deviations from the mean, $z$. That is, $t_1 = \mu_1 + z\sigma_1$ and $t_2 = \mu_2 + z\sigma_2$.

Examples

To focus on the joint right-tail events, we set both $t_1$ and $t_2$ at very high level. It doesn’t make too much sense to figure out the upper bound of $\Pr(X_1 \geq t_1, X_2 \geq t_2)$ when the two variables are negatively correlated. Because in this case, we will be asking for the highest probability of both variables being very large, which should be near zero above the mean of the random variables since they are negatively correlated. Therefore, the upper bounds on joint survival probabilities make more sense for two random variables that are positively correlated, or have a low negative correlation.

Considering the following example:

$$
\begin{align*}
E(X_1) &= 1 \\
E(X_2) &= 1 \\
E(X_1^2) &= 3 \\
E(X_2^2) &= 3 \\
E(X_1X_2) &= 1.5
\end{align*}
$$

The bounds of the above example is drawn in the left plot of Figure A.1. In this case, $\text{Cov}(X_1, X_2) = 0.5$. On the other hand, if $E(X_1X_2) = 0$, the covariance between $X_1$ and $X_2$ equals -1. We get the right plot in Figure A.1, which confirms our previous prediction that when the two variables are negatively corrected, upper bounds on the joint right-tail events does not make sense.
Appendix B

Bounds on $\Pr(w_1 X_1 + w_2 X_2 \geq a)$

Now consider the complementary problem of finding sharp upper and lower bounds on $\Pr(w_1 X_1 + w_2 X_2 \geq a)$. This problem can be (respectively) formulated as the following optimization problems, obtained by setting in problem (3.1) (Section 3.2) $\phi(X_1, X_2) = \mathbb{I}_{\{w_1 X_1 + w_2 X_2 \geq a\}}$, and $\mathcal{D} = \mathbb{R}^2$. The upper bound is

$$\bar{p}_{\text{VaR}} = \max \mathbb{E}_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \geq a\}})$$

subject to

$$\mathbb{E}_F(X_i) = \mu_i, \quad i = 1, 2,$$

$$\mathbb{E}_F(X_i^2) = \mu_i^{(2)}, \quad i = 1, 2,$$

$$\mathbb{E}_F(X_1 X_2) = \mu_{12},$$

$F(x_1, x_2)$ a probability distribution on $\mathbb{R}^2$.

The lower bound has the objective function:

$$\underline{p}_{\text{VaR}} = \min \mathbb{E}_F(\mathbb{I}_{\{w_1 X_1 + w_2 X_2 \geq a\}}),$$

with the same constraints as (B.1).

The necessary and sufficient feasibility condition in terms of the moment parameters is the same as that of its complementary problem of bounds on $\Pr(w_1 X_1 + w_2 X_2 \leq a)$ (see page 59). Next we derive SOS programs to numerically compute $\bar{p}_{\text{VaR}}$, and $\underline{p}_{\text{VaR}}$ by using SOS programming solvers.
APPENDIX B. BOUNDS ON $\Pr(w_1X_1 + w_2X_2 \geq a)$

Upper bound

We begin by stating the dual problem of (B.1):

$$\overline{d}_{\text{VaR}} = \min \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{20}\mu_1^{(2)} + y_{02}\mu_2^{(2)} + y_{11}\mu_1$$

subject to $p(x_1, x_2) \geq \mathbb{I}_{\{w_1x_1 + w_2x_2 \geq a\}}$, for all $x_1, x_2 \in \mathbb{R}$.

The constraint in (B.3) is equivalent to

$$p(x_1, x_2) \geq 1, \text{ for all } x_1, x_2 \text{ with } x_1 + x_2 \geq a$$

$$p(x_1, x_2) \geq 0, \text{ for all } x_1, x_2 \in \mathbb{R}.$$  

We can directly express the second constraint above as a SOS constraint by using Hilbert’s Theorem (Theorem 1). For the first constraint however, we need more work. Using the same coordinate transformation as in Section 3.3.2, i.e., substituting $x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha$ for $x_1$ and $x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha$ for $x_2$ to the first constraint above, we have

$$p[x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha, x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha] \geq 1, \text{ for all } x'_1 \geq 0, x'_2 \in \mathbb{R}.$$  

This is equivalent to two constraints:

$$p[x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha, x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha] \geq 1, \text{ for all } x'_1 \geq 0, x'_2 \geq 0$$

$$p[x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha, x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha] \geq 1, \text{ for all } x'_1 \geq 0, x'_2 \leq 0. \quad (B.4)$$

By substituting $x'_2 \rightarrow -x'_2$ in the last inequality, and applying Diananda’s Theorem (Theorem ??), it follows that (B.3) is equivalent to the following SOS program:

$$\overline{d}_{\text{VaR}} = \min \quad y_{00} + y_{10}\mu_1 + y_{01}\mu_2 + y_{20}\mu_1^{(2)} + y_{02}\mu_2^{(2)} + y_{11}\mu_1$$

subject to the following are SOS polynomials:

$$p[x'_1 \cos \alpha - x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha, x'_1 \sin \alpha + x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha] - 1 \quad (B.5)$$

$$p[x'_1 \cos \alpha + x'_2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha, x'_1 \sin \alpha - x'_2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha] - 1$$

$$p(x'_1, x'_2)$$

Again, we have dropped the primes in the variable labels since they are just variable labels. The SOS program (B.5) can be readily solved with a SOS programming solver. Thus, if problem (B.1) is feasible, then we can numerically obtain the semiparametric upper bound $\underline{p}_{\text{VaR}}$ on
APPENDIX B. BOUNDS ON $\Pr(w_1X_1 + w_2X_2 \geq a)$

VaR probability, by solving problem (B.5) with a SOS solver.

**Lower bound**

This is the dual of the lower bound problem (B.2):

$$d_{\text{VaR}} = \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}$$

subject to $p(x_1, x_2) \leq \mathbb{1}_{\{w_1x_1+w_2x_2 \geq a\}}$, for all $x_1, x_2 \in \mathbb{R}$.

Again, strong duality between (B.2) and (B.6) holds if (B.2) is feasible. Following analogous steps, we find that problem (B.6) is equivalent to the SOS programs, which can be solved with a SOS program solver.

$$d_{\text{VaR}} = \max \quad y_{00} + y_{10} \mu_1 + y_{01} \mu_2 + y_{20} \mu_1^{(2)} + y_{02} \mu_2^{(2)} + y_{11} \mu_{12}$$

while the following are SOS polynomials:

$$1 - p[x_1^2 \cos \alpha - x_2^2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha, \ x_1^2 \sin \alpha + x_2^2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha] \quad (B.7)$$

$$1 - p[x_1^2 \cos \alpha + x_2^2 \sin \alpha + \frac{a}{w_1} \sin^2 \alpha, \ x_1^2 \sin \alpha - x_2^2 \cos \alpha + \frac{a}{w_1} \sin \alpha \cos \alpha]$$

$$-p(x_1^2, x_2^2).$$

The SOS program (B.7) can be readily solved with a SOS programming solver. Thus, if problem (B.2) is feasible, it follows that we can numerically obtain the semiparametric lower bound $\underline{d}_{\text{VaR}}$ on VaR probability by solving problem (B.7) with a SOS solver.
Appendix C

Obtain bounds on stop-loss payments from a transformed problem

Let \( \psi(Z) = Z - \phi(Z) \) where \( \phi(Z) \) is the stop-loss payment function defined in problem (3.43) and \( \psi(Z) \) is the transform function (3.54).

Under the same moments constraints, if the moment matrix \( \Sigma \) satisfies the feasibility requirement (see page 68), the lower bound of \( \phi(Z) \), \( \underline{p}(\phi) = \min \{ E_F[\phi(Z)] : F \in \pi(\mu) \} \) can be obtained from the upper bound of \( \psi(Z) \), \( \overline{p}(\phi) = \overline{p}(\psi) \).

**Proof.** On the one side:

\[
\underline{p}(\phi) = \min \{ E_F[\phi(Z)] : F \in \pi(\mu) \}
\]
\[
\underline{p}(\phi) \leq E_F[\phi(Z)] = \mu_z - E_F[\psi(Z)], \text{ for all } F \in \pi(\mu)
\]
Therefore, \( \underline{p}(\phi) \leq \mu_z - \overline{p}(\psi) \). \( \tag{C.1} \)

On the other side:

\[
\overline{p}(\psi) = \max \{ E_F[\psi(Z)] : F \in \pi(\mu) \}
\]
\[
\overline{p}(\psi) \geq E_F[\psi(Z)] = \mu_z - E_F[\phi(Z)], \text{ for all } F \in \pi(\mu)
\]
Therefore, \( \overline{p}(\psi) \geq \mu_z - \underline{p}(\phi) \)
So, \( \overline{p}(\phi) \geq \mu_z - \overline{p}(\psi) \).

In order to satisfy (C.1) and (C.2) simultaneously, \( \underline{p}(\phi) \) must equal \( \mu_z - \overline{p}(\psi) \). \( \square \)

Similarly, we can prove that the upper bound of \( \phi(Z) \), \( \overline{p}(\phi) = \max \{ E_F[\phi(Z)] : F \in \pi(\mu) \} \), equals \( \mu_z \) minus the lower bound of \( \psi(Z) \). That is, \( \overline{p}(\phi) = \mu_z - \underline{p}(\psi) \).
Appendix D

Proof of CVaR Expression Transformation: Equation (4.9)

Denote $F(x, \alpha, \beta) = \alpha - \frac{1}{\beta} \mathbb{E}[(\alpha - R(x))^+]$. If we fix $x$, for $\lambda \in (0, 1)$,

\[
\mathbb{E}[(\lambda \alpha_1 + (1 - \lambda) \alpha_2 - R(x))^+] = \\
\mathbb{E}[(\lambda (\alpha_1 - R(x)) + (1 - \lambda) (\alpha_2 - R(x))^+)] \leq \\
\mathbb{E}[\lambda (\alpha_1 - R(x))^+ + (1 - \lambda) (\alpha_2 - R(x))^+] = \\
\lambda \mathbb{E}[(\alpha_1 - R(x))^+] + (1 - \lambda) \mathbb{E}[(\alpha_2 - R(x))^+].
\]

So $\mathbb{E}[(\alpha - R(x))^+]$ is convex on $\alpha$. The inequality above follows

$$\max\{a + b, 0\} \leq \max\{a, 0\} + \max\{b, 0\}.$$ 

Since $-\frac{1}{\beta} \leq 0$ and the first term in $F(x, \alpha, \beta)$ is linear, the function $F(x, \alpha, \beta)$ is concave. Thus the maximum of

$$\max_{\alpha} F(x, \alpha, \beta) = \max_{\alpha} \alpha - \frac{1}{\beta} \mathbb{E}[(\alpha - R(x))^+]$$

can be found by differentiating $F(x, \alpha, \beta)$ with respect to $\alpha$ and then setting differentiated function equal to zero.

$$\frac{\delta}{\delta \alpha} F(x, \alpha, \beta) = 1 - \frac{1}{\beta} \mathbb{E}[\mathbb{I}(R(x) \leq \alpha)] = 1 - \frac{1}{\beta} \text{Pr}(R(x) \leq \alpha).$$

So the maximizer $\alpha^*$ satisfies

$$1 - \frac{1}{\beta} \text{Pr}(R(x) \leq \alpha^*) = 0,$$

123
APPENDIX D. PROOF OF CVAR EXPRESSION TRANSFORMATION: EQUATION (4.9)

or

\[ \Pr(R(x) \leq \alpha^*) = \beta. \]

That is, \( \alpha^* \) is the \( \beta \)-level VaR or \( \alpha^* = \alpha(x, \beta) \). So

\[ \max_{\alpha} \alpha - \frac{1}{\beta} \mathbb{E}[(\alpha - R(x))^+] = \alpha(x, \beta) - \frac{1}{\beta} \mathbb{E}[(\alpha(x, \beta) - R(x))^+]. \]

To finish, we notice

\[ \mathbb{E}[(\alpha(x, \beta) - R(x))^+] = \mathbb{E}[(\alpha(x, \beta) - R(x))^+ | R(x) \geq \alpha(x, \beta)] \Pr(R(x) \geq \alpha(x, \beta)) \]
\[ + \mathbb{E}[(\alpha(x, \beta) - R(x))^+ | R(x) \leq \alpha(x, \beta)] \Pr(R(x) \leq \alpha(x, \beta)). \]

The first term on the right hand side of the above equation is zero and the second term becomes

\[ \mathbb{E}[(\alpha(x, \beta) - R(x))^+ | R(x) \leq \alpha(x, \beta)] \Pr(R(x) \leq \alpha(x, \beta)) \]
\[ = \mathbb{E}[(\alpha(x, \beta) - R(x)) | R(x) \leq \alpha(x, \beta)] \beta \]
\[ = \beta \alpha(x, \beta) - \beta \mathbb{E}[R(x) | R(x) \leq \alpha(x, \beta)] \]
\[ = \beta \alpha(x, \beta) - \beta \text{CVaR}(x, \beta). \]

Replacing it back, we get

\[ \max_{\alpha} \alpha - \frac{1}{\beta} \mathbb{E}[(\alpha - R(x))^+] = \alpha(x, \beta) - \frac{1}{\beta} (\beta \alpha(x, \beta) - \beta \text{CVaR}(x, \beta)) = \text{CVaR}(x, \beta). \]
## Appendix E

### Assets and Lines of Business

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSY: 1-3</td>
<td>Short-term US Treasury bills</td>
</tr>
<tr>
<td>TSY: 7-10</td>
<td>Long-term US Treasury bonds</td>
</tr>
<tr>
<td>MBS</td>
<td>Mortgage-backed securities</td>
</tr>
<tr>
<td>Crude</td>
<td>Crude oil futures</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>The S&amp;P 500 Index</td>
</tr>
<tr>
<td>Agcy 1-3</td>
<td>Short-term agency bonds</td>
</tr>
<tr>
<td>Agcy 7-10</td>
<td>Long-term agency bonds</td>
</tr>
<tr>
<td>Corp AAA 3-5</td>
<td>Mid-term AAA corporate bonds</td>
</tr>
<tr>
<td>Corp AA 3-5</td>
<td>Mid-term AA corporate bonds</td>
</tr>
<tr>
<td>Corp AAA 3-5</td>
<td>Mid-term A corporate bonds</td>
</tr>
<tr>
<td>Corp BBB 3-5</td>
<td>Mid-term BBB corporate bonds</td>
</tr>
<tr>
<td>Corp HYld</td>
<td>Corporate high yield bonds</td>
</tr>
<tr>
<td>Sovrgn: Int</td>
<td>International sovereign bonds</td>
</tr>
<tr>
<td>Yankee</td>
<td>Yankee bonds</td>
</tr>
<tr>
<td>ABS</td>
<td>Asset-backed securities</td>
</tr>
<tr>
<td>Muni 3-5 Yrs</td>
<td>Mid-term municipal bonds</td>
</tr>
<tr>
<td>Commodities</td>
<td>Commodity futures</td>
</tr>
<tr>
<td>Lumber</td>
<td>Lumber futures</td>
</tr>
<tr>
<td>Currency</td>
<td>Currency futures</td>
</tr>
<tr>
<td>ML Convert</td>
<td>Convertible bonds</td>
</tr>
</tbody>
</table>
# Lines of Business

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cml Prop</td>
<td>Commercial Property</td>
</tr>
<tr>
<td>Allied</td>
<td>Allied Lines</td>
</tr>
<tr>
<td>Hm/Fr</td>
<td>Farmowners/Farmers Multiple Peril</td>
</tr>
<tr>
<td>CMP</td>
<td>Commercial Multiple Peril</td>
</tr>
<tr>
<td>Comp</td>
<td>Workers Compensation</td>
</tr>
<tr>
<td>GL</td>
<td>General Liability</td>
</tr>
<tr>
<td>Med/Prof</td>
<td>Medical Professional Liability</td>
</tr>
<tr>
<td>PPAuto</td>
<td>Private Passenger Auto Liability</td>
</tr>
<tr>
<td>Cauto</td>
<td>Commercial Auto/Truck Liability</td>
</tr>
<tr>
<td>FSB</td>
<td>Fidelity/Surety</td>
</tr>
<tr>
<td>BC/BS</td>
<td>Blue Cross Blue Shield</td>
</tr>
<tr>
<td>PCHth</td>
<td>Public and Commercial Health Insurance</td>
</tr>
<tr>
<td>Reins</td>
<td>Reinsurance</td>
</tr>
<tr>
<td>Other</td>
<td>Other Insurance</td>
</tr>
</tbody>
</table>
Bibliography


