Persistency and Stein's Identity: Applications in Stochastic Discrete Optimization Problems

A Dissertation Presented

By

Zheng Zhichao

In Partial Fulfilment of the Requirements

For the Degree of

Doctor of Philosophy

In Management

Department of Decision Sciences

NUS Business School

National Universiry of Singapore

June, 2013

Acknowledgments

First and foremost, I would like to express my deepest gratitude towards my advisor, Professor Teo Chung-Piaw. His constant support, motivation, and guidance is the only reason that I can survive the P.D. program and complete this thesis. It is through him that I see the passion, responsibility, wisdom, humbleness, and above all, the integrity as a scholar and a person. He would take every opportunity to share with students his broad interests and deep insights in both research and life, which has greatly shaped who I am today. To me, he has been much more than a mentor. In times of need or trouble, he has always been there ready to offer any help. As a heartwarming episode in my P.D. life, it was a great privilege for me and my family to have him as a lawful witness of my marriage under the Registry of Marriage in Singapore.

I am also immensely indebted to Professor Karthik Natarajan, who has been a reliable source of support through various stages of my life. Karthik is the advisor for my undergraduate honours thesis. Even before that, I learnt to appreciate the beauty of operations research from his excellent courses. It was him, who led me, hand-in-hand, to the world of academic research. Over the years, he has kept providing new ideas and guidance to push the boundary of my research. He never turned away from me in case I needed any help. His passion for innovative research and dedication to students have greatly inspired me.

I am very grateful to my thesis committee members, Professor Sun Jie and Professor Toh Kim Chuan, for their invaluable advice on improving my thesis. I am particularly grateful for the help and guidance given by Professor Toh Kim Chuan on solving tough conic optimization problems encountered during my research.

I would like to express my very great appreciation to my coauthors, Professor Kong Qingxia, Professor Lee Chung-Yee, and Ms Xu Yunchao, for their contributions to my research and beyond. Qingxia guided me in the early stage of my research though various project collaborations. I benefited tremendously from her passion and compassion in life and work. Professor Lee Chung-Yee shared with me his lifelong experience as a successful researcher and respected teacher. Yunchao inspired me with her passion, initiative, and determination in pursuing academic life.

I am particularly grateful to the wonderful faculty members in our department. I would like to thank Professor Melvyn Sim for his encouragement when I just embarked on my P.D. journey and the consistent support throughout it. I am deeply indebted to Professor Mable Chou and Professor Jussi Keppo for their support in my research as well as my job hunting process. I am very grateful for many insightful and inspiring discussions with Professor Zhang Hanqin. I would also like to thank Professor Keith Carter and Professor Christopher Chia for enlightening me with their excellent communication and strategic thinking skills.

During my time in NUS Business School, I am very fortunate to have the opportunities to experience various teaching duties and learn from many excellent educators, including Professor Quek Ser Aik, Dr. Liu Qizhang, Dr. Qi Mei, Professor Hum Sin Hoon, and Professor Christopher Chia (in chronological order). I am grateful to their generous support and guidance in this early stage of my teaching journey.

I would also like to extend my appreciation to the staff in our P.D. office and

Decision Sciences department, Ms Lim Cheow Loo, Ms Hamidah Bte Rabu, Ms Lee Chwee Ming, Ms Dorothy Tan, and Ms Teng Siew Geok, for their commitment and support.

My time in our department would not have been so colourful without the group of wonderful friends, Huang Junfei, Long Zhuoyu, Qi Jin, Rohit Nishant, Vinit Mishra, Xiao Li, Yuan Xuchuan, Zhang Meilin, Zhong Yuanguang, etc. Visits by our seniors, like Shu Jia and Zheng Huan, have brought refreshing thoughts and joy to the group from time to time.

I wish to thank my friends who are also enjoying their P.D. lives in different fields, Liu Zhengning, Wang Ben, Xiao Hui, just to name a few. The sharing of research ideas and progresses among us helped me keep an open mind and learn to appreciate the subtleness in different areas of research. I am particularly grateful to Peter Dickinson, who has carefully read through my research papers and pointed out critical issues that I overlooked. I have learned a lot from his eye-opening examples and rigorous altitude towards research. I am also very grateful to Han Zhijin, who have been a true friend of mine and was always there to give me a hand whenever I needed it. I am also grateful to all my friends for their support and cordial friendship.

My project collaboration with EADS Innovation Works Singapore was an important component of my P.D. study. I am thankful to my supervisor in EADS, Ms Elaine Wong, for her patient guidance and strong support. I am also thankful to my friends and colleagues in EADS who made my stay there pleasant and productive. Special thanks go to my team members, Vinh Nguyen and Yann Rebourg, for all the helpful discussions and constructive feedback.

It is impossible to find words that describe my gratitude and love to my wife, Ye Lingzhu. Her love and faith in me have been my greatest motivation to advance in

academics and pursue this endeavor. My life will not be so complete and meaningful without her and our little baby, Yulong.

Finally, to my parents, for their unconditional love and support, as always.

May 2013

Singapore

Zheng Zhichao

Persistency and Stein's Identity:

Applications in Stochastic Discrete Optimization Problems

Abstract. This thesis is motivated by the connection between stochastic discrete optimization and classical probability theory. In a general stochastic discrete optimization problem, Bertsimas et al. (2006) defined the notion of persistency, which is a generalization of many well-known concepts in different fields, such as criticality index in a project management problem and choice probability in a discrete choice problem. On the other hand, there is a classical covariance identity in probability theory, namely Stein's Identity, which describes the covariance between a function of a vector of random variables and each individual random variable. If we view the stochastic optimization as a function over the uncertain parameters in the problem, persistency will appears as a critical component in the identity.

We exploit such connection to solve two classes of problems. The first is approximating the distribution of the optimal value of a mixed zero-one linear optimization problem under objective uncertainty. A typical example is to approximate the distribution of the completion time of a project when its individual activity completion times are stochastic. We propose a least squares approximation framework for the problem. By linking the framework to Stein's Identity, we show that the least squares normal approximation of the random optimal value can be computed by solving the corresponding persistency problem. We further extend our method to construct a quadratic least squares estimator to improve the accuracy of the approximation, in particular, to capture the skewness of the objective value. Computational studies show

that the new approach provides much more accurate estimates compared to existing methods, especially in predicting the variability of the project completion time.

The second problem is related to decision making under uncertainty. We propose a new decision criterion for stochastic discrete optimization problem under objective uncertainty, named quadratic regret. The proposed quadratic regret solution is selected by minimizing the expected squared deviation of its performance from the best alternative. We illustrate this decision criterion using the example of portfolio management problem, where it is equivalent to tracking-error minimization. We develop a new portfolio strategy that tracks the highest return from a set of benchmark portfolios. By resorting to Stein's Identity, we present a closed-form expression for the optimal portfolio position and relate them to the persistency. The connection between persistency and a common behavioural abnormality, probability matching, provides several interesting insights to the investment behaviour, which partially justifies our modeling framework. With the closed-form solution, we prove that our model has the flexibility to generate the entire mean-variance efficient frontier if the benchmark portfolios are two distinct mean-variance portfolios, a result similar to the Two-Fund Theorem. We also show that the linear combination rule would be inferior to our portfolio if the portfolio manager has a mean-variance utility with low risk aversion, which provides further motivation to our approach. In comparison to the single-benchmark tracking-error minimization approach, we show that the new model helps mitigate the agency issues due to the use of single benchmark, and provide several insights on benchmark selection for our multiple-benchmark model. We perform comprehensive numerical experiments with various empirical data sets to demonstrate that our approach can consistently provide higher net Sharpe ratio (after accounting for transaction cost), higher net aggregate return, and lower turnover rate, compared to ten different bench-

mark portfolios proposed in the literature, including the equally weighted portfolio.

Note that rather than solving the above two problems directly, we transform them into the problem of estimating persistency values by connecting them to Stein's Identity. This approach allows us to conduct many in-dept analysis of the problems as demonstrated above. Moreover, we can explore the existing results in persistency estimation literature to help tackle the original problems. In the last part of this thesis, besides commenting on potential future research, we also discuss an approach to refine the persistency estimation under normality assumption. Although most results in the thesis are derived under the normality assumption on the uncertainty due to the usage of Stein's Identity, there are several extensions of Stein's Identity to different distributions such that our results can be carried over to other situations.

Thesis Advisor. Professor Teo Chung-Piaw, Department of Decision Sciences, NUS Business School, National University of Singapore

Research Overview

This thesis originates from the author's summer paper for the P.D. qualifying examinations. The first version of the paper focused on the linear least squares model for distribution approximation and treated the portfolio management as an application of the theory. As suggested by some anonymous referees, the two parts contains disparate findings and there is a lack of unifying framework due to the different natures of the two problems, and it is better to separate them and involve more analytical depth for each part. Following the recommendations, we removed the portfolio management problem, and added more analysis on the distribution approximation problem, including the quadratic estimator, extension to skewed-normal distribution, as well as two more applications in maximum partial sum problem and statistical timing analysis. The part on portfolio management problem was repositioned to focus on tracking-error model for multiple benchmarks, and much more analysis has been included to make it a piece of research paper on its own. These two research papers form the two main chapters of this thesis. I would like to thank my coauthors, Karthik Natarajan and Yunchao Xu, for their contributions to these papers.

Besides the work presented in this thesis, I am also involved in another line of research on optimization under uncertainty, which focuses on conic reformulation of the distributionally robust optimization problem with applications in healthcare appointment scheduling and sequencing as well as liner shipping service planning. Indeed,

the problems addressed in this thesis are closely related to those works. The main persistency estimation model used in the distribution approximation problem in this thesis comes from those research efforts.

May 2013

Singapore

Zheng Zhichao

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Consider the following general mixed zero-one linear programming (LP) problem:

$$Z(\tilde{\boldsymbol{c}}) := \max_{\boldsymbol{x}\in\mathcal{P}} \sum_{j=1}^{n} \tilde{c}_j x_j, \tag{1.0.1}$$

where $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_n)^T$ is the random coefficient vector, and $\mathcal{P}$ is the domain of the feasible solutions defined by

$$\mathcal{P} := \{\boldsymbol{x} \in \mathbb{R}^n \,:\, \boldsymbol{a}_i^T \boldsymbol{x} = b_i,\, \forall i = 1, \ldots, m;\; x_j \in \{0,1\},\, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\};\; \boldsymbol{x} \geq \boldsymbol{0}\}.$$

Note that part of decision variables is bounded to be either 0 or 1, which is indexed by the set $\mathcal{B}$. We assume that $\mathcal{P}$ is nonempty and bounded so that $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ is finite. It is well-known that the general mixed zero-one LP problem is classified as $\mathcal{NP}$-hard. Nevertheless, it is one of the most useful tools to model the real world problems, ranging from engineering systems to business applications, for example, telecommunication networks, transportation systems, and production planning and scheduling, etc. Unfortunately, most of the input parameters to the model would contain errors

and/or noises either from estimation or prediction, and the most common approach to describe such uncertainty is probability distribution. In this thesis, we focus on the uncertainty inside the objective coefficient vector that follows a certain multivariate distribution.

In the rest of this chapter, we first discuss the concept of persistency in the context of our problem. Next, we review Stein's Identity, and point out its connection to persistency. Exploiting such connection between persistency and Stein's Identity, we solve two classes of problems in Chapter 2 and 3 by transforming them into persistency problems[1]. In Chapter 4, besides some concluding remarks, we also discuss how to solve the persistency problem better and consequently obtain better solution to the original problems.

1.1 Persistency

Bertsimas et al. (2006) introduced the notion of the persistency of a binary decision variable in Problem (1.0.1) as the probability that the variable is active (i.e., takes value of 1) in an optimal solution to Problem (1.0.1). We generalize this concept to include continuous variables as follows:

Definition 1.1. The persistency of the decision variable x_j in Problem (1.0.1)) is defined as $\mathbf{E}[x_j(\tilde{\boldsymbol{c}})]$, where $x_j(\tilde{\boldsymbol{c}})$ denotes an optimal value of x_j as a function of the random vector $\tilde{\boldsymbol{c}}$. If x_j is a binary variable, then $\mathbf{E}[x_j(\tilde{\boldsymbol{c}})] = \mathbf{P}(x_j(\tilde{\boldsymbol{c}}) = 1)$.

Remark 1.2. When $\tilde{\boldsymbol{c}}$ is continuous and spans the whole space of $\mathbb{R}^n$, the support of $\tilde{\boldsymbol{c}}$ over which Problem (1.0.1) has multiple optimal solutions has measure zero and $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is unique almost surely[2]. In other situations, if there exist multiple optimal solutions

[1]The definitions of persistency and persistency problem are provided in the next section.

[2]Note that the feasible region of Problem (1.0.1) is a bounded polytope, so it has multiple optimal

over a support of strictly positive measure, $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is defined to be an optimal solution randomly selected from the set of optimal solutions at $\tilde{\boldsymbol{c}}$.

The notion of persistency generalizes several popular concepts in different application domains, e.g., "criticality index" in project networks and "choice probability" in discrete choice models (cf. Bertsimas et al. (2006), Natarajan et al. (2009), Mishra et al. (2012)). In the rest of the thesis, by persistency problem, we mainly mean the problem of estimating the persistency values. Sometimes to avoid excessive exposition, we also include problem of estimating other stochastic parameters of the problem $Z(\tilde{\boldsymbol{c}})$ under the umbrella of persistency problem, which will be clear in the respective contexts.

Note that there is a very similar but different concept in literature, which is commonly referred to as persistence. Brown et al. (1997) brought up the issue of persistence and persistent modeling in optimization through a series of case studies. Although the idea of persistence conveyed in their paper is very broad and different from the persistency defined above, these two concepts are closely related through the issue of data uncertainty and robust optimization. The authors pointed out that from the perspective of persistence, robust optimization seeks a baseline solution that will persist as best as possible with a number of alternate forecast revisions. On the other hand, persistency describes the degree of persistence of each individual decision variable in an optimization problem with data uncertainty. Indeed, we can further generalize Definition 1.1 to the persistency of a feasible solution, i.e., the probability that this particular feasible solution is optimal. However, it is beyond the scope of this thesis,

solutions only when $\tilde{\boldsymbol{c}}$ realized to be a normal vector of a surface of the polytope. Since the number of facets is finite for a give polytope, the probability measure over all the normal vectors is zero. For example, consider a polytope in $\mathbb{R}^2$, for any polytope, its normal vectors are just lines in $\mathbb{R}^2$. If $\tilde{\boldsymbol{c}}$ is continuous and spans the whole space of $\mathbb{R}^2$, the probability measure over all these lines is zero, since the number of these lines is finite.

and we will not elaborate further.

In our problem setting, persistency describes an important characteristic of a stochastic optimization system, i.e., the impact of each individual random variable on the final outcome of the optimization process. Knowing the persistency values not only helps analyze the stochastic optimization systems, but also sheds some light on human being's decision making behaviour when interacting with such systems. As documented in extensive literature of decision making under uncertainty, human beings exhibit various predictably irrational decision patterns that deviate from those assumed by the conventional expected utility theory, which is commonly regarded as rational behaviour. One of such behavioural abnormality is called probability matching, and it is closely related to the concept of persistency we have described.

Probability matching refers to the suboptimal choice behaviour involving probabilistic outcomes in repeated events. By suboptimality, we mean that the choice decisions are consistently different from the strategy that maximizes the expected utility. A representative experiment involves one subject who is asked to repeatedly predict the outcome of two randomly flashing light bulbs. One of the light bulbs is red and the other is green. In each round, only one of them will flash, and the subject is asked to predict the colour of the flashing light bulb. The experiment is set up such that in every round, the red light will flash with probability 70% and the green one will flash with probability 30%. The subject is incentivized to maximize the number of corrected predictions when the game is repeated for a large number of times. Under such settings, it is obvious that the optimal strategy is to always predict the outcome of the more probable event, in this case, the red light. However, the empirical evidence suggests that people almost never choose the more probable outcome exclusively. More specifically, people tend to match the relative frequencies of their predictions to the

relative frequencies that the light bulbs flash. On average, people predict the red light approximately 70% of the time. Probability matching has been observed in various experiments under different settings. Although the exact behaviour of the subjects depends on many parameters, e.g., the amount of incentives, the length of experiments, etc., the pattern of probability matching appears to be quite robust. Similar experiments were carried out on various animals, and many interesting observations have been collected since 1950s. Typically, a rat or a monkey maximizes, i.e., they tend to choose the more frequently rewarded stimulus on almost all trials (cf. Hickson (1961), Wilson et al. (1964)). In the experiments with fish under the conditions in which the rat maximizes, by contrast, random probability matching appears to be the dominant behaviour (cf. Bullock & Bitterman (1961). The results with intermediate forms, e.g., pigeon, show mixing behaviours (cf. Bullock & Bitterman (1962)). Vulkan (2000) summarized and tabulated most experimental results related to probability matching on both human and animal subjects and provided a good review on the related literature.

It is worthwhile to point out that probability matching is not only observed in simple laboratory experiments, but also has many profound implications in real life decision making processes, e.g., medical diagnosis (cf. Friedman et al. (1995)), and law enforcement (cf. Guttel & Harel (2005)), etc. Moving beyond the binary choice, researchers have consistently observed that people tend to adopt mix strategies in more complex stochastic environment if they face the same problem repeatedly. Take the newsvendor problem as an example: when the future demand of the newspaper is uncertain, a newsvendor needs to decide how many copies of newspaper to order every morning before knowing the actual demand that will be realized only after the day ends. If the objective is to maximize the profit in the long run or the daily

expected profit, there is a well-known formula for the optimal order quantity based on a critical ratio and the demand distribution. Then the best strategy is to order this optimal quantity every day. However, this is not how human beings are going to behave, even in a laboratory setting under which all the environmental parameters fit the theoretical assumptions exactly and the optimal newsvendor order quantity is known to the subjects. It was observed in many experiments that the average order quantities from a pool of subjects in a series of repeated newsvendor games tend to fluctuate around a certain order level and form some distributions (cf. Schweitzer & Cachon (2000), Moritz et al. (2013), etc.). Though the reasons behind probability matching behaviour are still under intense debate, one of the most commonly accepted explanation is that human beings try to achieve the best possible outcome and believe that there is a way to perfectly predict the future. Besides understanding the origin of the behaviour, it is also important to incorporate such behaviour in any models that involve human decision making under uncertainty. There have been various attempts to model probability matching, but most of the models so far remain relatively quite preliminary and it was admitted that some ideas are extremely hard to formalize (cf. Vulkan (2000)).

In most settings when probability matching occurs, persistency values are exactly the underlying probabilities matched by the subjects. In the example of predicting flashing light bulbs, for any perfect prediction sequence, the proportion of prediction of red light is equal to the proportion of times that the red light flashes, which is around 70%. This is exactly the definition for persistency. In the context of the newsvendor problem, persistency is exactly the demand distribution because the best possible return comes from a perfect prediction of demand, and when demand is known, ordering the exact demand quantity maximizes the profit. Linking the theory of persistency

to the empirical phenomenon of probability matching may provide a better way to understand and model such probabilistic behaviour. Our first attempt is to analyze a portfolio selection problem under the uncertainty from asset returns. We propose a new decision criterion for decision making under uncertainty, namely least squares regret, which is equivalent to the popular benchmark tracking criterion in portfolio management practice. From the closed-form solution, we show that the persistency forms a basic component of the final decision. Connecting to the behaviour of probability matching, we gain new insights on the reasons of the behaviour. On the other hand, this also gives us a new way to model the behaviour, which is worth further exploration. We leave detailed discussion to Chapter 3 and 4.

Having discussed the importance of persistency, next we briefly review the existing generic methods for estimating the persistency. Note that since persistency is generalized from several popular concepts in different areas, there are specific methods that take advantage of the special problem structures to estimate the persistency in each area. We will leave the review of these specific methods to the place where the application examples are discussed in the rest of this thesis.

The most intuitive generic approach would be the Monte Carlo simulation. However, since the general mixed zero-one linear optimization problems are $\mathcal{NP}$-hard, simulation may require tremendous effort or resources to achieve satisfactory results. Moreover, the sensitivity of the approach to the samples generated also calls for other efficient estimation method. Over the past few years, a substream of research in the field of persistency estimation has yielded a series of semidefinite programming (SDP) models based on the connection between the moment cone and the semidefinite cone. A common feature of these models is that they only assume the knowledge of moment information of the uncertainty rather than the exact form of the distribu-

tion. Hence, they are also referred as distributionally robust stochastic programming (DRSP) models.

Bertsimas et al. (2006) introduced arguably the first generic computational approach to approximate the persistency by solving a class of SDPs called Marginal Moment Model (MMM) under the assumption that the random vector $\tilde{\boldsymbol{c}}$ is described only through the marginal moments of each $\tilde{c}_j$ and all the decision Problem (1.0.1) are binary. Natarajan et al. (2009) extended MMM to general mixed-integer LP problems, but their model formulation is based on the characterization of the convex hull of the binary reformulation, which is typically difficult to derive. Lasserre (2010) studied the class of parametric polynomial optimization problems, which includes the mixed zero-one linear programming problem as a special case. The author described the uncertainty using a combination of joint probability measure on the parameters and optimal solutions together with marginal probability measures on the parameters. A hierarchy of semidefinite relaxations was proposed to solve the problem. However, the size of the semidefinite relaxation grows rapidly, which makes solving the higher order semidefinite relaxations numerically challenging. Mishra et al. (2012) presented a SDP model named Cross Moment Model (CMM) for $\tilde{\boldsymbol{c}}$ described by both the marginal and cross moments. The formulation of CMM is based on the extreme point enumeration of Problem (1.0.1). Hence, the size of CMM becomes exponential for general LP problems. Inspired by a recent application of conic optimization on mixed zero-one LP problems due to Burer (2009), Natarajan et al. (2011) developed a parsimonious but $\mathcal{NP}$-hard convex conic optimization model to estimate the persistency of a general mixed zero-one LP problem when $\tilde{\boldsymbol{c}}$ is described by both the marginal and cross moments as well as nonnegative support. In this thesis, we mainly exploit this model to estimate the persistency values. Therefore, we will review it in greater detail next.

Natarajan et al. (2011) consider the following stochastic optimization problem:

$$Z_P := \sup_{\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right],$$

where $\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+$ means that the set of distributions of the random coefficient vector $\tilde{\boldsymbol{c}}$ is defined by the nonnegative support $\mathbb{R}^n_+$, finite mean vector $\boldsymbol{\mu}$, and finite covariance matrix Σ, i.e.,

$$\tilde{\boldsymbol{c}} \in \{\tilde{\boldsymbol{X}} \,:\, \mathbf{E}[\tilde{\boldsymbol{X}}] = \boldsymbol{\mu}, \mathbf{E}[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T] = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T, \mathbf{P}(\tilde{\boldsymbol{X}} \geq \mathbf{0}) = 1\}.$$

Furthermore, this set is assumed to be nonempty. The distribution that attains the value of Z_P is generally referred to as the worst case distribution. Natarajan et al. (2011) proved that Z_P can be solved as the following convex conic optimization problem:

$$\begin{array}{rll} Z_C := & \max & \sum_{j=1}^n Y_{j,j} \\ & s.t. & \boldsymbol{a}_i^T X \boldsymbol{a}_i - 2b_i \boldsymbol{a}_i^T \boldsymbol{x} + b_i^2 = 0,\ \forall i = 1, \ldots, m \\ & & X_{j,j} = x_j,\ \forall j \in \mathcal{B} \\ & & \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{x}^T \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T & Y^T \\ \boldsymbol{x} & Y & X \end{pmatrix} \succeq_{cp} 0 \end{array}$$

i.e., $Z_P = Z_C$, where the decision variables are $\boldsymbol{x} \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times n}$, and $Y \in \mathbb{R}^{n \times n}$. For a matrix $A \in \mathbb{R}^{n \times n}$, $A \succeq_{cp} 0$ means that A lies in the cone of completely positive matrices of dimension n defined as

$$\mathcal{CP}_n := \left\{A \in \mathbb{R}^{n \times n} \,:\, \exists V \in \mathbb{R}^{n \times k}_+, \text{ such that } A = VV^T\right\}.$$

The linear program over the convex cone of the completely positive matrices is called

a completely positive program (CPP), and Z_C is a typical CPP. Since the model is a CPP, and it captures the cross moment information, the authors named their model as Completely Positive Cross Moment Model (CPCMM). Furthermore, they extended CPCMM by relaxing the nonnegative support assumption on $\tilde{\boldsymbol{c}}$. A key reason that CPCMM is chosen for persistency estimation is its ability to capture correlations among random coefficients with its compact formulation. We will illustrate more on this point when discussing the specific applications later.

In the formulation of Z_C, the variables $\boldsymbol{x}$, Y and X attempt to encode the information $x_j = \mathbf{E}[x_j(\tilde{\boldsymbol{c}})]$, $Y_{i,j} = \mathbf{E}[\tilde{c}_j x_i(\tilde{\boldsymbol{c}})]$ and $X_{i,j} = \mathbf{E}[x_i(\tilde{\boldsymbol{c}})x_j(\tilde{\boldsymbol{c}})]$ under the worst case distribution. Thus, through solving Z_C, the optimal value of $\boldsymbol{x}$ is simply the persistency under the worst case distribution, which provides an estimate of the persistency under other distribution with the same moments.

An issue with CPCMM is that it is $\mathcal{NP}$-hard to solve despite the fact that the completely positive cone is closed, convex and pointed. Fortunately, there are various hierarchies of tractable approximations for the completely positive cone, e.g., Bomze et al. (2000), Parrilo (2000), and Klerk et al. (2002) etc. For all the computational studies in this thesis, we solve a simple SDP approximation of the completely positive constraint, i.e., $A \succeq_{cp} 0$ is relaxed to $A \succeq 0$ and $A \geq 0$, where $A \succeq 0$ means that A is positive semidefinite. Such relaxation is called doubly nonnegative relaxation.

1.2 Stein's Identity

In this section, we will introduce Stein's Identity, and briefly discuss its link to the discrete stochastic optimization problem and persistency. Stein's Identity is a well-known theorem of probability theory that is of interest primarily because of its applications to statistical inference and portfolio choice theory. The formal statement is presented

next together with its proof for completeness.

Lemma 1.3. *[Stein's Identity] Let the random vector* $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_n)^T$ *be multivariate normally distributed with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* Σ*. For any function* $h(c_1, \ldots, c_n) : \mathbb{R}^n \to \mathbb{R}$ *such that* $\partial h(c_1, \ldots, c_n)/\partial c_j$ *exists almost everywhere and* $\mathbf{E}[|\partial h(\tilde{\boldsymbol{c}})/\partial c_j|] < \infty$*,* $\forall j = 1, \ldots, n$*, denote* $\nabla h(\tilde{\boldsymbol{c}}) = (\partial h(\tilde{\boldsymbol{c}})/\partial \tilde{c}_1, \ldots, \partial h(\tilde{\boldsymbol{c}})/\partial \tilde{c}_n)^T$*. Then*

$$Cov(\tilde{\boldsymbol{c}}, h(\tilde{\boldsymbol{c}})) = \Sigma \mathbf{E}[\nabla h(\tilde{\boldsymbol{c}})].$$

Specifically,

$$Cov\left(\tilde{c}_k, h(\tilde{c}_1, \ldots, \tilde{c}_n)\right) = \sum_{j=1}^{n} Cov\left(\tilde{c}_k, \tilde{c}_j\right) \mathbf{E}\left[\frac{\partial}{\partial \tilde{c}_j} h(\tilde{c}_1, \ldots, \tilde{c}_n)\right], \forall k = 1, \ldots, n.$$

Proof. The proof is consolidated from Stein (1972), Stein (1981) and Liu (1994).

The first result is the univariate version of Stein's Identity (cf. Stein (1972) and Stein (1981)).

Let $\tilde{c}$ follow a standard normal distribution, $N(0,1)$, and $\phi(c)$ denote the standard normal density with the derivative satisfying $\phi'(c) = -c\phi(c)$. For any function $h : \mathbb{R} \to \mathbb{R}$ such that h' exists almost everywhere and $\mathbf{E}[|h'(\tilde{c})|] < \infty$,

$$\begin{aligned}
\mathbf{E}\left[h'\left(\tilde{c}\right)\right] &= \int_{-\infty}^{\infty} h'(c)\phi\left(c\right) dc \\
&= \int_{0}^{\infty} h'(c) \left[\int_{c}^{\infty} z\phi\left(z\right) dz\right] dc + \int_{-\infty}^{0} h'(c) \left[\int_{-\infty}^{c} -z\phi\left(z\right) dz\right] dc \\
&= \int_{0}^{\infty} z\phi\left(z\right) \left[\int_{0}^{z} h'(c) dc\right] dz - \int_{-\infty}^{0} z\phi\left(z\right) \left[\int_{z}^{0} h'(c) dc\right] dz \\
&= \left(\int_{0}^{\infty} + \int_{-\infty}^{0}\right) \left[z\phi\left(z\right) \left[h(z) - h(0)\right]\right] dz \\
&= \int_{-\infty}^{\infty} z\phi\left(z\right) h(z) dz \\
&= \mathbf{E}\left[\tilde{c}h\left(\tilde{c}\right)\right],
\end{aligned}$$

where the third equality is justified by Fubini's Theorem. Note that since $\mathbf{E}[\tilde{c}] = 0$ and $Var(\tilde{c}) = 1$, the equality proved above is essentially

$$Cov\left(\tilde{c}, h\left(\tilde{c}\right)\right) = Var(\tilde{c})\mathbf{E}\left[h'\left(\tilde{c}\right)\right]. \tag{1.2.1}$$

Next, we present the generalization of the result to the multivariate case (cf. Stein (1981) and Liu (1994)).

Let $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \ldots, \tilde{z}_n)^T$, where $\tilde{z}_j$'s are independent and identically distributed standard normal random variables. From Equation (1.2.1), it follows that for any function $\hat{h} : \mathbb{R}^n \to \mathbb{R}$ satisfying the same conditions as h in the theorem,

$$\mathbf{E}\left[\tilde{z}_1\hat{h}\left(\tilde{\boldsymbol{z}}\right)\middle|\left(\tilde{z}_2, \ldots, \tilde{z}_n\right)\right] = \mathbf{E}\left[\frac{\partial\hat{h}\left(\tilde{\boldsymbol{z}}\right)}{\partial z_1}\middle|\left(\tilde{z}_2, \ldots, \tilde{z}_n\right)\right].$$

Taking the expectation of both sides, we get

$$\mathbf{E}\left[\tilde{z}_1\hat{h}\left(\tilde{\boldsymbol{z}}\right)\right] = \mathbf{E}\left[\frac{\partial\hat{h}\left(\tilde{\boldsymbol{z}}\right)}{\partial z_1}\right].$$

Using a similar argument for the remaining random variables, we can show that

$$Cov\left(\tilde{\boldsymbol{z}}, \hat{h}\left(\tilde{\boldsymbol{z}}\right)\right) = \mathbf{E}\left[\nabla\hat{h}\left(\tilde{\boldsymbol{z}}\right)\right].$$

Note that the random vector $\tilde{\boldsymbol{c}}$ can be written as $\tilde{\boldsymbol{c}} = \Sigma^{1/2}\tilde{\boldsymbol{z}} + \boldsymbol{\mu}$. Consider $\hat{h}\left(\tilde{\boldsymbol{z}}\right) = h\left(\Sigma^{1/2}\tilde{\boldsymbol{z}} + \boldsymbol{\mu}\right)$, then $\nabla\hat{h}\left(\tilde{\boldsymbol{z}}\right) = \Sigma^{1/2}\nabla h\left(\tilde{\boldsymbol{c}}\right)$. Hence,

$$Cov\left(\tilde{\boldsymbol{c}}, h\left(\tilde{\boldsymbol{c}}\right)\right) = Cov\left(\Sigma^{1/2}\tilde{\boldsymbol{z}}, \hat{h}\left(\tilde{\boldsymbol{z}}\right)\right) = \Sigma^{1/2}\mathbf{E}\left[\nabla\hat{h}\left(\tilde{\boldsymbol{z}}\right)\right] = \Sigma\mathbf{E}\left[\nabla h\left(\tilde{\boldsymbol{c}}\right)\right].$$

Therefore, the proof is completed. □

Briefly speaking, we can view the optimization problem in (1.0.1) as a mapping on the random parameters, i.e., $Z(\tilde{\boldsymbol{c}})$ is a function of $\tilde{\boldsymbol{c}}$. Then Stein's Identity can be used to characterize the covariance between $Z(\tilde{\boldsymbol{c}})$ and each $\tilde{c}_j$ under the normality assumption on $\tilde{\boldsymbol{c}}$. Under certain conditions, the gradient of $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ is simply the persistency, i.e., $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$. More details will be provided when we discuss the problems in the next two chapters.

Lemma 1.3 holds under the normal uncertainty assumption on the random variables. Interestingly, there are many extensions of Stein's Identity for other distributions (cf. Adcock (2007), Barbour et al. (1992), Liu (1994), etc.), which allows the results discussed in this thesis to be extended further. We will illustrate one such case in the next chapter, where we take advantage of the Stein's Identity under multivariate skew-normal distribution.

Chapter 2

Least Squares Distribution Approximation – Application in Project Management Problem

ABSTRACT. This chapter addresses the following question: What is the best approximation of the distribution of the completion time of a project network with random activity durations? More generally, we consider any mixed zero-one linear optimization problem under objective uncertainty, and develop an approach to approximate the distribution of its optimal value when the random objective coefficients follow a multivariate normal distribution. Linking our model to the classical Stein's Identity, we show that the least squares normal approximation of the random optimal value can be computed by solving the corresponding persistency problem. We further extend our method to construct a quadratic least squares estimator to improve the accuracy of the approximation, in particular, to capture the skewness of the objective value. Computational studies show that the new approach provides more accurate estimates of the first and second moments of project completion time compared to existing methods.

2.1 Problem Overview

One of the fundamental problems in project management is to identify the project completion time when the activity durations are random. It is well-known that a project can be represented as a directed acyclic graph (DAG). We adopt the conventional activity-on-arc representation of the project network, where arcs represent activities and nodes represent the milestones that indicate the starting or ending of the activities. The length of an arc is the duration of the activity represented by that arc. Hence, if all the activities have deterministic durations, finding the project completion time is as easy as finding the longest path in a corresponding DAG, which can be solved as a linear programming (LP) problem[1]. However, when the activity durations are stochastic, the analysis of the random project completion time becomes nontrivial.

It has long been the interest of both researchers and practitioners to estimate the distribution of the project completion time. Over the past few decades, various methods have been proposed to approximate this distribution (cf. Dodin (1985); Cox (1995), etc.). Unfortunately, to the best of our knowledge, these approaches are derived using ad hoc heuristics or work on specific problem instances. In this research, we partially address this issue under the assumption that the activity durations follow a multivariate normal distribution, and construct a normal distribution approximation for the random project completion time that is optimal under the L^2-norm. In fact, our method applies to any general random mixed zero-one LP problem under objective uncertainty:

$$Z\left(\tilde{\boldsymbol{c}}\right) := \max_{\boldsymbol{x}\in\mathcal{P}} \sum_{j=1}^{n} \tilde{c}_j x_j, \tag{2.1.1}$$

where $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_n)^T$ is the random coefficient vector following a multivariate normal

[1]In fact, for the deterministic case, this problem can be solved in a more efficient way, which is dynamic programming, in effort proportional to the number of arcs in the DAG.

distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, denoted as $\tilde{\boldsymbol{c}} \sim N(\boldsymbol{\mu}, \Sigma)$, and $\mathcal{P}$ is the domain of the feasible solutions (assumed to be bounded) defined by

$$\mathcal{P} := \{\boldsymbol{x} \in \mathbb{R}^n \,:\, \boldsymbol{a}_i^T \boldsymbol{x} = b_i,\, \forall i = 1, \ldots, m;\; x_j \in \{0,1\}\,,\, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\}\,;\; \boldsymbol{x} \geq \boldsymbol{0}\}.$$

In the project management problem, $\mathcal{P}$ characterizes the incidence vector of paths in the project network, and $\tilde{c}_j$ is the random duration of activity j. To give a precise linear programming formulation of the project management problem, we index the variables in two dimensions. Consider a DAG (V, E) with origin node s representing the starting of the project and destination node t representing the ending of the project. For each arc (i, j) representing an activity, let $\tilde{c}_{i,j}$ be the activity duration and $x_{i,j}$ be the flow variable. Then the project completion time can be found by solving the following linear programming problem:

$$\begin{aligned} \max \quad & \sum_{(i,j)\in E} \tilde{c}_{i,j} x_{i,j} \\ s.t. \quad & \sum_{j:(i,j)\in E} x_{i,j} - \sum_{j:(j,i)\in E} x_{j,i} = \begin{cases} 1, & \text{if } i = s \\ -1, & \text{if } i = t \\ 0, & \text{otherwise} \end{cases} \\ & x_{i,j} \geq 0,\, \forall\, (i,j) \in E \end{aligned}$$

Throughout this chapter, we use bold face letters to denote column vectors. We use $\sigma_{j,k}$, $j, k = 1, \ldots, n$, to denote the covariance between $\tilde{c}_j$ and $\tilde{c}_k$, i.e., $(j,k)-$term of the covariance matrix Σ. We also use σ_j^2, $j = 1, \ldots, n$, to denote the variance of $\tilde{c}_j$, i.e., the j^{th} diagonal entry of Σ. We assume that $\tilde{\boldsymbol{c}}$ is non-degenerate, i.e., the covariance matrix Σ is symmetric positive definite (denoted as $\Sigma \succ 0$). Let $x_j\,(\tilde{\boldsymbol{c}})$ denote an optimal value of x_j as a function of the random vector $\tilde{\boldsymbol{c}}$. Together with the

normality assumption, we are sure that $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is unique almost surely.

There is by now a huge literature on finding the distribution of $Z(\tilde{\boldsymbol{c}})$ for various combinatorial optimization problems, including minimum assignment, spanning tree, and traveling salesman problem (cf. Aldous & Steele (2003)). These problems are notoriously hard, and often only partial results (e.g., asymptotic results with i.i.d random variables) are known. Finding the exact distribution for the general mixed zero-one LP problem appears to be almost impossible.

Back to the project management problem, under the Critical Path Method (CPM), which is often used by the project management community, the random project completion time is estimated by replacing $\tilde{c}_j$ with its expected value μ_j, i.e., $Z(\boldsymbol{\mu})$ is used to approximate the project completion time. In the classical Program Evaluation and Review Technique (PERT), this is taken one step further where the distribution of the project completion time is approximated by $\sum_{j=1}^{n} \beta_j \tilde{c}_j = \sum_{j=1}^{n} \beta_j (\tilde{c}_j - \mu_j) + Z(\boldsymbol{\mu})$[2], with

$$\beta_j = \begin{cases} 1, \text{ if arc } j \text{ is on the longest path when solving Z}(\boldsymbol{\mu}), \\ 0, \text{ otherwise.} \end{cases}$$

Due to the simplicity of the approach, PERT has gained a lot of popularity, and the random project networks are sometimes also called PERT networks. However, simply using the distribution of one critical path to approximate the distribution of the project completion time (or longest path) suffers from severe estimation errors. In particular, PERT has been widely criticized for significant underestimation of the mean project completion time and overestimation of the variability of the project completion time.

[2]Here we impose the conventional assumption that there is a unique optimal solution when we compute $Z(\boldsymbol{\mu})$.

This leads us to a natural estimation problem:

$$\text{(P)} \qquad \min_{\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^n} \ \mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha - \sum_{j=1}^{n}\beta_j(\tilde{c}_j - \mu_j)\right)^2\right],$$

where $\alpha \in \mathbb{R}$, and $\boldsymbol{\beta} \in \mathbb{R}^n$ are determined by minimizing the expected squared deviation of the linear approximation from the true distribution. Problem (P) is the central question addressed by this chapter. We are solving for the least squares normal approximation (or the best normal approximation in L^2-norm) to the random project duration, as an affine function of the individual task normally distributed activity durations. We also refer to this as the least squares linear estimator. In this thesis, I use these two terms interchangeably. We explicitly obtain the solution to this optimization problem, and it turns out that the solution is closely related to the persistency problem as reviewed in Chapter 1.

One critical drawback of the estimated distribution from solving Problem (P) is that it is restricted to be normal, which is symmetric about the mean. However, in most circumstances, $Z(\tilde{\boldsymbol{c}})$ is skewed. PERT also suffers from a similar issue. To strengthen the approximation, we propose to extend the estimator to include higher order terms on $\tilde{\boldsymbol{c}}$. In particular, we also find a quadratic estimator, $Q(\tilde{\boldsymbol{c}})$, to the distribution of $Z(\tilde{\boldsymbol{c}})$ of the following form:

$$Q(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n}\beta_j(\tilde{c}_j - \mu_j) + \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma_{j_1,j_2}(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2}),$$

where α, β_j and γ_{j_1,j_2} are adjustable parameters. Interestingly, the least squares quadratic estimator is also shown to be closely related to the persistency problem, and shares some common components with the least squares linear estimator.

Outline of this chapter: In the next section, we review the related literature. In

Section 2.3, we build our least squares linear approximation with an application to maximum partial sum problem, followed by the least squares quadratic estimation in Section 2.4. Two extensions are discussed in Section 2.5. In Section 2.6, we briefly review the methods for persistency estimation in the context of project management. In Section 2.7, we present the results from our computational studies and discuss the performance of our estimators.

2.2 Literature Review

2.2.1 Distribution Problem

Our problem of interest has a long history, and it is related to the classical "distribution problem of stochastic linear programming" literature (cf. Ewbank et al. (1974), Prekopa (1966) and the references therein). The distribution of the optimal value is often approximated by numerical methods such as the Cartesian integration method (cf. Bereanu (1963)). These methods have been studied under the general framework when the uncertain parameters may appear in the objective, constraint matrix, or the right hand side of the LP problem. However, the total number of random variables are very limited due to the numerical methods employed. In the case of project management, finding the distribution of completion time in a PERT network is still an active area of research with a rich literature (cf. Yao & Chu (2007) and the references therein). Most of the work in this area has been focused on using some graphical approaches to reduce the size of the graph and to reduce the complexity of estimating the distribution of the project completion time (e.g., Dodin (1985)). Another line of research tries to find a good normal approximation to the project completion time distribution using Central Limit Theorem and moment estimation methods (e.g., Cox (1995)). We

solve this problem and show that the best normal approximation to the completion time distribution, under L^2-norm, can be obtained by solving the related persistency problem introduced by Bertsimas et al. (2006), and further studied in Natarajan et al. (2009).

2.2.2 Correlation Issues

A recent paper by Agrawal et al. (2012) investigated the loss incurred by ignoring correlations in a DRSP model and proposed a new concept called price of correlations (POC). They showed that POC is bounded from above for a certain class of cost functions, suggesting that the intuitive approach of assuming independent distributions may actually work well for these problems. However, independence conditions can be extremely difficult to capture as well. One of the negative results is given by Hagstrom (1988), who showed that computing the expected value of the longest path in a directed acyclic graph is $\#\mathcal{P}$-complete when the arc lengths are restricted to taking two possible values and independent of each other. Perhaps a DRSP model with correlation conditions is more tractable. On the other hand, Agrawal et al. (2012) also show that for some cost functions, POC can be particularly large, indicating the need of DRSP models to capture correlations. Fortunately, CPCMM partially fills this gap, which in turns further strengthens our approximation method.

In the literature of project management, there is only limited sensitivity analysis with correlated activity times. For example, Banerjee & Paul (2008) showed that in the case of a project network with multivariate normal activity completion times and a covariance matrix characterized by only nonnegative terms, the completion times of activities are positively correlated. To the best of our knowledge, none of the previous studies address the issues of correlated activities for the project management problem

when approximating the distributions of the project completion times. Our research contributes to fill this gap by assuming a general non-degenerate multivariate normal distribution for the activity times when constructing the approximating distributions.

2.2.3 Statistical Timing Analysis

Indeed, there is another version of the stochastic longest path problem in the engineering domain, which has gained much research interest in recent years. The problem arises in analyzing signal delay in the digital circuit design, and it is known as "Statistical Timing Analysis" (STA). For a recent review on STA, please refer to Blaauw et al. (2008). The problem that STA tries to address is to estimate the delay time distribution of sending a signal through a digital circuit with various gates and connecting routes. There are many sources of uncertainty in delay times, and the delay of the signals passing through a gate is usually modeled as a normal random variable. The whole problem can be cast into a stochastic longest path problem in a DAG, just like the project management problem. Correlations in STA come from various sources, e.g., sharing of common physical parameters, or proximity in physical locations (a.k.a., spacial correlations), etc. There is a growing literature in STA to model these correlations, e.g., Tsukiyama et al. (2001), Agrawal et al. (2003), Le et al. (2004), Chang & Sapatnekar (2005), Zhan et al. (2005), Zhang et al. (2005), Li et al. (2007), Khandewal & Srivastava (2007), Tang et al. (2012), etc. Commonly used methods for STA are based on CPM, PERT and approximation methods developed by Clark (1961). They are very similar to the approach adopted by Cox (1995) from project management literature (cf. Blaauw et al. (2008)). Clark (1961) studied the moment estimation problem for the maximum of a finite set of random variables following a multivariate normal distribution. The method is iterative in the sense

that $\max\{\tilde{c}_1, \tilde{c}_2, \tilde{c}_3\} \cong \max\{\max\{\tilde{c}_1, \tilde{c}_2\}, \tilde{c}_3\}$, where $\max\{\tilde{c}_1, \tilde{c}_2\}$ is assumed to be normal. Clark (1961) presented a set of analytical expressions to compute the moment estimates by taking into account of correlations among the set of random variables.

More recently, several researchers in this field have begun to look beyond the simple normality assumption and tried to model the delay time at the gate as a quadratic function of normal random variables (cf. Zhang et al. (2005), Zhan et al. (2005), Khandewal & Srivastava (2007), etc.). The key reason to extend to quadratic models is to capture the skewness in the delay time distribution in the project management problem. These methods usually try to estimate the coefficients of the quadratic function through moment matching equalities and topologically go through every node in the network applying approximation techniques on "sum" and "max" operations. Our approach is significantly different from them. By linking the problem to Stein's Identity, we can explicitly characterize the expression of the least square quadratic estimator. Besides the obvious theoretical elegance, our method is indeed more accurate computationally, which we will illustrate using examples in Section 2.7.

Contributions

We summarize the key contributions of this work as below:

- We systematically study the distribution approximation problem under the least squares framework and take into account correlations among the random coefficients.

- Linking our problems to Stein's Identity, we explicitly derive the expressions of both the least squares linear and quadratic approximations.

- We provide a new perspective to the distribution approximation problem by

transforming it into the related persistency estimation problem, for which there exist many well-established results to provide good estimates.

- In the context of project management problem, we show that knowing the criticality indices of arcs is the key to estimate the variability in the project completion time.

- By comparing against existing methods through extensive numerical studies, we demonstrate the superiority of bringing persistency into distribution approximation problem.

2.3 Least Squares Linear Estimator

As discussed before, our main idea is to approximate the distribution of $Z(\tilde{\boldsymbol{c}})$ by a normal distribution, $W(\tilde{\boldsymbol{c}})$, with the following form:

$$W(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n} \beta_j (\tilde{c}_j - \mu_j), \tag{2.3.1}$$

where α and β_j's are adjustable parameters. Note that the linear estimator in Equation (2.3.1) is also a normal distribution. The objective is to choose α and β_j's such that the expected squared deviation between $W(\tilde{\boldsymbol{c}})$ and $Z(\tilde{\boldsymbol{c}})$ is minimized. In particular, we aim to solve:

$$\text{(P)} \qquad \min_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n} \mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha - \sum_{j=1}^{n} \beta_j (\tilde{c}_j - \mu_j)\right)^2\right],$$

i.e., we want to find the least squares normal approximation to the distribution of $Z(\tilde{\boldsymbol{c}})$. It turns out that the solution to Problem (P) under the normality assumption

of $\tilde{\boldsymbol{c}}$ is related to the concept of persistency in a straightforward manner as shown in the following theorem.

Theorem 2.1. *When $\tilde{\boldsymbol{c}} \sim N(\boldsymbol{\mu}, \Sigma)$ and $\Sigma \succ 0$, the unique solution to Problem (P) is*

$$\alpha^* = \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right],\ \beta_k^* = \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right],\ k = 1, \ldots, n.$$

Proof. It is obvious that Problem (P) is convex. Then the necessary and sufficient optimality conditions of Problem (P) are

$$\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right) - \alpha^* - \sum_{j=1}^{n} \beta_j^* \left(\tilde{c}_j - \mu_j\right)\right] = 0, \text{ and}$$

$$\mathbf{E}\left[\left(Z\left(\tilde{\boldsymbol{c}}\right) - \alpha^* - \sum_{j=1}^{n} \beta_j^* \left(\tilde{c}_j - \mu_j\right)\right)\left(\tilde{c}_k - \mu_k\right)\right] = 0,\ \forall k = 1, \ldots, n.$$

Hence, an optimal solution to (P), $(\alpha^*, \boldsymbol{\beta}^*)$ should satisfy

$$\alpha^* = \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right], \text{ and}$$

$$\mathbf{E}\left[\left(Z\left(\tilde{\boldsymbol{c}}\right) - \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right] - \sum_{j=1}^{n} \beta_j^* \left(\tilde{c}_j - \mu_j\right)\right)\left(\tilde{c}_k - \mu_k\right)\right] = 0,\ \forall k = 1, \ldots, n.$$

Rearranging the second set of conditions, we get

$$Cov\left(\tilde{c}_k, Z\left(\tilde{\boldsymbol{c}}\right)\right) = \sum_{j=1}^{n} \beta_j^* \sigma_{j,k},\ \forall k = 1, \ldots, n. \tag{2.3.2}$$

The optimal objective value $Z(\tilde{\boldsymbol{c}})$ satisfies the conditions in Stein's Identity since $\partial Z(\tilde{\boldsymbol{c}})/\partial c_k = x_k(\tilde{\boldsymbol{c}})$ almost everywhere. By applying Stein's Identity on $\tilde{\boldsymbol{c}}$ and $Z(\tilde{\boldsymbol{c}})$,

we have

$$Cov\left(\tilde{c}_k, Z(\tilde{\boldsymbol{c}})\right) = \sum_{j=1}^{n} \sigma_{j,k} \mathbf{E}\left[\frac{\partial Z(\tilde{\boldsymbol{c}})}{\partial \tilde{c}_j}\right], \forall k = 1, \ldots, n.$$

Observe that

$$\begin{aligned}
\mathbf{E}\left[\frac{\partial Z(\tilde{\boldsymbol{c}})}{\partial \tilde{c}_j}\right] &= \mathbf{E}\left[\frac{\partial}{\partial \tilde{c}_j}\left(\sum_{k=1}^{n} \tilde{c}_k x_k(\tilde{\boldsymbol{c}})\right)\right] \\
&= \mathbf{E}\left[\sum_{k=1}^{n} \tilde{c}_k \frac{\partial x_k(\tilde{\boldsymbol{c}})}{\partial \tilde{c}_j} + x_j(\tilde{\boldsymbol{c}})\right] \\
&= \mathbf{E}\left[x_j(\tilde{\boldsymbol{c}})\right], \qquad \forall j = 1, \ldots, n.
\end{aligned}$$

The last equality follows from our assumptions on $\tilde{\boldsymbol{c}}$, i.e., normal and non-degenerate, so that for all $j, k = 1, \ldots, n$ $\partial x_k(\tilde{\boldsymbol{c}})/\partial c_j$ exists almost everywhere and equals to zero whenever it exists[3]. Thus, we get $\beta_j^* = \mathbf{E}\left[x_j(\tilde{\boldsymbol{c}})\right]$, $j = 1, \ldots, n$ as one solution to Equation (2.3.2), which is also unique since Σ is positive definite. Thus, the proof is complete. □

With Theorem 2.1, the problem of finding the least squares normal approximation to the distribution of $Z(\tilde{\boldsymbol{c}})$ is transformed into computing the persistency in Problem (2.1.1) as well as estimating $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. From these results, we know that the mean of estimated distribution $W(\tilde{\boldsymbol{c}})$ is the same as the mean of $Z(\tilde{\boldsymbol{c}})$. However, the variance of $W(\tilde{\boldsymbol{c}})$ is governed by the persistency values, and it is not necessarily equal to the variance of $Z(\tilde{\boldsymbol{c}})$. Indeed, the variance of $W(\tilde{\boldsymbol{c}})$ is a lower bound of the variance of

[3]Note that $\partial x_k(\tilde{\boldsymbol{c}})/\partial c_j$ is not defined when there are multiple optimal solutions to Problem (2.1.1), but in other situations, $x_k(\tilde{\boldsymbol{c}})$ does not change with a small perturbation of c_j. Please refer to the footnote in Remark 1.2 for the detailed discussion on the probability measure over the set of $\tilde{\boldsymbol{c}}$ that leads to multiple optimal solutions. Precisely, we should write the derivation process in integral form, i.e., expressing all the expectations in integral form. Then it will be clear that $\partial x_k(\tilde{\boldsymbol{c}})/\partial c_j$ can only be integrated over the support of $\tilde{\boldsymbol{c}}$ where it is defined, and hence only zero values remain in the integration expression for $\mathbf{E}[\tilde{c}_k \partial x_k(\tilde{\boldsymbol{c}})/\partial c_j]$.

$Z(\tilde{\boldsymbol{c}})$, i.e.,

$$\begin{aligned} Var\left(W\left(\tilde{\boldsymbol{c}}\right)\right) &= Var\left(\sum_{j=1}^{n}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\tilde{c}_j\right) \\ &= \left(\mathbf{E}\left[\boldsymbol{x}\left(\tilde{\boldsymbol{c}}\right)\right]\right)^T\Sigma\left(\mathbf{E}\left[\boldsymbol{x}\left(\tilde{\boldsymbol{c}}\right)\right]\right) \\ &\leq Var\left(\sum_{j=1}^{n}x_j\left(\tilde{\boldsymbol{c}}\right)\tilde{c}_j\right) \\ &= Var\left(Z\left(\tilde{\boldsymbol{c}}\right)\right). \end{aligned}$$

The inequality above is due to Cacoullos (1982), where equality holds if and only if $\mathbf{E}[x_j(\tilde{\boldsymbol{c}})]$ is constant for every $j = 1, \ldots, n$. Note that although Cacoullos' inequality,

$$Var\left(g\left(\tilde{\boldsymbol{c}}\right)\right) \geq \left(\mathbf{E}\left[\nabla g\left(\tilde{\boldsymbol{c}}\right)\right]\right)^T\Sigma\left(\mathbf{E}\left[\nabla g\left(\tilde{\boldsymbol{c}}\right)\right]\right),$$

holds for any absolutely continuous real-valued function $g\left(\tilde{\boldsymbol{c}}\right)$ with finite variance, we still need those properties of $Z\left(\tilde{\boldsymbol{c}}\right)$ and $\mathbf{E}\left[\boldsymbol{x}\left(\tilde{\boldsymbol{c}}\right)\right]$ as used in the proof of Theorem 2.1 to derive the above result. Though a lower bound, the variance of the least squares linear estimator is significantly closer to the true variance than those estimated from existing distribution approximation methods. We will illustrate this point using more examples in Section 2.7.

Although Theorem 2.1 is established under the normality assumption on $\tilde{\boldsymbol{c}}$, the result can be generalized to other distributions given the knowledge of many extensions of Stein's Identity for other distributions (cf. Adcock (2007), Barbour et al. (1992), Liu (1994), etc.). In particular, we will discuss an extension of Theorem 2.1 in Section 2.5.2 under the assumption that $\tilde{\boldsymbol{c}}$ follows a multivariate skew-normal distribution, which was originally developed in the finance area to capture the skewness in financial data (cf. Adcock (2007)).

Remark 2.2. Empirically, instead of using the observed persistency values to estimate the values for $\boldsymbol{\beta}$, we can also use $Cov(\tilde{c}_j, Z(\tilde{\boldsymbol{c}}))/\sigma_j^2$ to estimate β_j when $\tilde{c}_j$'s are inde-

pendent of each other (cf. Equation (2.3.2)). This is exactly the formula used in linear regression. One such example is estimating the beta coefficient of a risky asset under the capital asset pricing model (CAPM) in finance. This approach comes in handy when only $Z(\tilde{\boldsymbol{c}})$ is observed but not the optimal choices made, as is the case in linear regression.

In the next example, we describe an immediate application of our result on an important statistical problem: approximating the distribution of the maximum partial sum of normal random variables. The problem is critical in many areas of application, including hydrology and testing for a change-point (cf. Hurst (1951), James et al. (1987), Conniffe & Spencer (2000)). Combining our result with some classical results in probability theory, we present a closed-form expression for the least squares normal approximation of the maximum partial sum of normal random variables, from which many interesting statistics can be easily calculated, including its variance.

Example 2.3. Suppose $\tilde{c}_j$'s $(j = 1, \ldots, n)$ are independent and identically distributed (i.i.d.) normal random variables with zero mean and finite standard deviation σ . Let $S_0 = 0$ and

$$\tilde{S}_k = \tilde{c}_1 + \cdots + \tilde{c}_k,\ k = 1, \ldots, n.$$

The problem is to estimate the distribution of $\tilde{S}_{max} := \max_{k\in\{0,\ldots,n\}} \tilde{S}_k$, i.e., the maximum partial sum of $\tilde{c}_j$'s (or the maximum value of the random walk from $\tilde{c}_j$'s). Note that

$$\begin{aligned} S_{max} = \max_{\sum_{k=0}^n y_k=1, \boldsymbol{y}\geq 0} \sum_{k=0}^n S_k y_k &= \max_{\sum_{k=0}^n y_k=1, \boldsymbol{y}\geq 0} \sum_{k=1}^n \left(\sum_{j=1}^k c_j\right) y_k \\ &= \max_{\sum_{k=0}^n y_k=1, \boldsymbol{y}\geq 0} \sum_{j=1}^n \left(\sum_{k=j}^n y_k\right) c_j \end{aligned} \tag{2.3.3}$$

Applying Theorem 2.1, we get the following expression of the least squares normal

approximation to $\tilde{S}_{max}$:

$$\mathbf{E}\left[\tilde{S}_{max}\right] + \sum_{j=1}^{n}\left(\sum_{k=j}^{n}\mathbf{E}\left[y_k\left(\tilde{\boldsymbol{c}}\right)\right]\right)\tilde{c}_j, \tag{2.3.4}$$

where $\mathbf{E}[y_k(\tilde{\boldsymbol{c}})]$ is the persistency in Problem (2.3.3), i.e., the probability that the partial sum attains its maximum value at step k. The classical finite arcsine law (cf. Andersen (1953)) states that this probability does not depend on the distribution of $\tilde{c}_j$ provided that $\tilde{c}_j$ is symmetric around the mean 0:

$$\mathbf{E}\left[y_k(\tilde{\boldsymbol{c}})\right] = \begin{pmatrix} 2k \\ k \end{pmatrix}\begin{pmatrix} 2n-2k \\ n-k \end{pmatrix}\frac{1}{2^{2n}},\ k = 1, \ldots, n.$$

Observe that the variance of our approximation in Equation (2.3.4) is solely determined by the second term through persistency. Hence, we get the following closed-form lower bound to the variance of the maximum partial sum:

$$\frac{\sigma^2}{2^{4n}}\sum_{j=1}^{n}\left[\sum_{k=j}^{n}\begin{pmatrix} 2k \\ k \end{pmatrix}\begin{pmatrix} 2n-2k \\ n-k \end{pmatrix}\right]^2.$$

The above result expands the current literature by providing a different way to estimate the variance of the maximum partial sum of i.i.d. normal random variables. Note that there exists various methods in literature to compute or estimate $\mathbf{E}[\tilde{S}_{max}]$, with which we get a complete characterization of the least squares normal approximation as shown in Equation (2.3.4). For instance, when $\mu = 0$, $\sigma = 1$, i.e., $\tilde{c}_i$'s are independent standard normal random variables, Spitzer (1956) showed that

$$\mathbf{E}\left[\tilde{S}_{max}\right] = \sum_{k=1}^{n}\frac{1}{k}\mathbf{E}\left[\tilde{S}_k^+\right] = \mathbf{E}\left[\tilde{c}_1^+\right]\sum_{k=1}^{n}\frac{1}{\sqrt{k}},$$

where $\tilde{S}_k^+ = \max\{0, \tilde{S}_k\}$, and $\tilde{c}_1^+ = \max\{0, \tilde{c}_1\}$.

Remark 2.4. Note that Brownian motion can be treated as the limit of symmetric random walks. Hence, it is natural to use the results for Brownian motion to asymptotically approximate the distribution of the maximum partial sum. It is well-known that the running maximum $M_t = \max\{B_s : s \in [0,t]\}$ of standard Brownian motion B_t has the density $2\phi(m/\sqrt{t})$, where ϕ is the standard normal density function. However, this method is inaccurate except for very large samples (cf. Conniffe & Spencer (2000)).

2.4 Least Squares Quadratic Estimator

In the previous section, we show how to approximate the distribution of $Z(\tilde{\boldsymbol{c}})$ using a linear estimator $W(\tilde{\boldsymbol{c}})$. By "linear", we mean that $W(\tilde{\boldsymbol{c}})$ is a linear function in $\tilde{\boldsymbol{c}}$. As discussed in the introduction, to address the problem of skewness in $Z(\tilde{\boldsymbol{c}})$, we propose to extend our estimator to incorporate higher order terms on $\tilde{\boldsymbol{c}}$. The estimator we consider is denoted as $Q(\tilde{\boldsymbol{c}})$ with the following form:

$$Q(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n} \beta_j(\tilde{c}_j - \mu_j) + \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2}),$$

where α, β_j's and γ_{j_1,j_2}'s are adjustable parameters. Then the least squares quadratic estimation problem can be formulated as:

$$\text{(Q)} \qquad \min_{\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^n,\Gamma\in\mathbb{R}^{n\times n}} \mathbf{E}\left[\Big(Z(\tilde{\boldsymbol{c}}) - \alpha - \textstyle\sum_{j=1}^{n} \beta_j(\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\Big)^2\right],$$

where the matrix Γ is defined in the way that makes our notation compact, $\Gamma_{j_1,j_2} := (1/2)\gamma_{j_1,j_2}$, for $1 \leq j_1 < j_2 \leq n$, $\Gamma_{j_1,j_2} := (1/2)\gamma_{j_2,j_1}$, for $1 \leq j_2 < j_1 \leq n$, and $\Gamma_{j_1,j_2} := \gamma_{j_1,j_2}$, for $j_1 = j_2 = 1, \ldots, n$.

Following a similar approach as in Section 2.3, we can also derive the solution to Problem (Q). Interestingly, adding the quadratic term does not affect the solution of $\boldsymbol{\beta}$, which is still the persistency, as presented in the following theorem. Notation-wise, we use "$\bullet$" to denote the inner product of two matrices.

Theorem 2.5. *When $\tilde{\boldsymbol{c}} \sim N(\boldsymbol{\mu}, \Sigma)$, a solution $(\alpha^*, \boldsymbol{\beta}^*, \Gamma^*)$ to Problem (Q) can be characterized as follows:*

$$\alpha^* = \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*, \quad \beta_k^* = \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right], \, k = 1, \ldots, n,$$

and Γ^* is symmetric and satisfies the following system of $(n^2+n)/2$ linear equations:

$$\begin{aligned} &\sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^* \left(\sigma_{j_1,k_1}\sigma_{j_2,k_2} + \sigma_{j_1,k_2}\sigma_{k_1,j_2}\right) \\ &\quad = \sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right)\sigma_{j,k_2}, \, \forall 1 \leq k_1 \leq k_2 \leq n. \end{aligned}$$

Proof. Since Problem (Q) is convex, its necessary and sufficient optimality conditions are

$$\mathbf{E}\left[Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n} \beta_j^*(\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^*(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\right] = 0,$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n} \beta_j^*(\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^*(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\right)(\tilde{c}_k - \mu_k)\right] = 0,$$

$$\forall k = 1, \ldots, n, \text{ and}$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}})-\alpha^*-\sum_{j=1}^{n}\beta_j^*(\tilde{c}_j-\mu_j)\right.\right.$$
$$\left.\left.-\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma_{j_1,j_2}^*(\tilde{c}_{j_1}-\mu_{j_1})(\tilde{c}_{j_2}-\mu_{j_2})\right)(\tilde{c}_{k_1}-\mu_{k_1})(\tilde{c}_{k_2}-\mu_{k_2})\right]=0,$$
$$\forall 1\le k_1\le k_2\le n.$$

Hence, an optimal solution $(\alpha^*,\boldsymbol{\beta}^*,\Gamma^*)$ should satisfy

$$\alpha^*=\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]-\Sigma\bullet\Gamma^*,$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}})-\alpha^*-\sum_{j=1}^{n}\beta_j^*(\tilde{c}_j-\mu_j)\right)(\tilde{c}_k-\mu_k)\right]$$
$$-\mathbf{E}\left[\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma_{j_1,j_2}^*(\tilde{c}_{j_1}-\mu_{j_1})(\tilde{c}_{j_2}-\mu_{j_2})(\tilde{c}_k-\mu_k)\right]=0,\ \forall k=1,\ldots,n,\ \text{and}$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}})-\alpha^*-\sum_{j=1}^{n}\beta_j^*(\tilde{c}_j-\mu_j)\right)(\tilde{c}_{k_1}-\mu_{k_1})(\tilde{c}_{k_2}-\mu_{k_2})\right]$$
$$-\mathbf{E}\left[\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma_{j_1,j_2}^*(\tilde{c}_{j_1}-\mu_{j_1})(\tilde{c}_{j_2}-\mu_{j_2})(\tilde{c}_{k_1}-\mu_{k_1})(\tilde{c}_{k_2}-\mu_{k_2})\right]=0,$$
$$\forall 1\le k_1\le k_2\le n.$$

From Isserlis' Theorem, if random variable $(\tilde{z}_1,\ldots,\tilde{z}_n)$ follows a zero mean multivariate normal distribution, then

$$\mathbf{E}\left[\prod_{i=1}^{n}\tilde{z}_i\right]=\begin{cases}0, & \text{if } n \text{ is odd},\\ \sum\prod\mathbf{E}\left[\tilde{z}_i\tilde{z}_j\right], & \text{if } n \text{ is even},\end{cases}$$

where $\sum\prod$ means summing over all distinct ways of partitioning $(\tilde{z}_1,\ldots,\tilde{z}_n)$ into pairs (cf. Isserlis (1918)). In particular, when $n=3,4$,

$$\mathbf{E}\left[\tilde{z}_1\tilde{z}_2\tilde{z}_3\right]=0,\ \text{and}$$

$$\mathbf{E}\left[\tilde{z}_1\tilde{z}_2\tilde{z}_3\tilde{z}_4\right] = \mathbf{E}\left[\tilde{z}_1\tilde{z}_2\right]\mathbf{E}\left[\tilde{z}_3\tilde{z}_4\right] + \mathbf{E}\left[\tilde{z}_1\tilde{z}_3\right]\mathbf{E}\left[\tilde{z}_2\tilde{z}_4\right] + \mathbf{E}\left[\tilde{z}_1\tilde{z}_4\right]\mathbf{E}\left[\tilde{z}_2\tilde{z}_3\right].$$

Applying Isserlis' Theorem, we can reduce the optimality conditions into

$$\alpha^* = \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*,$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n}\beta_j^*(\tilde{c}_j - \mu_j)\right)(\tilde{c}_k - \mu_k)\right] = 0,\ \forall k = 1,\ldots,n, \tag{2.4.1}$$

and

$$\begin{aligned}&\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^*\right)(\tilde{c}_{k_1} - \mu_{k_1})(\tilde{c}_{k_2} - \mu_{k_2})\right]\\ &\quad - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma_{j_1,j_2}^*\left(\sigma_{j_1,j_2}\sigma_{k_1,k_2} + \sigma_{j_1,k_1}\sigma_{j_2,k_2} + \sigma_{j_1,k_2}\sigma_{k_1,j_2}\right) = 0,\ \forall 1 \le k_1 \le k_2 \le n.\end{aligned} \tag{2.4.2}$$

Further simplifying Equation (2.4.1), we get

$$\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})(\tilde{c}_k - \mu_k)\right] = \sum_{j=1}^{n}\beta_j^*\sigma_{j,k},\ \forall k = 1,\ldots,n.$$

Since $\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})(\tilde{c}_k - \mu_k)\right] = Cov\left(\tilde{c}_k, Z(\tilde{\boldsymbol{c}})\right)$, we arrive at the same conditions as Equation (2.3.2) in Theorem 2.1. Therefore, following the same argument, we have $\beta_k^* = \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right]$, $k = 1,\ldots,n$, which is unique if Σ is positive definite.

Consider a part of the first term in Equation (2.4.2),

$$\begin{aligned}\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})(\tilde{c}_{k_1} - \mu_{k_1})(\tilde{c}_{k_2} - \mu_{k_2})\right] = &\ \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\tilde{c}_{k_2}\right] - \mu_{k_1}\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_2}\right]\\ &-\mu_{k_2}\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\right] + \mu_{k_1}\mu_{k_2}\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]\\ = &\ \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\tilde{c}_{k_2}\right] - \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\right]\mu_{k_2}\\ &-\mu_{k_1}\left(\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_2}\right] - \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]\mu_{k_2}\right)\\ = &\ Cov\left(Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}, \tilde{c}_{k_2}\right) - \mu_{k_1}Cov\left(Z(\tilde{\boldsymbol{c}}), \tilde{c}_{k_2}\right).\end{aligned}$$

It is straightforward to apply Stein's Identity on $Cov\left(Z(\tilde{\boldsymbol{c}}), \tilde{c}_{k_2}\right)$ as we have done before, i.e.,

$$Cov\left(Z(\tilde{\boldsymbol{c}}), \tilde{c}_{k_2}\right) = \sum_{j=1}^{n} \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2}.$$

For the other term, $Cov\left(Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}, \tilde{c}_{k_2}\right)$, we can also use Stein's Identity,

$$\begin{aligned} Cov\left(Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}, \tilde{c}_{k_2}\right) = & \sum_{j=1}^{n} \mathbf{E}\left[\frac{\partial Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}}{\partial c_j}\right] Cov\left(\tilde{c}_j, \tilde{c}_{k_2}\right) \\ = & \sum_{j=1}^{n} \mathbf{E}\left[\tilde{c}_{k_1}\frac{\partial Z(\tilde{\boldsymbol{c}})}{\partial c_j} + Z(\tilde{\boldsymbol{c}})\frac{\partial \tilde{c}_{k_1}}{\partial c_j}\right] \sigma_{j,k_2} \\ = & \sum_{j=1}^{n} \mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2} + \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] \sigma_{k_1,k_2}, \end{aligned}$$

where the last equality follows from the same argument as in the proof of Theorem 2.1. Therefore,

$$\begin{aligned} \mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^*\right)\left(\tilde{c}_{k_1} - \mu_{k_1}\right)\left(\tilde{c}_{k_2} - \mu_{k_2}\right)\right] = & \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\left(\tilde{c}_{k_1} - \mu_{k_1}\right)\left(\tilde{c}_{k_2} - \mu_{k_2}\right)\right] - \alpha^*\sigma_{k_1,k_2} \\ = & \sum_{j=1}^{n} \mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2} + \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] \sigma_{k_1,k_2} \\ & -\mu_{k_1} \sum_{j=1}^{n} \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2} \\ & -\left(\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*\right) \sigma_{k_1,k_2} \\ = & \sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right) \sigma_{j,k_2} \\ & +\sigma_{k_1,k_2}\Sigma \bullet \Gamma^*. \end{aligned}$$

Substituting this into Equation (2.4.2), we get a system of $(n^2+n)/2$ linear equations on Γ^*,

$$\begin{aligned} & \sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right) \sigma_{j,k_2} + \sigma_{k_1,k_2}\Sigma \bullet \Gamma^* \\ & \quad - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma^*_{j_1,j_2}\left(\sigma_{j_1,j_2}\sigma_{k_1,k_2} + \sigma_{j_1,k_1}\sigma_{j_2,k_2} + \sigma_{j_1,k_2}\sigma_{k_1,j_2}\right) = 0, \ \forall 1 \le k_1 \le k_2 \le n, \end{aligned}$$

which reduces to

$$\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma^*_{j_1,j_2}\left(\sigma_{j_1,k_1}\sigma_{j_2,k_2}+\sigma_{j_1,k_2}\sigma_{k_1,j_2}\right)$$
$$=\sum_{j=1}^{n}\left(\mathbf{E}\left[\tilde{c}_{k_1}x_j\left(\tilde{\boldsymbol{c}}\right)\right]-\mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right)\sigma_{j,k_2},\ \forall 1\le k_1\le k_2\le n.$$

Thus, we complete the proof. □

From Theorem 2.5, the problem of finding the least squares quadratic estimator for the distribution of $Z(\tilde{\boldsymbol{c}})$ is again transformed into a persistency problem, i.e., estimating $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, and $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. The additional requirement to estimate $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, i.e., the interaction between random coefficients and the optimal solution, can be interpreted as the increased difficulty of adding the quadratic terms in the estimation. However, as shown in Chapter 1, the estimation of $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$ can be obtained as a by-product when we estimate the persistency using semidefinite programming methods.

In general, Γ^* may not be unique due to the correlation structures. However, when $\tilde{c}_j$'s are uncorrelated and not degenerate, we do have a simple and unique solution.

Corollary 2.6. *When $\tilde{c}_j$'s are uncorrelated and each follows a normal distribution with $\sigma_j^2>0$, there is a unique solution to Problem (Q) as follows:*

$$\begin{aligned}
\alpha^* &= \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]-\Sigma\bullet\Gamma^*,\\
\beta_k^* &= \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right],\ k=1,\ldots,n,\\
\gamma^*_{k_1,k_2} &= \frac{\mathbf{E}\left[\tilde{c}_{k_1}x_{k_2}\left(\tilde{\boldsymbol{c}}\right)\right]-\mu_{k_1}\mathbf{E}\left[x_{k_2}\left(\tilde{\boldsymbol{c}}\right)\right]}{\sigma^2_{k_1}},\ \forall 1\le k_1<k_2\le n,\\
\gamma^*_{k,k} &= \frac{\mathbf{E}\left[\tilde{c}_k x_k\left(\tilde{\boldsymbol{c}}\right)\right]-\mu_k\mathbf{E}\left[x_k\left(\tilde{\boldsymbol{c}}\right)\right]}{2\sigma_k^2},\ \forall k=1,\ldots,n.
\end{aligned}$$

It would be interesting to know whether the least quadratic estimation is convex in $\tilde{\boldsymbol{c}}$. Unfortunately, Hertog et al. (2002) observed that the least squares quadratic

approximation of a multivariate convex function in a finite set of points is not necessarily convex even though it is convex for a univariate convex function. Similarly for our problem, we cannot guarantee that the least quadratic estimation is convex. It is however possible to enforce convexity through imposing a semidefinite constraint on Γ, but the resulting problem will not exhibit a nice and explicit characterization of the solution as the unconstrained version.

2.5 Extensions

In this section, we discuss two interesting extensions of our approximation models. The first extension applies to the situation when only partial information on the random coefficients is available, and the second extension addresses the case when the random coefficient vector follows a multivariate skew-normal distribution.

2.5.1 Distribution Approximation Using Partial Information

In both least squares linear and quadratic approximations, we make use of all the distributional information on the random coefficient vector $\tilde{\boldsymbol{c}}$ in constructing the approximating distributions. What would happen if we only have the information on a subset of the random coefficients? Suppose the set of random coefficients are divided into two independent components, $\tilde{\boldsymbol{c}} \in \mathbb{R}^{n_1}$ and $\tilde{\boldsymbol{d}} \in \mathbb{R}^{n_2}$, both of which follow multivariate normal distributions, and $n_1 + n_2 = n$. However, we only know the mean and covariance of $\tilde{\boldsymbol{c}}$, but not $\tilde{\boldsymbol{d}}$. Let $\boldsymbol{\mu}^c$ and Σ^c denote the mean vector and covariance matrix of $\tilde{\boldsymbol{c}}$, respectively, and assume $\Sigma^c \succ 0$. We use $Z(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}})$ to represent the optimal

objective value as defined below:

$$Z\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right) := \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{P}} \sum_{j=1}^{n_1} \tilde{c}_i x_i + \sum_{k=1}^{n_2} \tilde{d}_i y_i,$$

where $\mathcal{P}$ is the domain of $\boldsymbol{x}$ and $\boldsymbol{y}$ defined by linear and nonnegative constraints as well as some binary conditions.

With the limited information on $\tilde{\boldsymbol{c}}$, we would like to find the least squares normal approximation of $Z(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}})$, denoted as $W^c(\tilde{\boldsymbol{c}})$ of the following form:

$$W^c\left(\tilde{\boldsymbol{c}}\right) = \alpha + \sum_{j=1}^{n_1} \beta_j(\tilde{c}_j - \mu_j^c).$$

The least squares normal approximation problem with partial information is then formulated as follows:

$$(\mathrm{P^C}) \qquad \min_{\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^{n_1}} \mathbf{E}\left[\left(Z\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right) - \alpha - \sum_{j=1}^{n_1} \beta_j\left(\tilde{c}_j - \mu_j^c\right)\right)^2\right].$$

Surprisingly, the optimal parameters turn out to be independent on the moments of the unknown component $\tilde{\boldsymbol{d}}$ provided that $\tilde{\boldsymbol{d}}$ is independent of $\tilde{\boldsymbol{c}}$. Instead, they are related only through persistency values.

Proposition 2.7. *With the assumptions on $\tilde{\boldsymbol{c}}$ and $\tilde{\boldsymbol{d}}$ described as above, the unique solution to Problem (P^C) is*

$$\alpha^* = \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right)\right],\ \beta_j^* = \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right)\right],\ j = 1, \ldots, n_1.$$

The proof of above proposition is a straightforward extension of that of Theorem 2.1. One only needs the independence assumption on $\tilde{\boldsymbol{c}}$ and $\tilde{\boldsymbol{d}}$ to simplify the deriva-

tion, i.e., $Cov(\tilde{c}_j, \tilde{d}_k) = 0$, $\forall\ j = 1, \ldots, n_1,\ k = 1, \ldots, n_2$. Indeed, the uncorrelated assumption is sufficient, but for multivariate normally distributed random variables, uncorrelatedness is equivalent to independence.

Besides the project management problem, another potential application of this extended model is the choice prediction problem, where a manufacturer decides on the design parameters of products and consumers make the purchasing decision after observing the products by maximizing their utilities. The consumer's utility function certainly depends on some parameters that are observable to the manufacturer, like the product design parameters and demographic characteristics. However, there might also exist some private decision criteria only known by the consumer, so the resulting purchasing behaviour will appear random to the manufacturer. Therefore, the manufacturer's task is to predict the consumer's choice based on limited information. Under such settings, $Z(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}})$ can be interpreted as the random utility function of the consumer, and $\tilde{\boldsymbol{c}}$ represents the observable parameters, and $\tilde{\boldsymbol{d}}$ represents the consumer's private decision criteria. The manufacturer wants to find the best prediction of the consumer's choice as the observable parameters change. Our result in Proposition 2.7 says that under some assumptions on the decision parameters, the least squares normal approximation can be constructed by observing the consumer's preference over those observable parameters, $\mathbf{E}[x_j(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}})]$. Such information can be obtained from carefully designed surveys and conjoint analysis. Mishra et al. (2013) analyzed an automotive design problem which falls in the category of our problem description, although they treated the problem using a different model with different assumptions.

It is also possible to extend our least squares quadratic estimator for the partial

information problem with the following distributional form:

$$Q^c(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n_1} \beta_j \left(\tilde{c}_j - \mu_j^c\right) + \sum_{j_1=1}^{n_1} \sum_{j_2=j_1}^{n_1} \gamma_{j_1,j_2} (\tilde{c}_{j_1} - \mu_{j_1}^c)(\tilde{c}_{j_2} - \mu_{j_2}^c).$$

Defining Γ in the same manner as before, we present the solution to the least squares quadratic estimation problem with partial information in the following proposition. The detailed proof is omitted, since it is similar to that of Theorem 2.5.

Proposition 2.8. *With the same assumptions on $\tilde{\boldsymbol{c}}$ and $\tilde{\boldsymbol{d}}$ as in Proposition 2.7, the least squares quadratic estimator's parameters $(\alpha^*, \boldsymbol{\beta}^*, \Gamma^*)$ can be characterized as follows:*

$$\alpha^* = \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right)\right] - \Sigma^c \bullet \Gamma^*, \ \beta_j^* = \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right)\right], \ j = 1, \ldots, n_1,$$

and Γ^* is symmetric and satisfies the following system of $(n_1^2 + n_1)/2$ linear equations:

$$\begin{aligned} &\sum_{j_1'=1}^{n_1} \sum_{j_2'=j_1}^{n_1} \gamma_{j_1',j_2'}^* \left(\Sigma_{j_1',j_1}^c \Sigma_{j_2',j_2}^c + \Sigma_{j_1',j_2}^c \Sigma_{j_1,j_2'}^c\right) \\ &\quad = \sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{j_1} x_j\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right)\right] - \mu_{j_1} \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{d}}\right)\right]\right) \sigma_{j,j_2}, \ \forall 1 \le j_1 \le j_2 \le n_1. \end{aligned}$$

2.5.2 Multivariate Skew-Normal Distribution

To address the shape limitation of normal distribution, i.e., symmetry, another research field currently under intense development is on skew-normal distribution. The idea is to introduce a skewness shock to normally distributed random variables. For a comprehensive review of this topic, please refer to Azzalini (2005). The idea of skew-normal distribution attracts our attention because it allows a nice generalization of Stein's Identity contributed by Adcock (2007). Following our notation, suppose the random vector $\tilde{\boldsymbol{c}}$ can be expressed as $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{c}}^N + \boldsymbol{\lambda}\tilde{u}$, where $\tilde{\boldsymbol{c}}^N$ has a multivariate

normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma \succ 0$, and $\tilde{u}$ is a normal random variable independent of $\tilde{\boldsymbol{c}}^N$ with mean τ and variance 1 truncated from below at zero, and $\boldsymbol{\lambda} \in \mathbb{R}^n$ is a vector of skewness parameters, which may take any real values. Then the probability distribution of $\tilde{\boldsymbol{c}}$ is multivariate skew-normal with parameters $\boldsymbol{\mu}$, Σ, $\boldsymbol{\lambda}$, and τ, denoted as $\tilde{\boldsymbol{c}} \sim MSN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \tau)$. Adcock (2007) extended Stein's Identity for $\tilde{\boldsymbol{c}} \sim MSN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \tau)$ as follows:

Lemma 2.9. *[Extension of Stein's Identity for MSN Distribution] Let $\tilde{\boldsymbol{c}}$ be an n dimensional random vector that has the distribution $MSN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \tau)$. For any function $h(c_1, \ldots, c_n) : \mathbb{R}^n \to \mathbb{R}$ such that $\partial h(c_1, \ldots, c_n)/\partial c_j$ is continuous almost everywhere and $\mathbf{E}[|\partial h(\tilde{\boldsymbol{c}})/\partial c_j|] < \infty$, $\forall j = 1, \ldots, n$. Then*

$$Cov(\tilde{\boldsymbol{c}}, h(\tilde{\boldsymbol{c}})) = \left(\Sigma + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right) \mathbf{E}\left[\nabla h(\tilde{\boldsymbol{c}})\right] + \boldsymbol{\lambda}\left(\mathbf{E}[h(\tilde{\boldsymbol{c}}^N)] - \mathbf{E}\left[h(\tilde{\boldsymbol{c}})\right]\right)\xi(\tau),$$

where $\tilde{\boldsymbol{c}}^N \sim N(\boldsymbol{\mu}, \Sigma)$, and $\nabla h(\tilde{\boldsymbol{c}})$ is as defined in Lemma 1.3, and $\xi(\tau) = \phi(\tau)/\Phi(\tau)$. $\phi(.)$ and $\Phi(.)$ are the probability density function and cumulative distribution function of the standard normal random variable, respectively.

With Lemma 2.9, we can easily extend our results in Theorem 2.1 for $\tilde{\boldsymbol{c}} \sim MSN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \tau)$. The proof is a simple modification of that of Theorem 2.1 and hence omitted here.

Proposition 2.10. *When $\tilde{\boldsymbol{c}} \sim MSN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \tau)$ and $\Sigma \succ 0$, the unique solution to Problem (P) is*

$$\begin{aligned}
\alpha^* = &\ \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right], \\
\boldsymbol{\beta}^* = &\ \mathbf{E}\left[\boldsymbol{x}(\tilde{\boldsymbol{c}})\right] + \Sigma^{-1}\boldsymbol{\lambda}\left\{\boldsymbol{\lambda}^T\mathbf{E}\left[\boldsymbol{x}(\tilde{\boldsymbol{c}})\right] + \left(\mathbf{E}[Z(\tilde{\boldsymbol{c}}^N)] - \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]\right)\xi(\tau)\right\}.
\end{aligned}$$

Again, we transform the distribution estimation problem into the related persistency problem. The only difference is that we need to estimate one more parameter,

$\mathbf{E}[Z(\tilde{\boldsymbol{c}}^N)]$, the expected objective value without skewness shock. Observe that when $\boldsymbol{\lambda} = 0$, we get back the solution to normally distributed $\tilde{\boldsymbol{c}}$.

We do not generalize the quadratic estimator for the multivariate skew-normal distribution because the purpose of introducing the quadratic terms can be partially fulfilled by modeling the uncertainty with the skew-normal distribution. This not only introduces skewness to the random coefficients, $\tilde{\boldsymbol{c}}$, but also the final approximating distribution, $W(\tilde{\boldsymbol{c}})$.

2.6 Approximating Persistency Values

Theorem 2.1 and 2.5 transform the distribution approximation problems under the least squares framework into the problems of estimating certain stochastic parameters, i.e., $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, and $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. Note that we are not bound to use just one method to estimate all these parameters. Indeed, we can choose any methods deemed appropriate for each parameter. From this point of view, there is a huge literature we can make use of to estimate these parameters, especially for $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. What we show next is only one possible approach. To avoid the criticism of speculation, we only choose some basic and generic estimation methods without sophisticated modifications to tailor to our test problems. Hence, we leave plenty of room for users to improve the approximation accuracy for specific applications and better demonstrate the power of our least squares approximations.

In literature, the problem of estimating the expected objective value of a stochastic optimization problem has been studied for a long time. In case of the project management problem, the search for the expected project completion time started half century ago (cf. Clark (1961), and Fulkerson (1962), etc.) and is still an active research topic (cf. Yao & Chu (2007)). In all the numerical studies presented next, we choose to use

the classical estimation method proposed by Clark (1961). The technique developed in Clark (1961) forms the building block of most modern distribution approximation methods, especially for the project management problem and statistical timing analysis. Note that we only use the original estimation methods from Clark (1961) to estimate $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ without considering any further extensions and refinements.

On the other hand, although the concept of persistency was only brought into the optimization area since Bertsimas et al. (2006), it has long existed as criticality index in the project management area. Similar to the case of approximating the distribution of the project completion time, the majority of the research work on estimating criticality has been focusing on developing heuristics algorithms based on the topological properties of the project networks, and the uncertainty is usually treated by discretization and/or stochastic dominance (cf. Dodin (1984), Dodin & Elmaghraby (1985), etc.). More advanced method combines the strength of different approaches to obtain new hybrid approach. For example, Bowman (1995) utilized the geometric properties of the networks to reduce the computational requirement of simulation. The common limitation of these methods is the lack of consideration of correlations among different activity completion times.

Besides these specific estimation methods for the project management method, there is a series of generic conic programming based models for persistency estimation as reviewed before (cf. Natarajan et al. (2009), Lasserre (2010), Mishra et al. (2012), Natarajan et al. (2011), Kong et al. (2013) etc.). By "generic" we mean that these methods work on any optimization problems and do not exploit any specific problem structure like the network flow in the project management problem. We will adopt one of the conic programming method, CPCMM, to estimate $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$ and $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$. As reviewed in Chapter 1, the matrix decision variable encodes both information for

$\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$ and $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$. Besides its generic nature, another key reason that we choose this model is its ability to capture correlations among random coefficients.

However, a key drawback of CPCMM is that it ignores the distributional information. Hence, when $\tilde{\boldsymbol{c}}$ is normally distributed, CPCMM only gives an upper bound on $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ and estimates of the persistency and $\mathbf{E}[\tilde{c}_j x_i(\tilde{\boldsymbol{c}})]$. We will discuss some ways to improve the persistency estimation in Chapter 4. In this stage, however, we do not implement any improvement methods, because the persistency estimates from CPCMM are good enough for most examples, and we want to keep the focus of this chapter on distribution approximation rather than persistency estimation. As discussed in Chapter 1, computationally we will solve the doubly nonnegative relaxation of the completely positive programs. Despite all these numerical inaccuracies, we will show that our approximation methods are still practically attractive due to the use of persistency in the approximation and the flexibility of our methods.

2.7 Computational Study

We divide the computational analysis in this section into two main parts. In the first part, we assess the quality of our linear and quadratic estimations by assuming the availability of exact persistency values. In the second part, we build approximating distributions based on the estimated persistency values discussed in the previous section. For both parts, we mainly focus on the application in project management problems, and the performance of our estimators are gauged against various existing methods. In particular, we are interested in comparing our methods with PERT. All the computational studies are conducted in MATLAB environment, and LPs are solved using CPLEX solver with YALMIP interface in MATLAB (cf. Löfberg (2004)). SDPs are solved using SDPT3 solver (cf. Toh et al. (1999), Tutuncu et al. (2003)).

Before presenting the results, we would like to discuss the performance measures first.

Note that all the examples studied in this section are simple test projects from the literature, which are meant for illustration of the least squares approximation framework. As reviewed in the previous section, there are many more sophisticated methods available to improve the approximations in terms of both accuracy and efficiency for larger problems. However, since it would digress the focus of the current chapter, we opt not to go into details in this direction here and leave this issue to Chapter 4.

2.7.1 Performance Measures

For different instances of the project management problem, we try to compare our approaches with as many existing approximation methods as possible, neglecting approaches that require the activity duration distribution to be discrete, e.g., bounding distribution method by Kleindorfer (1971).

The key performance indicator we consider is the expected square deviation (ESD), which is also the objective function we try to minimize in obtaining our least square approximations. Unfortunately, almost all the approximating distributions derived using previous methods do not reside in the same probability space as $Z(\tilde{\boldsymbol{c}})$, which makes it impossible to compute the squared deviation from $Z(\tilde{\boldsymbol{c}})$. This problem arises since the traditional approaches solely focus on the distribution (like tail probabilities, etc.) but overlook the approximation error between the approximated completion time and the true completion time under a specific realization of the random activity durations. For example, Cox (1995) assumed the project completion time to be normally distributed at first, and then tried to estimate the moments of the completion time. Hence, we have to resort to other measures to compare the performance of different approximation methods including descriptive statistics, like mean, standard deviation, and

skewness. In addition, we also employ the following measure to quantify the distance between two distributions:

$$\text{Square Norm Distance}(F,G) = \text{SND}(F,G) := \int_0^1 \left[F^{-1}(y) - G^{-1}(y)\right]^2 dy$$

where F and G are the cumulative distribution functions of two distributions.

On the other hand, PERT approach simply considers the expected duration of each activity when choosing the critical path (i.e., the path with longest expected completion time), and then use the mean and variance of this critical path to approximate the mean and variance of the project completion time, respectively. Finally, resorting to the Central Limit Theorem, PERT assumes that the project completion time is normally distributed (cf. MacCrimmon & Ryavec (1964)). When the project activities follow a multivariate normal distribution, PERT uses the distribution of the critical path to approximate the distribution of the project completion time. Thus, it admits the computation of squared deviation.

2.7.2 With Exact Persistency Values

The purposes of this analysis are twofold. First, we would like to know how accurate the least squares linear approximation can be and how the least squares quadratic approximation can improve the estimation accuracy. Using the exact persistency values, we rule out the impact of errors from estimating persistency values, which might either increase or decrease the accuracy of our least squares approximations and complicate the analysis. Second, the results from such analysis serve as a benchmark for the following subsection where we discuss the approximations based on estimated persistency values. It helps to decide how much estimation error in persistency values is tolerable for constructing our approximating distributions.

By "exact", we mean the persistency values are directly computed from simulation (i.e., sample estimates of $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$ and $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$) rather than some persistency estimation models. For small project management problems, simulation is still possible because their deterministic versions are solvable in polynomial time.

With the following example, besides comparing the performance of different approximation methods, we also illustrate the importance of considering correlations among activities.

Example 2.11. The project network consists of four nodes and five arcs as shown in Figure 2.7.1. All activities are independent and normally distributed with mean and variance both equal to one.

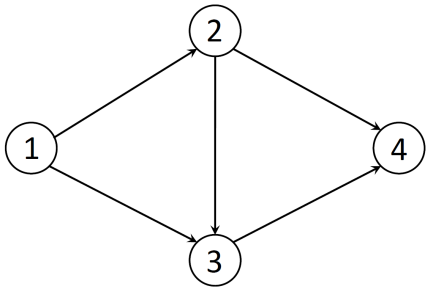


Figure 2.7.1: The project network in Example 2.11

The network in Example 2.11 is the "Wheatstone bridge" network from Lindsey (1972), and later regarded as the "forbidden graph" by Dodin (1985), since it is the basic evidence of graph irreducibility. Ord (1991) summarized the results for this graph documented in literature with normally distributed activity durations, and also provided the results from his discrete approximation method with a parameter k indicating the number of discrete points used to approximate the normal distribution. In fact, the approximated distributions obtained by Ord (1991) should be a discrete distribution. However, we extend his theory in computing the square norm distance

by assuming the final approximated distribution follows a normal distribution with the moments derived from his original procedure. All these results are presented in Table 2.7.1, where T denotes the project completion time, and $\sigma(T)$ denotes its standard deviation, and $sk(T)$ denotes its skewness. "Error on $\sigma(T)$" is computed as the absolute relative error against the simulation result. The new result from our method is also presented in Table 2.7.1 under "LSN" and "LSQ", where "LSN" stands for "Least Squares Normal" and "LSQ" stands for "Least Squares Quadratic". We conducted 10^6 simulation runs to estimate the persistency values. To avoid repetition, the estimation results of our least square approximations using estimated persistency values are also included in this table labeled as "LSNe" and "LSQe". We will not discuss those results until next subsection.

Approximation Method		$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND
10^6 simulation		3.516	1.39	-	0.28	-	-
Numerical integration		3.483	1.47	5.76%	0	-	0.017
Ord (1991)	$k=2$	3.261	0.70	49.64%	0	-	0.543
	$k=3$	3.485	1.04	25.18%	0	-	0.128
	$k=4$	3.525	1.08	22.32%	0	-	0.101
	$k=5$	3.582	1.15	17.27%	0	-	0.068
	$k=6$	3.594	1.15	17.27%	0	-	0.069
Cox (1995)		3.639	1.69	21.58%	0	-	0.116
PERT		3.000	1.73	24.46%	0	0.973	0.395
LSN		3.515	1.27	8.63%	0	0.311	0.021
LSQ		3.515	1.36	2.16%	0.47	0.078	0.005
LSNe		3.518	1.26	8.80%	0	0.311	0.022
LSQe		3.519	1.44	3.76%	0.60	0.124	0.014

Table 2.7.1: Estimation results for Example 2.11

Theoretically, if we have exact persistency values, the mean estimates from both least squares approximations should be the same as the true mean project completion time. However, in simulation, the sampling errors are magnified through linear and quadratic terms in the estimation formulas. Even though we use the sample mean

project completion time from simulation as the value of α in least quadratic estimators, the sampling errors from linear and quadratic terms make the final mean estimates of our approximations different from that of simulation.

From Table 2.7.1, we can see that except the numerical integration approach and our quadratic estimator, the least squares linear estimation gives the best estimate for the standard deviation, in terms of absolute relative error. Regardless of the high accuracy, the integration approach would be too tedious to be applicable for even medium-size networks. This suggests that using persistency is a promising way to estimate the variability in the project completion time. Recall that in our approximation model, the variance is solely determined by the persistency values, i.e., β_j's in Equation(2.3.1). Adding the quadratic terms not only helps capture the right direction of skewness, but more interestingly, it also significantly improves the estimation on variance. The added variability comes from the quadratic components of the estimator, as the linear term in the least squares quadratic estimation shares the same coefficients as the least squares linear estimator, i.e., persistency. Overall, the least squares linear approximation is remarkably effective with extremely low ESD and SND, and the least squares quadratic approximation even pushes the SND below the numerical integration approach. Figure 2.7.2 plots the density and cumulative distribution functions of PERT and our least squares estimations together with the simulation results. It is obvious from the plots that both least square estimators fit closely with simulation results. With the right skewness direction, the cumulative distribution function of the quadratic estimator almost overlaps with that of simulation.

To demonstrate the impact of having correlated activities, let us consider a simple variation of Example 2.11.

Example 2.12. Consider the same project network as in Example 2.11 with the same

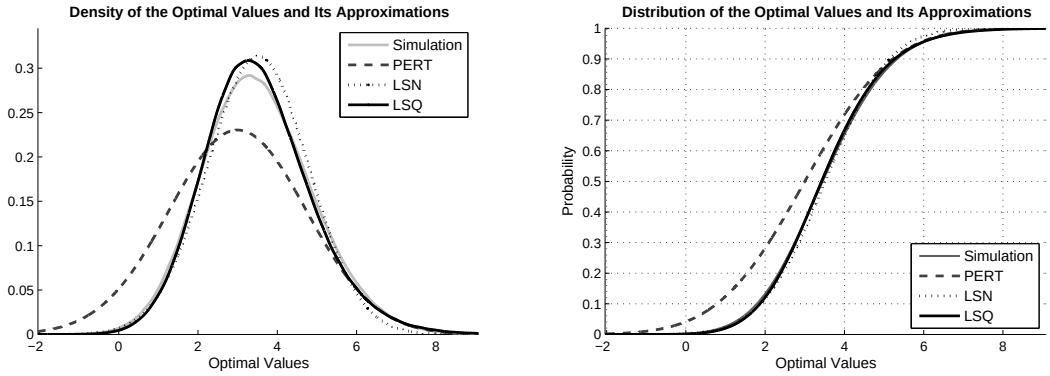


Figure 2.7.2: Distributions for Example 2.11

mean and variance for each activity, but the activities are correlated with each other. The correlation matrix is randomly generated using MATLAB function "randcorr".

Note that all the previous methods except PERT give the same approximations since they ignore correlations among activities. The distribution approximated by PERT is simply the distribution of the critical path, i.e., Path $1 \to 2 \to 3 \to 4$.

Table 2.7.2 summarizes the results for nine instances of Example 2.12[4]. "SIMU" stands for "simulation", and the sample size for all the simulations is 10^6. Same as before, the results from estimated persistency values are also included in the table to save space but will only be discussed in the next subsection. First of all, it is observed that both mean and variance of the project completion time can vary a lot from the independent case, especially the variance. Similar to Example 2.11, the variance estimates from both least square approximations cope nicely with the true variances. The least squares quadratic estimator constantly produces standard deviation estimates with less than 3% error. The skewness estimates from the least squares quadratic estimator tend to be larger than the true values, but the direction is correct, which helps improve ESD and SND substantially. The ESD of the least squares linear estimator is

[4]We have conducted this experiment for hundreds of random correlation matrices, and the findings are the same as those discussed here. Thus, we only report nine instances here for succinctness.

	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND
SIMU	3.581	1.55	-	0.61	-	-	3.643	1.38	-	0.40	-	-	3.675	1.36	-	0.41	-	-
PERT	3.000	2.08	33.90%	0	1.307	0.683	3.000	1.82	31.49%	0	1.398	0.624	3.000	1.81	33.30%	0	1.510	0.682
LSN	3.581	1.43	7.95%	0	0.367	0.062	3.643	1.22	11.46%	0	0.412	0.040	3.675	1.18	13.06%	0	0.450	0.047
LSQ	3.581	1.53	1.58%	0.83	0.075	0.010	3.643	1.35	2.64%	0.66	0.099	0.008	3.676	1.32	2.86%	0.72	0.107	0.011
LSNe	3.597	1.38	11.32%	0	0.374	0.076	3.644	1.20	13.06%	0	0.414	0.047	3.680	1.13	16.93%	0	0.455	0.067
LSQe	3.597	1.56	0.75%	0.97	0.159	0.019	3.644	1.44	4.37%	0.78	0.168	0.020	3.681	1.41	4.01%	0.89	0.207	0.025
SIMU	3.578	1.33	-	0.47	-	-	3.512	1.44	-	0.30	-	-	3.513	1.32	-	0.20	-	-
PERT	3.000	1.78	33.88%	0	1.224	0.566	3.000	1.80	25.16%	0	0.975	0.406	3.000	1.53	15.76%	0	0.901	0.312
LSN	3.578	1.18	10.76%	0	0.360	0.040	3.512	1.33	7.59%	0	0.303	0.022	3.514	1.17	11.61%	0	0.384	0.027
LSQ	3.578	1.30	2.41%	0.76	0.084	0.009	3.512	1.41	1.83%	0.48	0.075	0.005	3.514	1.30	2.20%	0.48	0.077	0.008
LSNe	3.581	1.14	13.76%	0	0.363	0.052	3.514	1.31	8.87%	0	0.304	0.026	3.527	1.13	14.36%	0	0.393	0.040
LSQe	3.582	1.38	3.67%	0.92	0.157	0.021	3.513	1.47	2.27%	0.58	0.120	0.011	3.527	1.36	2.87%	0.71	0.156	0.025
SIMU	3.509	1.43	-	0.34	-	-	3.566	1.61	-	0.29	-	-	3.637	1.47	-	0.41	-	-
PERT	3.000	1.81	26.22%	0	0.969	0.416	3.000	1.99	23.37%	0	1.108	0.477	3.000	1.92	31.04%	0	1.411	0.639
LSN	3.509	1.32	7.60%	0	0.300	0.024	3.565	1.49	7.46%	0	0.373	0.026	3.637	1.32	10.22%	0	0.416	0.041
LSQ	3.509	1.41	1.82%	0.52	0.074	0.005	3.565	1.58	1.79%	0.47	0.089	0.006	3.637	1.43	2.27%	0.65	0.097	0.008
LSNe	3.511	1.28	10.45%	0	0.305	0.034	3.568	1.44	10.68%	0	0.379	0.040	3.641	1.27	13.01%	0	0.419	0.054
LSQe	3.511	1.44	0.20%	0.58	0.124	0.007	3.568	1.65	2.09%	0.54	0.194	0.011	3.641	1.51	2.99%	0.80	0.179	0.019

(Simulation results for independent case: $\mathbf{E}[T] = 3.516$, $\sigma(T) = 1.39$, $sk(T) = 0.28$)

Table 2.7.2: Estimation results for Example 2.12

usually less than 30% of that from PERT, and the least squares quadratic estimator offers a further 70% reduction. The improvement on SND is even larger.

For the example problems we studied above, the skewness in the optimum distribution is not very strong. In order to better demonstrate the impact of the quadratic estimator, we study a simple problem discussed by Zhan et al. (2005) in the next example.

Example 2.13. Approximate the distribution of the maximum of two normal random variables, $N(0, 0.5^2)$ and $N(1, 3^2)$. In this case, the persistency values can be accurately obtained from integration.

The results are plotted in Figure 2.7.3, and the improvement from the quadratic estimator is obvious. We can conclude that the advantage of adding quadratic terms is larger if the true distribution is suspect to be very skewed.

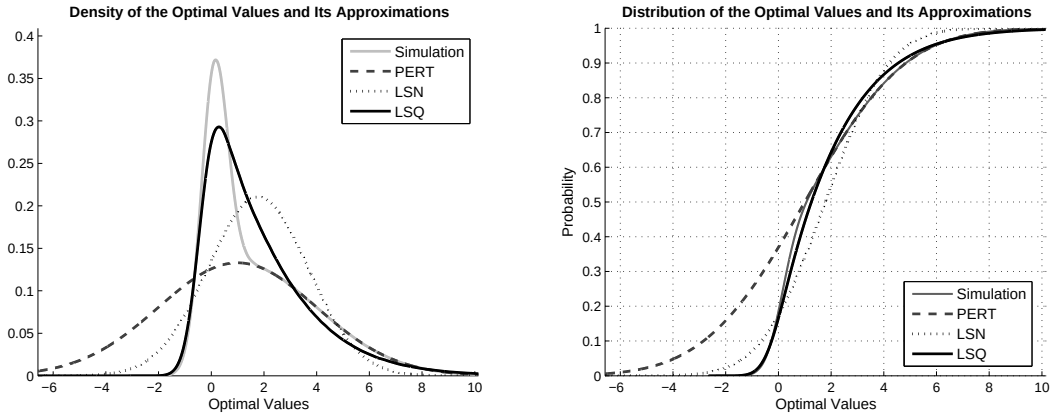


Figure 2.7.3: Distributions for Example 2.13

2.7.3 With Estimated Persistency Values

Consider Example 2.11 again, and we will construct our least squares approximating distributions using estimated persistency values. As discussed above, we implement

the estimation scheme from Clark (1961) to estimate the mean project completion time, i.e., the parameter α in our models. For persistency estimates, we solve the SDP relaxation of CPCMM reviewed in Section 2.6. The results are summarized in Table 2.7.1 on Page 67, where we add a lower case letter "e" after "LSN" and "LSQ" to indicate that the results are from estimated persistency parameters.

From the table, we can see that when estimated parameters are used instead of the exact ones, the distributions constructed from our least squares method still perform very well. For the least squares linear approximation, the estimated variance only deteriorates a little bit, which highlights the accuracy of persistency estimates from CPCMM and the power of using persistency in distribution approximation. Although the estimation error on $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$ has some impact on the least squares quadratic approximation, it still improves the performance from the least squares linear approximation. In particular, the variability estimate still outperforms the numerical integration approach, and the SND is below the numerical integration approach and much better than any other existing methods.

Next, we examine the approximating distributions with estimated persistency values using Example 2.12. New results are included in Table 2.7.2 on Page 70. Both least squares approximations with estimated persistency parameters perform as well as their counterparts with exact persistency parameters, and completely dominate PERT. From ESD and SND, the least squares quadratic approximation suffers a bit more from estimation errors since there are more parameters to estimate to construct the distribution. Nevertheless, the least squares linear approximation offers a very robust performance, given the fact that only n persistency values need to be estimated.

Now, let us revisit Example 2.13 with severe skewness in the optimum distribution. We present the approximation results with both exact and estimated persistency values

in Table 2.7.3. Although the estimation errors in the persistency parameters cause the variance estimate from the least squares quadratic estimator to become worse, the right skewness characterization still makes it better than the least squares linear estimation in terms of lowest ESD and SND. To highlight the impact of quadratic estimators, we also report the results from Clark's approximation method, which computes the exact moments including mean and variance in this simple setting. However, the failure to capture skewness makes Clark's method even worse than the least squares linear approximation in terms of higher SND. This also gives prominence to the advantage of our least squares approximation method over the traditional moment matching methods, since Clark's method in this case represents the perfect moment matching methods, i.e., zero estimation errors in moments.

Approximation Method	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND
10^6 simulation	1.777	2.09	-	1.13	-	-
PERT	1.000	3.00	43.54%	0	2.655	2.091
Clark (1961)	1.778	2.09	0.00%	0	-	0.463
LSN	1.778	1.90	9.21%	0	0.776	0.453
LSQ	1.779	2.07	1.23%	1.49	0.114	0.057
LSNe	1.777	1.98	5.48%	0	0.782	0.445
LSQe	1.777	2.23	6.50%	1.72	0.166	0.115

Table 2.7.3: Estimation results for Example 2.13

Before we end this section, we present a final example on STA considering correlations between delays. The example is taken from Tsukiyama et al. (2001). As we shall see later, the input values for this example are quite extreme. However, our least squares approximation method can still provide accurate estimations, which also demonstrate the robustness of the new approach.

Example 2.14. Consider the digital circuit and its network representation as shown in Figure 2.7.4. All the delay times (i.e., arc lengths) follow normal distributions.

Arc $(1,3)$ and $(1,4)$ have distributions of $N(20, 1.4^2)$, while Arc $(2,3)$, $(2,4)$, $(3,5)$, and $(4,5)$ follow $N(10, 0.7^2)$. Arc $(1,2)$ is normally distributed with mean μ_{12} and standard deviation σ_{12}. Correlation exists only between Arc $(3,5)$ and Arc $(4,5)$, and their correlation coefficient is denoted as ρ.

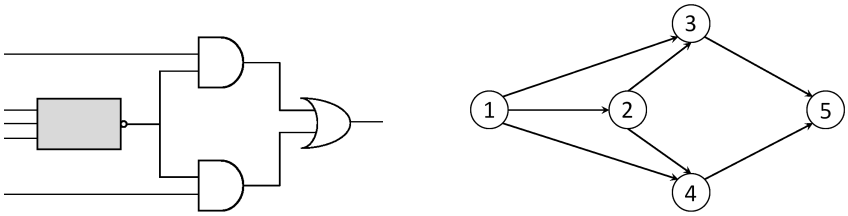


Figure 2.7.4: The digital circuit and its network representation in Example 2.14

For this example, we directly use the mean delay estimates from Tsukiyama et al. (2001) as the value of α for our least squares approximations. Tsukiyama et al. (2001) studied four scenarios of different input parameter values, i.e., μ_{12}, σ_{12}, and ρ. We report our approximation results under different scenarios in Table 2.7.4.

As mentioned above, the numerical values are extreme in this example. The coefficients of variation for all the delay times are only 0.07. Furthermore, with the normality assumption on delay times, PERT is expected to provide excellent approximation results, and the impact of correlations is minimal and hence difficult to capture. From Table 2.7.4, we find that despite the extreme inputs, our least squares quadratic approximation still gives the best estimate on the variance of circuit delay. Overall, both least squares approximations perform much better than PERT in terms of lower ESD and SND. Although the numerical errors cause the least squares quadratic approximation to have a higher SND than the least squares linear approximation, the least squares quadratic approximation still provides the lowest ESD.

Approximation Method	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND
$\mu_{12} = 10$, $\sigma_{12} = 0.7$, $\rho = 0.122$						
10^6 simulation	31.32	1.08	-	0.26	-	-
PERT	30.00	1.21	12.28%	0	3.220	1.769
LSNe	31.34	0.86	20.74%	0	0.432	0.054
LSQe	31.34	1.18	8.93%	0.62	0.162	0.018
$\mu_{12} = 20$, $\sigma_{12} = 1.4$, $\rho = 0.800$						
10^6 simulation	40.43	1.66	-	0.01	-	-
PERT	40.00	1.71	3.43%	0	0.585	0.191
LSNe	40.56	1.60	3.43%	0	0.124	0.020
LSQe	40.56	1.66	0.06%	0.26	0.073	0.026
$\mu_{12} = 40$, $\sigma_{12} = 2.8$, $\rho = 0.941$						
10^6 simulation	60.41	2.94	-	0.00	-	-
PERT	60.00	2.97	0.96%	0	0.520	0.171
LSNe	60.57	2.91	0.96%	0	0.118	0.024
LSQe	60.57	2.94	0.07%	0.17	0.062	0.038
$\mu_{12} = 80$, $\sigma_{12} = 5.6$, $\rho = 0.980$						
10^6 simulation	100.40	5.67	-	0.00	-	-
PERT	100.00	5.69	0.27%	0	0.500	0.158
LSNe	100.60	5.64	0.56%	0	0.132	0.041
LSQe	100.60	5.66	0.26%	0.18	0.122	0.097

Table 2.7.4: Estimation results for Example 2.14 with estimated parameters for least squares approximating distributions

2.8 Conclusion

In this chapter, we show that the distribution approximation problem under least squares framework and normality assumption can be transformed into the related persistency problem. Extensive computational experiments are presented to demonstrate the advantages of our approximation method, especially the benefits of introducing persistency into the distribution approximation problem. Better estimation on persistency values is then becoming critical and hence worth more exploration, especially under the normality assumption.

Chapter 3

Quadratic Regret Strategy

– Application in Portfolio Selection Problem

ABSTRACT. We propose a new multiple-benchmark tracking-error model for portfolio selection problem. The tracking error of a portfolio from a set of benchmark portfolios is defined as the difference between its return and the highest return from the set of benchmarks. We derive closed-form solution of our portfolio strategy, whose main component is the sum of the benchmark portfolios weighted by their respective probabilities of attaining the highest return among the portfolios in the benchmark. These probabilities, also known as the persistency values, are less sensitive to estimation errors in the means and covariances. These features help to stabilize the computational performance of our portfolio strategy against estimation errors.

We use the proposed model to address several pertinent issues in active portfolio management: (1) What are the benefits in tracking performance of multiple benchmarks? We demonstrate that under suitable conditions, multiple-benchmark tracking-error model can actually produce portfolio strategy that has less variability in portfolio returns, compared to the portfolio strategy constructed using single benchmark model, given a fixed target rate of returns. This addresses the agency issue in the investmen-

t problem, as portfolio managers are more concerned with variability of the excess returns above the benchmark, whereas the investors are more concerned with the variability of the total returns. (2) How and when to rebalance the portfolio allocation when prices and asset returns change over time, taking into account transaction cost? We show that our model can control for transaction cost by adding the buy-and-hold strategy into the set of benchmark portfolios. This approach reduces drastically the transaction volume of several popular static portfolio rules executed dynamically over time.

Last but not least, we perform comprehensive numerical experiments with various empirical data sets to demonstrate that our portfolio can consistently provide higher net Sharpe ratio (after accounting for transaction cost), higher net aggregate return, and lower turnover rate, compared to ten different benchmark portfolios proposed in the literature, including the equally weighted portfolio (the $1/n$ strategy).

3.1 Problem Overview

In practice, an institutional investor often evaluates the performance of a portfolio manager against a benchmark (e.g., an index fund). In active portfolio management, the portfolio manager makes specific investment with the goal of outperforming the benchmark, as his fees and compensations are directly linked to the excess returns above the benchmark. For a given target rate of returns, the portfolio manager would often seek to minimize the volatility of the deviation of the portfolio return from the benchmark return, i.e., the tracking-error volatility. The portfolio selection models that minimize the tracking-error volatility are referred to as benchmark tracking-error models (cf. Roll (1992) and Jorion (2003)).

Roll (1992) investigated the benchmark tracking-error model that minimizes the tracking-error volatility subject to the full investment constraint and the constraint on target expected return, i.e.,

$$\min_{\boldsymbol{e}^T\boldsymbol{x}=1,\boldsymbol{\mu}^T\boldsymbol{x}=K} \mathbf{E}\left[\left(\tilde{\boldsymbol{r}}^T\boldsymbol{p}-\tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right].$$

where $\tilde{\boldsymbol{r}} \in \mathbb{R}^n$ is the random return vector of the financial assets; $\boldsymbol{\mu}$ is the expect return of the assets, i.e., $\boldsymbol{\mu} = \mathbf{E}[\tilde{\boldsymbol{r}}]$; $\boldsymbol{e}$ is the column vector with all entries equal to 1; K is the target expected return; $\boldsymbol{p}$ is the benchmark portfolio; and $\boldsymbol{x} \in \mathbb{R}^n$ is the investment decision.

The deficiency with this tracking-error approach is however well-known. Roll (1992) observed that the optimal trading decision $(\boldsymbol{x}-\boldsymbol{p})$ does not depend on the benchmark at all. Furthermore, with this setup the portfolio manager will focus solely on the tracking-error volatility but ignore the total portfolio risk. This creates an agency problem, since the investor is more concerned with the latter. The tracking-error model may thus produce seriously inefficient portfolios for the investor. To address these issues, Roll (1992) proposed to constrain the portfolio's beta; Jorion (2003) proposed to constrain the portfolio's total variance; and Alexander & Baptista (2008) proposed to constrain the portfolio's Value-at-Risk (VaR). However, most of these proposals are difficult to implement in practice, and do not address directly the connection with the benchmark based approach to portfolio management.

The choice of the proper benchmark is also a problem in practice. Poor active portfolio management could lead to less than perfectly diversified portfolio, and incur heavy transaction costs and assumes high total portfolio risk. El-Hassan & Kofman (2003) observed from their empirical analysis that in reality, the selected benchmark is often inefficient, and its expected return could fall below the expected return of

the well-known minimum-variance portfolio. The immediate consequence is that during bear market conditions the benchmark tracking-error models will call for a huge amount of short selling, which can substantially increase the total portfolio risk. This problem is compounded by the fact that tracking-error measurement does not differentiate between over-performing and under-performing vis-à-viz the benchmark portfolio, and hence the performance of the tracking-error model can be adversely affected by a poorly selected benchmark.

To mitigate this problem of finding the right unique benchmark, one natural solution is to use multiple benchmarks to evaluate the performance of a portfolio manager. By choosing benchmarks that can counter-balance the performance of each other in different market environments, we can track the performance of our portfolio strategy in a more accurate and reliable manner. The literature on multiple-benchmark tracking-error strategy is however comparatively sparse. Wang (1999) extended the single-benchmark tracking-error model to track multiple benchmarks simultaneously. The tracking error of the portfolio with respect to a set of m benchmarks is defined by a weighted sum of single-benchmark tracking errors, i.e.,

$$\min_{\boldsymbol{e}^T\boldsymbol{x}=1,\boldsymbol{\mu}^T\boldsymbol{x}=K} \sum_{j=1}^{m} w_j \mathbf{E}\left[\left(\tilde{\boldsymbol{r}}^T\boldsymbol{p}^j - \tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right],$$

where w_j is the weight on the tracking error of the j^{th} benchmark portfolio, $\boldsymbol{p}^j$, $j = 1, \ldots, m$. Rustem & Howe (2002) considered an alternative model. Their objective is to minimize the maximum tracking-error volatility across all benchmarks, i.e.,

$$\min_{\boldsymbol{e}^T\boldsymbol{x}=1,\boldsymbol{\mu}^T\boldsymbol{x}=K} \max_{j\in\{1,\ldots,m\}} \mathbf{E}\left[\left(\tilde{\boldsymbol{r}}^T\boldsymbol{p}^j - \tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right].$$

However, it is not clear how a portfolio manager should choose the weights in Wang's

model. The minimax approach, on the other hand, is often considered to be too conservative. A more critical issue is that these models still rely on the evaluation of single-benchmark tracking error and only combine them in the aggregate level. They fail to distinguish between over-performing and under-performing vis-à-viz the selected benchmarks, and that the performance of the different benchmarks may be correlated. Hence they do not fully capture the concerns arising from the real investment activities as discussed above.

In this chapter, we propose a new class of tracking-error models for multiple benchmarks. This problem arises naturally when multiple natural benchmarks (e.g., risk-free returns, S&P 500 index etc.) are readily available in the market that can be used to evaluate the performance of the portfolio managers. It also arises when the portfolio manager is managing funds for different clients, each with a unique benchmark that will be used to evaluate the performance of the manager. Instead of managing different pools of funds, one for each client, we explore the possibility of pooling the funds and benchmarks together to derive a better portfolio strategy. Our target performance is to match the highest return among all the benchmarks, i.e.,

$$Z_B(\tilde{\boldsymbol{r}}) := \max_{j\in\{1,\ldots,m\}} \tilde{\boldsymbol{r}}^T \boldsymbol{p}^j.$$

Note that since the asset returns are random, the highest benchmark return is also random, and it depends on the realization of the asset returns. That is why we use the notation $Z_B(\tilde{\boldsymbol{r}})$ to represent the highest benchmark return. Our multiple-benchmark tracking error is defined as the difference between the portfolio return and the highest return induced from the benchmark portfolios, i.e., $\tilde{\boldsymbol{r}}^T\boldsymbol{x} - Z_B(\tilde{\boldsymbol{r}})$. We are interested

in finding a portfolio $\boldsymbol{x}$ whose tracking-error volatility is minimized, i.e.,

$$\text{(T)} \qquad \min_{\boldsymbol{x}\in\mathcal{X}} \mathbf{E}\left[\left(Z_B(\tilde{\boldsymbol{r}}) - \tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right],$$

where $\mathcal{X}$ is a set of feasible portfolios. The constraints in the set $\mathcal{X}$ includes the full investment constraint $\boldsymbol{e}^T\boldsymbol{x} = 1$. It is possible to capture additional constraints on the portfolio vector in the set $\mathcal{X}$, e.g., target expected return constraint, short-sale constraints, etc. Note that when there is only one benchmark portfolio, this reduces to the single-benchmark tracking-error model of Roll (1992).

Intuitively, as the financial asset returns are very volatile, it is almost impossible for a single benchmark to consistently perform well in every situation. Tracking the best return from a set of benchmarks appears to be a more attractive and practical objective, as it addresses partially the concern of a particularly bad benchmark dragging down the performance of the portfolio. The investor can also control the aggressiveness of the active investment by choosing an appropriate pool of benchmarks that suit the style and risk profile of the investor. Surprisingly, this approach can also be used to address the agency issue concerning the conflicting objectives between the investor and portfolio manager - the portfolio constructed using the multiple-benchmark tracking-error model may actually resulted in lower total returns variability, compared to the single benchmark approach.

Our definition of multiple-benchmark tracking error is similar to the definition of external regret in the on-line portfolio selection and machine learning problems. In those problems, the goal is to minimize the predictor's cumulative loss with respect to the best cumulative loss in a pool of "experts", and the external regret measures the difference between the predictor's cumulative loss and that of the best expert (cf.Stoltz & Lugosi (2005) and references therein). There are several key differences

between the on-line portfolio selection problem and our problem. First, our problem happens in a single period, but the on-line portfolio selection deals with multi-period investment problem. Consequently, the objective for the on-line portfolio selection problem is usually to maximize the wealth growth over a given number of periods, and most research work in literature focused on providing bounds on the external regret. In the on-line portfolio selection problem, the distribution of asset return is unknown and must be gradually learned from investment activities, but in our problem, we assume that all the distributional information is known before making the investment decision. Last but not least, even though the on-line portfolio selection problem is a multi-period investment problem, the impact of transaction costs was largely ignored in this substream of literature. On the other hand, we pay a lot of attention on controlling for transaction costs when analyzing our model, and also build an extension that can explicitly penalize high transaction volumes. Although not purposely designed for the multi-period problem, our model still demonstrates superior performance when it is used in multiple periods and the estimation errors are properly controlled by the unique feature of our model.

Remark 3.1. Note that Problem (T) can be interpreted as a quadratic regret minimization model, where the regret is measured as the performance deviation from the maximum benchmark return. Mathematically, Problem (T) is very similar to the distribution approximation problem discussed in Chapter 2. In this case, we want to find a distribution given as a linear function of the random vector $\tilde{\boldsymbol{r}}$ to approximate the distribution of $Z_B(\tilde{\boldsymbol{r}})$, the outcome of an optimization system depending on $\tilde{\boldsymbol{r}}$. The linear coefficients, $\boldsymbol{x}$, are obtained from minimizing the expected squared deviation of the two distributions subject to some constraints. The main difference is the lack of constant term in Problem (T), which makes the solution of $\boldsymbol{x}$ different from the persis-

tency under the normality assumption. However, as we shall see later, persistency still plays an important role in the optimal portfolio position. This not only leads to the robustness of our portfolio, but also provides some insights to the regret minimization model and probability matching behaviour, which we will illustrate further in Section 3.2.3 and 4.1.2.

Our main contributions in the chapter are as follows:

1. Under the assumption of a normally distributed return vector, we derive the closed-form solution of our portfolio model without short-sale constraints, and show that the optimal multiple-benchmark tracking-error portfolio relies on the probabilities that the benchmarks attain the highest return. This helps to stabilize the performance of our portfolio strategy in numerical experiments, as those probabilities are generally less prone to estimation errors on means and covariances.

2. Using two suitably chosen benchmarks, we prove that one can generate the entire mean-variance efficient frontier using our model. This result is similar to the well-known Two-Fund Theorem in classical portfolio theory.

3. We also compare the performance of our multiple-benchmark tracking-error model with the traditional single-benchmark tracking-error model, for fixed target expected return K. While the portfolio manager focuses on minimizing the variability of the excess return against the benchmark(s), we show that the total portfolio variance can be lower in the multiple-benchmark environment. We identify the environments under which the multiple-benchmark portfolio strategy will dominate the single-benchmark approach in terms of lower total portfolio variance at all levels of target expected return. This result exploits the fact

that the variance of the returns of a linear combination of portfolio rules can be smaller than the variance of the returns of each individual portfolio rule.

4. We also show that the portfolio strategy constructed using our multiple-benchmark tracking-error model will be preferred over using simple linear combination of the benchmark portfolios, in the environment when the portfolio managers have mean-variance utility functions with low risk aversion parameters.

5. More importantly, we show that our portfolio strategy performs well even with estimation errors and when transaction costs are properly accounted for [1]. We show that our model can be extended to penalize for transaction volumes. Alternately, we can also simply incorporate the buy-and-hold strategy into the set of benchmarks to reduce the transaction volumes. We show via extensive numerical experiments that this approach can significantly reduce transaction costs while not sacrificing the performance on returns. For instance, in the multi-period empirical tests, when we combine the partial minimum-variance (PARR) portfolio proposed by DeMiguel et al. (2009) with the buy-and-hold strategy as two benchmarks, our multiple-benchmark tracking-error portfolio incurs turnover rates that are less than half of those from the PARR portfolio. In terms of out-of-sample Sharpe ratio net of 50 basis point, our portfolio is significantly higher than the PARR portfolio. Our strategy also beats the equally weighted investment strategy (also known as the $1/n$ strategy) comprehensively when transaction costs are properly accounted for.

Outline of the chapter: In the next section, we solve our multiple-benchmark

[1]The most common approach in existing literature is to include either a penalty term in the objective function or a budget constraint in the portfolio models. For example, Brodie et al. (2009) proposed to add an additional penalty term to the classical Markowitz mean-variance framework, where the penalty is proportional to the sum of the absolute values of the portfolio weights.

tracking-error model and analyze the properties of its solution. We present and discuss the results of the numerical studies in Section 3.3 with a focus on including the buy-and-hold strategy as a benchmark to beat the other benchmark portfolios, especially the equally weighted portfolio. Finally, we provide some concluding remarks in Section 3.4.

3.2 Multiple-Benchmark Tracking-Error Portfolio

In this section, we will derive the solution to our model, i.e., Problem (T), and analytically investigate its features. Especially, we will compare our portfolio with the well-known Markowitz mean-variance efficient portfolio and the linear combination rule proposed by Tu & Zhou (2011). An extension of out model to penalize transaction cost is presented in the final part of this section.

To derive the closed-form solution of Problem (T), we first simplify the problem by linking it to the concept of persistency and Stein's identity.

3.2.1 Persistency and Stein's Identity

In the context of the benchmark tracking problem, the persistency of a benchmark portfolio refers to the probability that this portfolio outperforms the rest benchmark portfolios. Formally, we present the definition next.

Definition 3.2. Define the m dimensional random vector

$$\boldsymbol{p}(\tilde{\boldsymbol{r}}) = \left(\mathbb{I}_{Z_B(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^T\boldsymbol{p}^1}, \ldots, \mathbb{I}_{Z_B(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^T\boldsymbol{p}^m}\right)^T,$$

where the indicator function $\mathbb{I}_{Z_B(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^T\boldsymbol{p}^j}$ takes a value of 1 if the j^{th} benchmark portfolio produces the highest return in the set of benchmark portfolios and 0 otherwise.

The persistency vector is an m dimensional vector whose j^{th} component is the probability that the j^{th} benchmark portfolio is the best performing portfolio in the set of benchmark portfolios, i.e.,

$$\mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})] = \left(\mathbf{P}\left(Z_B(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^T\boldsymbol{p}^1\right), \ldots, \mathbf{P}\left(Z_B(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^T\boldsymbol{p}^m\right)\right)^T.$$

Define the $n \times m$ benchmark portfolio matrix $P = [\boldsymbol{p}^1, \ldots, \boldsymbol{p}^m]$. The persistency weighted benchmark portfolio is defined as the n dimensional vector $P\mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})]$.

Remark 3.3. In this chapter, we assume that $\tilde{\boldsymbol{r}}$ is a nondegenerate multivariate continuous random vector with a positive definite covariance matrix. The support of $\tilde{\boldsymbol{r}}$ over which more than one benchmark attains the maximum return has measure zero. Then $\boldsymbol{p}(\boldsymbol{r})$ is unique almost surely, and $\mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})]$ satisfies $\sum_{j=1}^{m}\mathbf{E}[p_j(\tilde{\boldsymbol{r}})] = 1$.

As we will see later, the solution to the multiple-benchmark tracking-error minimization problem is related to the persistency when the return follows a multivariate normal distribution, i.e.,

(A) The random return vector, $\tilde{\boldsymbol{r}}$, follows a multivariate normal distribution with a finite mean, $\boldsymbol{\mu} \neq \mathbf{0}$[2], and a finite positive definite covariance matrix, Σ, denoted as $\tilde{\boldsymbol{r}} \sim N(\boldsymbol{\mu}, \Sigma)$.

This result is established by appealing to Stein's Identity in probability theory, as discussed in Chapter 1.

3.2.2 Tracking-Error Minimization

Intuitively, if we treat $Z_B(\tilde{\boldsymbol{r}})$ as a function on $\tilde{\boldsymbol{r}}$, we can apply Stein's Identity to derive the covariance between the individual asset return and the highest benchmark return.

[2]Note that the assumption of $\boldsymbol{\mu} \neq \mathbf{0}$ is required only for the model analysis, especially on efficient frontiers. For our basic model, we can still obtain the solution when $\boldsymbol{\mu} = \mathbf{0}$.

In particular, applying Stein's Identity to $Cov(Z_B(\tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}}^T)$, we get

$$Cov(Z_B(\tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}}^T) = (\Sigma P \mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})])^T.$$

Then our problem can be simplified as shown in the next proposition.

Proposition. *Under Assumption (A), the multiple-benchmark tracking-error portfolio in Problem (T) can be found by solving the following convex quadratic minimization problem:*

$$(T') \qquad \min_{\boldsymbol{x}\in\mathcal{X}} \ \boldsymbol{x}^T\big(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T\big)\boldsymbol{x} - 2\Big(\Sigma P \mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})] + \mathbf{E}[Z_B(\tilde{\boldsymbol{r}})]\boldsymbol{\mu}\Big)^T\boldsymbol{x}.$$

Proof. Expanding the expectation term in (T), we get the equivalent formulation as follows:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \ \boldsymbol{x}^T\big(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T\big)\boldsymbol{x} - 2\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\tilde{\boldsymbol{r}}^T\right]\boldsymbol{x} + \mathbf{E}\left[\left(Z_B(\tilde{\boldsymbol{r}})\right)^2\right].$$

Since the last term is independent of $\boldsymbol{x}$, we can exclude it from the minimization problem. Note that

$$\mathbf{E}[Z_B(\tilde{\boldsymbol{r}})\tilde{\boldsymbol{r}}^T] = Cov(Z_B(\tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}}^T) + \mathbf{E}[Z_B(\tilde{\boldsymbol{r}})]\mathbf{E}[\tilde{\boldsymbol{r}}^T] = Cov(Z_B(\tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}}^T) + \mathbf{E}[Z_B(\tilde{\boldsymbol{r}})]\boldsymbol{\mu}^T.$$

Using differentiation by parts, we get

$$\begin{aligned}
\mathbf{E}\left[\frac{\partial Z_B(\tilde{\boldsymbol{r}})}{\partial \tilde{r}_l}\right] &= \mathbf{E}\left[\frac{\partial}{\partial \tilde{r}_l}\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{r}_i P_{i,j} p_j(\tilde{\boldsymbol{r}})\right)\right] \\
&= \mathbf{E}\left[\sum_{j=1}^{m} P_{l,j} p_j(\tilde{\boldsymbol{r}}) + \sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{r}_i P_{i,j}\frac{\partial p_j(\tilde{\boldsymbol{r}})}{\partial \tilde{r}_l}\right] \\
&= P_{(l)}\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right],
\end{aligned}$$

where $P_{(l)}$ denotes the l^{th} row of P, and $P_{i,j} = p_i^j$, $\forall i = 1, \ldots, n$, $\forall j = 1, \ldots, m$. The

last equality follows from our assumption on $\tilde{\boldsymbol{r}}$ so that $\partial p_j(\tilde{\boldsymbol{r}})/\partial \tilde{r}_l$ exists almost everywhere and equals zero wherever it exists. Applying Stein's Identity to $Cov(Z_B(\tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}}^T)$, we get

$$Cov(Z_B(\tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}}^T) = (\Sigma P \mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})])^T,$$

and thus Problem (T′). □

Remark 3.4. Suppose that the random vector $\tilde{\boldsymbol{r}}$ is not normally distributed. It is still possible to find the multiple-benchmark tracking-error portfolio by solving the following convex quadratic programming problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \ \boldsymbol{x}^T\left(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T\right)\boldsymbol{x} - 2\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\tilde{\boldsymbol{r}}^T\right]\boldsymbol{x}.$$

This requires the estimation of $\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\tilde{\boldsymbol{r}}^T\right]$. The advantage of resorting to Stein's Identity for the multivariate normal distribution is twofold. First, by using Stein's Identity, we need to estimate (a) The persistency vector $\mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})]$, and (b) The expectation of the highest benchmark return $\mathbf{E}[Z_B(\tilde{\boldsymbol{r}})]$. Estimation of the benchmark portfolio that different experts believe will outperform the rest is inherently easier to elicit from managers. Second, the transformed problem provides a simple characterization with a closed-form solution that allows for more in-depth analysis of the model. We elaborate on this issue in the next several subsections.

By re-writing the expression in Proposition 3.2.2, we can reinterpret our model as a variant of single-benchmark tracking-error model: Problem (T′) is equivalent to

$$\min_{\boldsymbol{x}\in\mathcal{X}} \ \left(\boldsymbol{x} - P\mathbf{E}\left[\boldsymbol{p}\left(\tilde{\boldsymbol{r}}\right)\right]\right)^T \boldsymbol{\Sigma}\left(\boldsymbol{x} - P\mathbf{E}\left[\boldsymbol{p}\left(\tilde{\boldsymbol{r}}\right)\right]\right) + \left(\mathbf{E}\left[Z_B\left(\tilde{\boldsymbol{r}}\right)\right] - \boldsymbol{\mu}^T\boldsymbol{x}\right)^2.$$

The first term is exactly the tracking-error volatility, measured against the persistency

weighted portfolio. The second term penalizes the shortfall of the portfolio return from the highest benchmark return. When there is a constraint that fixes the expected portfolio return, the second term will vanish in the minimization problem, and our model reduces to a single-benchmark tracking-error model with the persistency weighted benchmark portfolio as the only benchmark. In general, our model anchors in the persistency weighted benchmark portfolio, and it is adjusted to recover the loss in the expected portfolio return from the highest benchmark return. This result shows that Problem (T) is related to the single-benchmark tracking-error literature in the following ways:

- If we fixed a target expected return, Problem (T) reduces to a single-benchmark tracking-error minimization problem, where the benchmark tracked is $P\mathbf{E}\left[\boldsymbol{p}\left(\tilde{\boldsymbol{r}}\right)\right]$, the persistency strategy formed by the set of portfolio used as benchmarks.

- If we fixed a budget for the tracking-error volatility, i.e.,

$$\left(\boldsymbol{x} - P\mathbf{E}\left[\boldsymbol{p}\left(\tilde{\boldsymbol{r}}\right)\right]\right)^{T} \boldsymbol{\Sigma} \left(\boldsymbol{x} - P\mathbf{E}\left[\boldsymbol{p}\left(\tilde{\boldsymbol{r}}\right)\right]\right),$$

  then Problem (T) will find a strategy that has expected return as close as possible to the highest benchmark return. Our model therefore uses $\mathbf{E}\left[Z_B\left(\tilde{\boldsymbol{r}}\right)\right]$ to anchor the selection of the portfolio strategy in the tracking-error model, to avoid excessive risk, instead of limiting the total risk (variance of the returns), as commonly used. The selection of the benchmarks used in our model is thus crucial to the performance of the portfolio strategy.

3.2.3 Closed-Form Solution

In this subsection, we present the closed-form expression of the multiple-benchmark tracking-error portfolio when the return vector $\tilde{\boldsymbol{r}}$ satisfies the multivariate normality assumption. To simplify the expression, we introduce three constants, $A = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$, $B = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{e}$, and $C = \boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}$. These constants are also used to describe the closed-form expression of the Markowitz mean-variance portfolio (cf. Steinbach (2001)). Note that by Assumption (A), $A > 0$ and $C > 0$.

Theorem 3.5. *Define the set of feasible portfolios as* $\mathcal{X} = \left\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{e}^T\boldsymbol{x} = 1\right\}$. *Under Assumption (A), the optimal multiple-benchmark tracking-error portfolio in (T) is given by*

$$P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \Sigma^{-1}\left(\frac{C\boldsymbol{\mu} - B\boldsymbol{e}}{(A+1)C - B^2}\right)\left(\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] - \boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]\right). \tag{3.2.1}$$

Proof. Since Problem (T′) is strictly convex, the first-order optimality conditions are both necessary and sufficient. In particular, the Lagrangian of Problem (T′) is given by

$$\mathcal{L}(\boldsymbol{x}, \pi) = \boldsymbol{x}^T\left(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T\right)\boldsymbol{x} - 2\left(\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]\boldsymbol{\mu}\right)^T\boldsymbol{x} + 2\pi\left(1 - \boldsymbol{e}^T\boldsymbol{x}\right).$$

The first-order conditions yield

$$2\left(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T\right)\boldsymbol{x}^* - 2\left(\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]\boldsymbol{\mu}\right) - 2\pi\boldsymbol{e} = 0,$$

and

$$\sum_{i=1}^{n} x_i^* = 1.$$

Multiplying Σ^{-1} to both sides of the first set of equalities, we get

$$\begin{pmatrix} \Sigma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^T + I_n & -\Sigma^{-1}\boldsymbol{e} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^* \\ \pi \end{pmatrix} = \Sigma^{-1}\boldsymbol{\mu}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] + P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right], \tag{3.2.2}$$

where I_n denotes the identity matrix of dimension $n \times n$. Multiplying $\boldsymbol{\mu}^T$ to both sides, we have

$$\begin{pmatrix} \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}^T I_n & -\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{e} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^* \\ \pi \end{pmatrix} = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] + \boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right].$$

Making the substitution with A, B, and C and dividing both sides by $(A+1)$, we get

$$\begin{pmatrix} \boldsymbol{\mu}^T & -\dfrac{B}{A+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^* \\ \pi \end{pmatrix} = \frac{A}{A+1}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] + \frac{\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]}{A+1}.$$

Subtracting $\mathbf{E}[Z_B(\tilde{\boldsymbol{r}})]$ from both sides gives

$$\begin{pmatrix} \boldsymbol{\mu}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^* \\ \pi \end{pmatrix} - \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] = \frac{\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] + \pi B}{A+1}. \tag{3.2.3}$$

Back to Equation (3.2.2), we can rewrite it as follows:

$$\Sigma^{-1}\boldsymbol{\mu}\left[\begin{pmatrix} \boldsymbol{\mu}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^* \\ \pi \end{pmatrix} - \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]\right] + \begin{pmatrix} I_n & -\Sigma^{-1}\boldsymbol{e} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^* \\ \pi \end{pmatrix} = P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]. \tag{3.2.4}$$

Substituting Equation (3.2.3) into (3.2.4), we have

$$\left(\begin{array}{cc} I_n & \dfrac{\Sigma^{-1}\boldsymbol{\mu}B}{A+1} - \Sigma^{-1}\boldsymbol{e} \end{array} \right) \left(\begin{array}{c} \boldsymbol{x}^* \\ \pi \end{array} \right) = P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]}{A+1}\Sigma^{-1}\boldsymbol{\mu}. \tag{3.2.5}$$

Multiplying $\boldsymbol{e}^T$ to both sides of the above equation, we get

$$\boldsymbol{e}^T\boldsymbol{x}^* + \left(\frac{B^2}{A+1} - C\right)\pi = \boldsymbol{e}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{B}{A+1}\left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]\right).$$

Note that $\boldsymbol{e}^T\boldsymbol{x}^* = \boldsymbol{e}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] = 1$. Canceling these two terms from both side, we have

$$\pi = \frac{B}{(A+1)C - B^2}\left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]\right).$$

Substituting the above formula for π into Equation (3.2.5), we get

$$\boldsymbol{x}^* = P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]\right)\Sigma^{-1}\left(\frac{B\boldsymbol{e} - C\boldsymbol{\mu}}{(A+1)C - B^2}\right),$$

which is the closed-form solution as shown in the theorem. □

Define $\mu_p := \boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]$, the mean return of the persistency weighted portfolio. It is well-known that if the returns of the portfolios in the set of benchmarks are negatively correlated, then it is possible for the variance of the persistency portfolio to be smaller than the variance of the individual portfolio. Our strategy builds on the persistency portfolio, and adjusts for higher mean returns through the term,

$$\Sigma^{-1}\left(\frac{C\boldsymbol{\mu} - B\boldsymbol{e}}{(A+1)C - B^2}\right)\left(\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] - \boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]\right).$$

In this way, we can ensure that the mean returns of our strategy is at least as good as

the persistency weighted portfolio.

Proposition 3.6. *Under Assumption (A), the expected return of our multiple-benchmark tracking-error portfolio is not less than μ_p. In particular, when $\boldsymbol{\mu} \neq \boldsymbol{e}$, the portfolio has a strictly higher expected return than μ_p.*

Proof. From the closed-form solution, the expected return of our multiple-benchmark tracking-error portfolio is

$$\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \left(1 - \frac{C}{(A+1)C - B^2}\right)\left(\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] - \boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]\right).$$

Note that

$$\begin{aligned} & \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] - \boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] \\ = \; & \mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}}) - \tilde{\boldsymbol{r}}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]\right] \\ = \; & \mathbf{E}\left[\max_{j\in\{1,\ldots,m\}} \tilde{\boldsymbol{r}}^T \boldsymbol{p}^j - \sum_{j=1}^{m} \mathbf{E}\left[p_j(\tilde{\boldsymbol{r}})\right]\left(\tilde{\boldsymbol{r}}^T \boldsymbol{p}^j\right)\right] \\ \geq \; & 0. \end{aligned}$$

If

$$1 \geq \frac{C}{(A+1)C - B^2}, \tag{3.2.6}$$

then

$$\boldsymbol{\mu}^T \boldsymbol{x} \geq \boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] = \mu_p.$$

Now we shall show (3.2.6) holds. Let $\boldsymbol{\alpha} = \Sigma^{-1/2}\boldsymbol{e}$, and $\boldsymbol{\beta} = \Sigma^{-1/2}\boldsymbol{\mu}$. By Cauchy-Schwartz Inequality,

$$B^2 = \left(\boldsymbol{e}^T \Sigma^{-1} \boldsymbol{\mu}\right)^2 = \left(\boldsymbol{\alpha}^T \boldsymbol{\beta}\right)^2 \leq \left(\boldsymbol{\alpha}^T \boldsymbol{\alpha}\right)\left(\boldsymbol{\beta}^T \boldsymbol{\beta}\right) = \left(\boldsymbol{e}^T \Sigma^{-1} \boldsymbol{e}\right)\left(\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}\right) = AC.$$

We have

$$\frac{1}{1+\dfrac{AC-B^2}{C}} \leq 1,$$

i.e.,

$$\frac{C}{(A+1)C-B^2} \leq 1.$$

The equality holds if and only if $\boldsymbol{\mu}$ is proportional to $\boldsymbol{e}$, since Σ is positive definite and so is $\Sigma^{-1/2}$. □

As with all tracking-error models in the literature, the improvement on mean returns comes from an associated increase in risks. The variability of our portfolio returns will be higher than the persistency weighted portfolio. Fortunately, with properly selected benchmarks, the variability of the returns from the persistency portfolio can be lower than the variability of the returns from each individual benchmark. Hence, the performance of our portfolio strategy may still dominate the performance of some of the benchmarks used in our model, in terms of both mean and variance of the portfolio return. This partially explains the superior performance of our approach in numerical experiments conducted in Section 3.3.

Remark 3.7. One interesting observation from the closed-form expression of the optimal portfolio position is that it is anchored in the linear combination of the benchmark portfolios weighted by their persistency values, $\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]$. Although probability matching is usually observed in a multi-period setting with some mixed strategy, in a single period problem with continuous decision, probability matching could appear as a decision that consolidates the mixed strategy according to its mixing probabilities. In this sense, our model has some flavour of probability matching, where the investment decision tries to match the underlying probability of achieving the best possible benchmark return. Furthermore, Simon (1956) has shown that probability matching behaviour

"would be exhibited by a rational subject intent on minimaxing his regret". Our result tells that probability matching strategy is optimal if the decision maker tries to minimize his expected quadratic regret, which from another perspective reassures the finding from Simon (1956) that connects the behaviour with regret criterion. This in turn partially justifies that our multiple-benchmark tracking-error model may be a good model that describes investor's behaviour. For more discussion on the quadratic regret decision criterion and the probability matching behaviour, please refer to Chapter 4.

3.2.4 Comparison with the Markowitz Mean-Variance Portfolio

The pioneering work of modern portfolio theory by Markowitz (1952) quantified the relationship between the expected return and risk of portfolios, which is measured by the variance in portfolio returns. Markowitz introduced the notion of an efficient portfolio as the portfolio with minimal variance at a given level of expected return. The continuum of such portfolios forms an efficient frontier in the mean-variance space of the portfolios.

In this subsection, we exploit the advantage of the closed-form solution and compare our portfolio with the Markowitz mean-variance portfolio.

3.2.4.1 Optimal Portfolio Weights

Consider the Markowitz portfolio optimization model of the following form:

$$\min_{\boldsymbol{e}^T\boldsymbol{x}=1} \ \frac{1}{2}\boldsymbol{x}^T\Sigma\boldsymbol{x} - \gamma\boldsymbol{\mu}^T\boldsymbol{x},$$

where γ is the risk aversion parameter. The closed-form solution (cf. Steinbach (2001)) is

$$\Sigma^{-1}\left(\frac{(1-\gamma B)\boldsymbol{e}+\gamma C\boldsymbol{\mu}}{C}\right)=\frac{1}{C}\Sigma^{-1}\boldsymbol{e}-\frac{\gamma}{C}\Sigma^{-1}\left(B\boldsymbol{e}-C\boldsymbol{\mu}\right).$$

Rearranging the closed-form expression of the multiple-benchmark tracking-error portfolio in Theorem 3.5 helps to make the comparison more explicit as follows:

$$P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]+\frac{\boldsymbol{\mu}^{T}P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]-\mathbf{E}\left[Z_{B}(\tilde{\boldsymbol{r}})\right]}{(A+1)C-B^{2}}\Sigma^{-1}\left(B\boldsymbol{e}-C\boldsymbol{\mu}\right).$$

From these formulas, it is clear that both portfolios consist of two components: (a) a baseline portfolio, and (b) an adjustment term with a common factor, $\Sigma^{-1}\left(B\boldsymbol{e}-C\boldsymbol{\mu}\right)$. The baseline portfolio of the Markowitz mean-variance portfolio is the minimum-variance portfolio, $\Sigma^{-1}\boldsymbol{e}/C$, and the adjustment is related to the risk aversion parameter, γ. For the multiple-benchmark tracking-error portfolio, the baseline portfolio is the persistency weighted benchmark portfolio. The adjustment term accounts for the impact of the random return on the performance of the benchmark portfolios, in other words, the selection of the best performer as the target expected return. The multiple-benchmark tracking-error portfolio thus incorporates information on the relative performance of the competing portfolios in the market, which is absent in the information set while deciding the Markowitz mean-variance portfolio. Please refer Section 4.1.2 for more discussions on the interpretation of the adjustment term in the optimal multiple-benchmark tracking-error portfolio from the perspective of probability matching.

To provide some intuition on the difference between these two portfolios, consider a simple example of investment between a risk-free asset and a risky asset.

Example 3.8. Suppose an investor has to decide a portfolio among two uncorrelated

assets, one of which is risk-free with zero variance, and the other is risky with a variance of 1. Both assets have zero-mean returns. In this case, the investor who follows the Markowitz strategy will always choose the risk-free asset for any nonnegative risk aversion parameter. On the other hand, suppose we choose two extreme strategies as the benchmark portfolios – each strategy investing solely in one of the two assets. The multiple-benchmark tracking-error portfolio we obtained is one that divides the capital among the two assets with equal weights (under the normality assumption). This is simply the equally weighted investment strategy often used by practitioners.

3.2.4.2 Volatility

It has been observed that the Markowitz mean-variance portfolio suffers from severe volatility in portfolio returns due to estimation errors in mean and covariance (cf. Michaud (1989) and Best & Grauer (1991)). Our multiple-benchmark tracking-error portfolio tends to exhibit less volatility since the risk of estimation errors is mitigated by the persistency values, which are more robust to estimate. This difference is indeed observed in the numerical studies we did using the real data in Section 3.3. In what follows, we use a set of simulated data to illustrate such difference. We include the equally weighted portfolio (a.k.a. the $1/n$ portfolio) as a reference portfolio, since it is known to be effective in minimizing volatility, in particular, for a large pool of assets.

In the experiment, we simulate the monthly returns of 48 risky assets under multivariate normality assumption. We use the estimated mean, variance and covariance of the 48 Industry Portfolios from the Fama French online data library (in the period from 1981 to 2010) as the underlying distributional parameters. We adopt a rolling horizon method with an estimation window of 80 periods and investment horizon of 400 periods. In particular, we simulate 480 samples from the underlying distribution-

al parameters, and use the first 80 sample points to obtain sample mean, variance and covariance, based on which the portfolios are determined. Then the returns of the portfolios are evaluated using the 81^{st} sample point, which is an out-of-sample return. Next, the whole process moves one period forward, i.e., now the sample mean, variance and covariance are computed using sample points from the 2^{nd} to the 81^{st} sample points, and the returns are calculated using the 82^{nd} point. We continue such experiment for 400 periods. The out-of-sample returns of the three portfolios over the whole investment horizon are plotted in Figure 3.2.1. In this experiment, there were 48 benchmark portfolios, each corresponding to an individual industry portfolio.

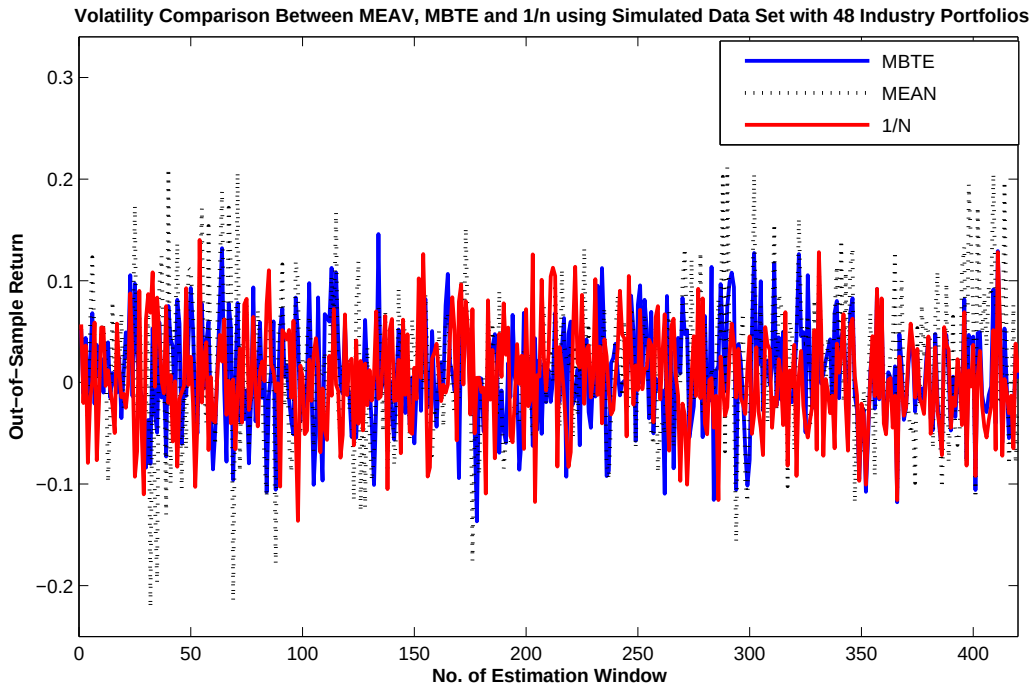


Figure 3.2.1: Out-of-sample returns of the $1/n$, Markowitz mean-variance (MEAV), and multiple-benchmark tracking-error (MBTE) portfolios over an investment horizon of 400 periods

The out-of-sample return of the $1/n$ portfolio is stable and shows only slight fluctuation over the course of the experiment. The Markowitz mean-variance portfolio how-

ever exhibits much larger volatility. In comparison, our multiple-benchmark tracking-error portfolio's performance is close to that of the $1/n$ portfolio, with significantly smaller fluctuations than the Markowitz mean-variance portfolio.

3.2.4.3 Efficient Frontier (In-Sample)

Next, we compare our multiple-benchmark tracking-error frontier with the mean-variance efficient frontier. To give an immediate idea, we first plot the multiple-benchmark tracking-error frontier and the efficient frontier assuming the full knowledge on the distributional parameters of the random returns. Here, we consider the following multiple-benchmark tracking-error model:

$$(\mathrm{T}'') \qquad \min_{\boldsymbol{e}^T\boldsymbol{x}=1,\, \boldsymbol{\mu}^T\boldsymbol{x}=K} \mathbf{E}\left[\left(Z_B(\tilde{\boldsymbol{r}}) - \tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right],$$

which is a variant of Problem (T) with additional target expected return constraint. Similarly, to obtain the mean-variance efficient frontier, we consider the Markowitz model with a target expected return constraint as follows:

$$(\mathrm{M}) \qquad \min_{\boldsymbol{e}^T\boldsymbol{x}=1,\, \boldsymbol{\mu}^T\boldsymbol{x}=K} \frac{1}{2}\boldsymbol{x}^T\Sigma\boldsymbol{x}$$

We use an experiment to illustrate the different in the two frontiers. We simulate the monthly returns of 10 risky assets under multivariate normality assumption. We use the estimated mean, variance and covariance of the monthly returns of the 10 Industry Portfolios from the Fama French online data library (in the period from 1981 to 2010) as the underlying distributional parameters. We assume the complete knowledge of means and variances of returns when solving (T″) and (M). We consider a sequence of target expected returns, K, from 0 to 0.2 with a step size of 0.0001.

For each K, we solve (T″) and (M) to obtain our multiple-benchmark tracking-error portfolio and the Markowitz mean-variance portfolio, respectively. Then we compute the variance of the two portfolios. The continuum of such K-variance pairs constitutes the frontier for each portfolio selection model, as plotted in Figure 3.2.2. Similarly, in this experiment, the benchmark portfolios are chosen to be all the extreme portfolios that invest solely in individual assets.

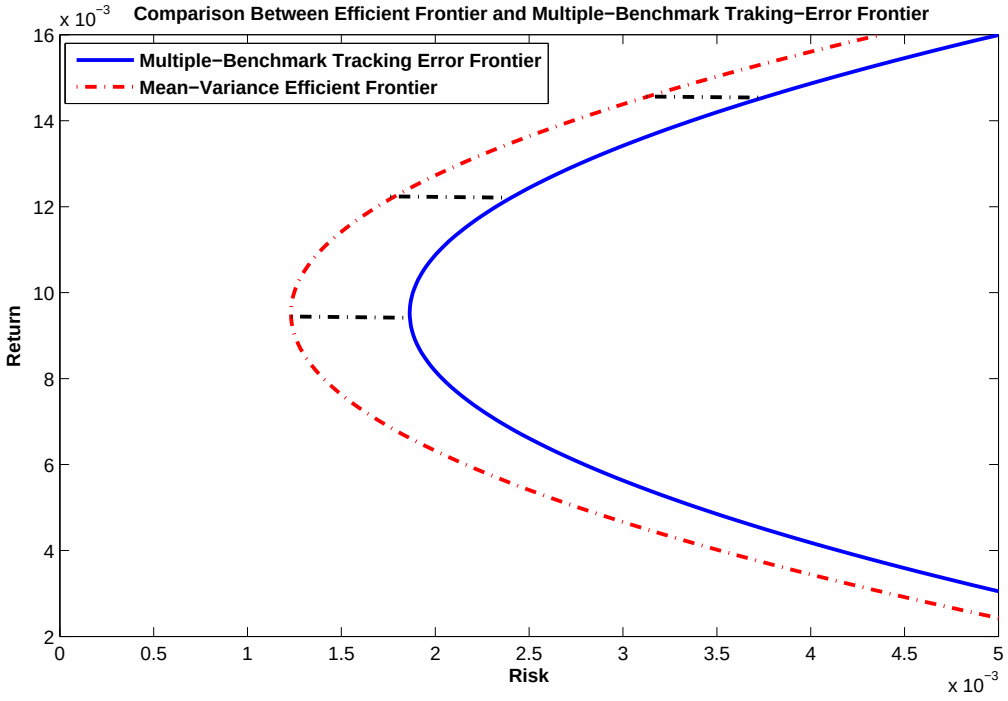


Figure 3.2.2: Risk and return with known distributional parameters and simulated data

Comparing the two frontiers in the risk-return plot, we observe a constant shift of the multiple-benchmark tracking-error frontier from the mean-variance efficient frontier. Note that a similar feature was observed for single-benchmark tracking-error portfolios by Roll (1992). The magnitude of the shift corresponds to the magnitude of the agency problem in this environment - when the portfolio manager focuses on min-

imizing the variability of the tracking error, the resulting portfolio is often inefficient and lies in the interior of the efficient frontier. We give an analytical expression for the constant shift next.

Proposition 3.9. *Under Assumption (A), the multiple-benchmark tracking-error frontier obtained from solving Problem (T″), is a constant shift from the mean-variance efficient frontier. More specifically, the distance between the two frontiers at each level of expected return is*

$$\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]^T P^T \Sigma P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{B}{C}\right)^2 - \frac{1}{C}.$$

Proof. Similar to Theorem 3.5, we solve problem (T″) and obtain the closed-form solution as follows:

$$\boldsymbol{x}_{MBTE} = P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \left(\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - K\right)\Sigma^{-1}\left(\frac{B\boldsymbol{e} - C\boldsymbol{\mu}}{AC - B^2}\right). \tag{3.2.7}$$

The corresponding portfolio variance is

$$\begin{aligned}\sigma^2_{MBTE} = \; & \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]^T P^T \Sigma P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]\right)^2 \\ & + \frac{2B}{AC - B^2}\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \frac{CK^2 - 2KB}{AC - B^2}.\end{aligned}$$

At the return level of K, the optimal Markowitz mean-variance portfolio is given by

$$\boldsymbol{x}_{MEAV} = \frac{(A - BK)\Sigma^{-1}\boldsymbol{e} + (CK - B)\Sigma^{-1}\boldsymbol{\mu}}{AC - B^2}.$$

Its variance is

$$\sigma^2_{MEAV} = \frac{1}{C} + \frac{(KC - B)^2}{C(AC - B^2)}. \tag{3.2.8}$$

At the expected return level of K, the difference in portfolio variances of the two models is given by

$$\sigma^2_{MBTE} - \sigma^2_{MEAV} = \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]^T P^T \Sigma P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{B}{C}\right)^2 - \frac{1}{C}.$$

Note that the above difference is independent of K, which indicates that the multiple-benchmark tracking-error frontier is a constant shift from the mean-variance efficient frontier. □

Note that the first term in the summand corresponds to the variance of the returns of the persistency weighted portfolio strategy, and the term B/C corresponds to the mean returns of the minimum variance strategy. We can assume that the mean returns of each portfolio used in the set of benchmarks generate higher mean returns than the minimum variance strategy. Hence,

$$\left(\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{B}{C}\right)^2$$

corresponds to the square of the excess returns of the persistency weighted portfolio strategy above the minimum variance strategy. We can now use this result to rank the performance of the tracking-error models using different benchmarks.

Proposition 3.10. *Under assumption (A), if*

$$\begin{aligned} Var(\tilde{\boldsymbol{r}}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]) - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{B}{C}\right)^2 \\ \leq Var(\tilde{\boldsymbol{r}}^T \boldsymbol{p}^j) - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T \boldsymbol{p}^j - \frac{B}{C}\right)^2, \end{aligned}$$

then the frontier constructed from the multiple-benchmark tracking-error model (T'') dominates the frontier for the single-benchmark tracking-error model constructed using

benchmark $\boldsymbol{p}^j$ *only.*

The proof of Proposition 3.10 is omitted as it is straightforward from the above analysis. This result can be used to identify complementary benchmark portfolio that can help improve the performance of the single-benchmark tracking-error model using $\boldsymbol{p}^j$. For instance, if there exists portfolio $\boldsymbol{q}^j$ such that

- $\boldsymbol{\mu}^T\boldsymbol{p}^j = \boldsymbol{\mu}^T\boldsymbol{q}^j$,
- $Var(\tilde{\boldsymbol{r}}^T\boldsymbol{p}^j) = Var(\tilde{\boldsymbol{r}}^T\boldsymbol{q}^j)$, and
- $\tilde{\boldsymbol{r}}^T\boldsymbol{p}^j$ and $\tilde{\boldsymbol{r}}^T\boldsymbol{q}^j$ are independent or negatively correlated,

then

$$Var\left(\tilde{\boldsymbol{r}}^T(\kappa\boldsymbol{p}^j + (1-\kappa)\boldsymbol{q}^j)\right) \leq \kappa^2 Var(\tilde{\boldsymbol{r}}^T\boldsymbol{p}^j) + (1-\kappa)^2 Var(\tilde{\boldsymbol{r}}^T\boldsymbol{q}^j) \leq Var(\tilde{\boldsymbol{r}}^T\boldsymbol{p}^j),$$

for any κ in $[0, 1]$. Thus $\boldsymbol{q}^j$ can be used in our multiple-benchmark model to improve the performance of the single-benchmark tracking-error model. This result shows the potential of the multiple-benchmark tracking-error model in reducing the impact of the agency problem for the investor, as it can bring the frontier of the tracking-error model closer to the mean-variance efficient frontier.

We can also show an interesting result similar to the famous Two-Fund Theorem, which says that any affine combination of two distinct mean-variance efficient portfolios is itself a mean-variance efficient portfolio.

Proposition 3.11. *Under Assumption (A), when the set of benchmark portfolios contains exactly two distinct mean-variance efficient portfolios, our multiple-benchmark tracking-error frontier coincides with the mean-variance efficient frontier. Consequent-*

ly, the multiple-benchmark tracking-error portfolio obtained from solving Problem (T) falls on the mean-variance efficient frontier.

Proof. In order to prove this result, it suffices to show that the gap between the two frontiers is zero.

Let $\boldsymbol{p}^1$ and $\boldsymbol{p}^2$ be two distinct portfolios on the mean-variance efficient frontier, and they serve as the benchmark portfolios for our multiple-benchmark tracking-error model. Their persistency values satisfy

$$\mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right] + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right] = 1.$$

From the Two-Fund Theorem, we know that the persistency weighted benchmark portfolio,

$$\mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^1 + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^2 = P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right],$$

is also a mean-variance portfolio with expected return of $\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]$. From Equation (3.2.8), the variance of this portfolio is

$$\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]^T P^T\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] = \frac{1}{C} + \frac{\left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]C - B\right)^2}{C\left(AC - B^2\right)}.$$

By Proposition 3.9, the gap between the multiple-benchmark tracking-error frontier and the mean-variance efficient frontier is

$$\begin{aligned}
& \sigma^2_{MBTE} - \sigma^2_{MEAV} \\
= \;& \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]^T P^T\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{B}{C}\right)^2 - \frac{1}{C} \\
= \;& \frac{1}{C} + \frac{\left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]C - B\right)^2}{C\left(AC - B^2\right)} - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{B}{C}\right)^2 - \frac{1}{C} \\
= \;& 0.
\end{aligned}$$

Therefore, we have completed the proof. □

Proposition 3.11 shows that our multiple-benchmark tracking-error model has the flexibility to generate the entire mean-variance efficient frontier if the benchmark portfolios are chosen properly. It is obvious from the proof that this result can be extended to the case with more than two mean-variance efficient benchmark portfolios.

3.2.4.4 Efficient Frontier (Out-of-Sample)

For the purpose of completeness, we conduct further numerical analysis by drawing the frontiers for both portfolios based on out-of-sample estimation in Figure 3.2.3. We simulate 130 samples from the underlying distributional parameters same as the previous experiment, and use the first 120 sample points to obtain sample mean, variance and covariance. We consider the sequence of values for K as before. For each K, we determine our multiple-benchmark tracking-error portfolio with the sample mean and covariance, and calculate the out-of-sample mean return and variance using the last 10 periods of data. By drawing such return-variance pairs for all K's, we get an out-of-sample multiple-benchmark tracking-error frontier. The out-of-sample mean-variance efficient frontier is obtained in a similar way.

Although the theoretical frontier of the Markowitz mean-variance portfolio could be more efficient in-sample, in the out-of-sample experiment, the estimation errors in mean and variance leads to much less efficient Markowitz portfolios.

3.2.5 Comparison with the Linear Combination Rule

To improve the performance of the Markowitz mean-variance portfolio under estimation errors, Tu & Zhou (2011) proposed to combine more sophisticated strategies with the naive $1/n$ rule. They found that the optimal affine combination of the estimated

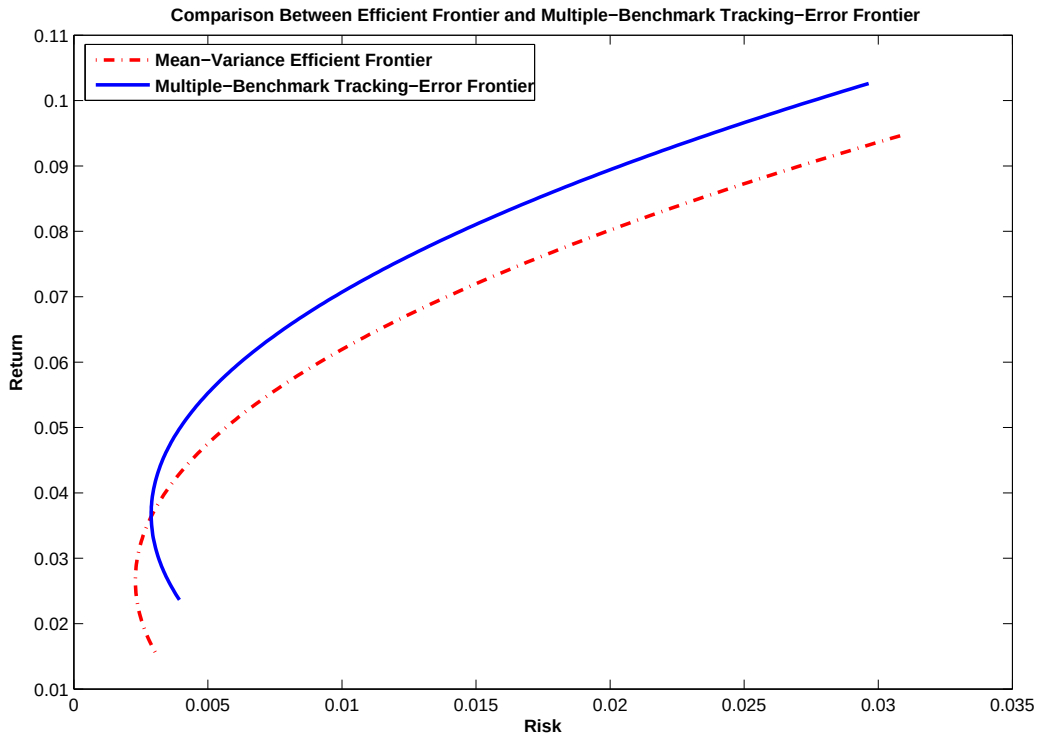


Figure 3.2.3: Risk and return with out-of-sample estimates of distributional parameters and simulated data

Markowitz portfolio and the $1/n$ portfolio often outperforms both portfolios in terms of expected mean-variance utility. In order to derive the desired result, the authors focused on the unconstrained version of the Markowitz model, i.e., without the requirement that the sum of portfolio weights equals to one. In a later study, Kirby & Ostdiek (2012) pointed out the importance of research design in driving the performance of the Markowitz portfolio. In particular, high target expected return will significantly inflate the estimation errors and result in extremely risky position for the Markowitz portfolio.

We investigate next the relationship between our multiple-benchmark tracking-error model and the linear combination rule, for different ranges of the target expected

return. Let $\boldsymbol{p}^1$ and $\boldsymbol{p}^2$ be two distinct portfolios with different expected return, i.e., $\boldsymbol{\mu}^T\boldsymbol{p}^1 \neq \boldsymbol{\mu}^T\boldsymbol{p}^2$. The linear combination rule generates a series of portfolios of the form,

$$\delta\boldsymbol{p}^1 + (1-\delta)\,\boldsymbol{p}^2,$$

where δ is the linear combination coefficient. To facilitate the comparison, for our model, we use the same two portfolios to construct our benchmark. Note that given a target expected return, the linear combination coefficient rule can be uniquely determined. Similarly, Equation (3.2.7) gives the closed-form solution of our multiple-benchmark tracking-error portfolio at the target expected return K.

In general, the variance of the linear combination portfolio will usually increase at a faster rate as the target expected return increases. On the other hand, as demonstrated earlier, the frontier of our multiple-benchmark tracking-error portfolio is a constant shift to the right from the mean-variance efficient frontier. As the target expected return increases, we expect our portfolio to be more efficient, i.e., having smaller variance than the solution produced by linear combination rule. We have the following result.

Proposition 3.12. *Under Assumption (A), the multiple-benchmark tracking-error frontier will dominate the linear combination rule frontier when the target expected return is high enough.*

Proof. At the target expected return K, the linear combination coefficient is given by

$$\delta = \frac{K - \boldsymbol{\mu}^T\boldsymbol{p}^2}{\boldsymbol{\mu}^T\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right)}.$$

Then the linear combination portfolio has the closed-form expression as follows:

$$\boldsymbol{x}_{LCR} = \frac{K}{\boldsymbol{\mu}^T\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right)}\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right) + \frac{\left(\boldsymbol{\mu}^T\boldsymbol{p}^1\right)\boldsymbol{p}^2 - \left(\boldsymbol{\mu}^T\boldsymbol{p}^2\right)\boldsymbol{p}^1}{\boldsymbol{\mu}^T\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right)}.$$

To emphasize the portfolio variance's dependence on the target expected return, we denote the variance of the linear combination portfolio as $\sigma^2_{LCR}(K)$. Then

$$\begin{aligned}\sigma^2_{LCR}(K) = &\ \boldsymbol{x}^T_{LCR}\Sigma\boldsymbol{x}_{LCR}\\ = &\ a_{LCR}K^2 + b_{LCR}K + c_{LCR},\end{aligned}$$

where

$$\begin{aligned}a_{LCR} = &\ \tfrac{1}{\left[\boldsymbol{\mu}^T\left(\boldsymbol{p}^1-\boldsymbol{p}^2\right)\right]^2}\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right)^T\Sigma\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right),\\ b_{LCR} = &\ \tfrac{1}{\left[\boldsymbol{\mu}^T\left(\boldsymbol{p}^1-\boldsymbol{p}^2\right)\right]^2}\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right)^T\Sigma\left(\left(\boldsymbol{\mu}^T\boldsymbol{p}^1\right)\boldsymbol{p}^2 - \left(\boldsymbol{\mu}^T\boldsymbol{p}^2\right)\boldsymbol{p}^1\right)\\ &+\tfrac{1}{\left[\boldsymbol{\mu}^T\left(\boldsymbol{p}^1-\boldsymbol{p}^2\right)\right]^2}\left(\left(\boldsymbol{\mu}^T\boldsymbol{p}^1\right)\boldsymbol{p}^2 - \left(\boldsymbol{\mu}^T\boldsymbol{p}^2\right)\boldsymbol{p}^1\right)^T\Sigma\left(\boldsymbol{p}^1 - \boldsymbol{p}^2\right), \text{ and}\\ c_{LCR} = &\ \tfrac{1}{\left[\boldsymbol{\mu}^T\left(\boldsymbol{p}^1-\boldsymbol{p}^2\right)\right]^2}\left(\left(\boldsymbol{\mu}^T\boldsymbol{p}^1\right)\boldsymbol{p}^2 - \left(\boldsymbol{\mu}^T\boldsymbol{p}^2\right)\boldsymbol{p}^1\right)^T\Sigma\left(\left(\boldsymbol{\mu}^T\boldsymbol{p}^1\right)\boldsymbol{p}^2 - \left(\boldsymbol{\mu}^T\boldsymbol{p}^2\right)\boldsymbol{p}^1\right).\end{aligned}$$

At the target expected return K, the multiple-benchmark tracking-error portfolio is given by Equation (3.2.7), and its variance is

$$\sigma^2_{MBTE}(K) = a_{MBTE}K^2 + b_{MBTE}K + c_{MBTE},$$

where

$$\begin{aligned}a_{MBTE} = &\ \frac{C}{AC - B^2},\\ b_{MBTE} = &\ -\frac{2B}{AC - B^2}, \text{ and}\\ c_{MBTE} = &\ \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]^T P^T\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] - \frac{C}{AC - B^2}\left(\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]\right)^2\\ &+\frac{2B}{AC - B^2}\boldsymbol{\mu}^T P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right].\end{aligned}$$

From Equation (3.2.8), the variance of the Markowitz mean-variance portfolio at target expected return K is

$$\sigma^2_{MEAV}(K) = a_{MEAV}K^2 + b_{MEAV}K + c_{MEAV},$$

where

$$\begin{aligned} a_{MEAV} &= \frac{C}{AC - B^2}, \\ b_{MEAV} &= -\frac{2B}{AC - B^2}, \text{ and} \\ c_{MEAV} &= \frac{1}{C}. \end{aligned}$$

Observe that all three variances are quadratic functions of the target expected return, and

$$a_{MBTE} = a_{MEAV},\ b_{MBTE} = b_{MEAV}.$$

Before proving the main result, we will first establish two claims.

Claim 1. *The multiple-benchmark tracking-error frontier intersects the linear combination rule frontier at the target expected return equal to the expected return of the persistency weighted benchmark portfolio, i.e., when* $K = \boldsymbol{\mu}^T \left(\mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^1 + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^2\right)$.

When $K = \boldsymbol{\mu}^T \left(\mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^1 + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^2\right)$, the linear combination portfolio is exactly the persistency weighted benchmark portfolio, i.e., $\mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^1 + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^2$. From Equation (3.2.7), the multiple-benchmark tracking-error portfolio is

$$\begin{aligned} \boldsymbol{x}_{MBTE} = {} & \mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^1 + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^2 \\ & + \left(\boldsymbol{\mu}^T \left(\mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^1 + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^2\right) - K\right)\Sigma^{-1}\left(\frac{B\boldsymbol{e} - C\boldsymbol{\mu}}{AC - B^2}\right) \\ = {} & \mathbf{E}\left[p_1(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^1 + \mathbf{E}\left[p_2(\tilde{\boldsymbol{r}})\right]\boldsymbol{p}^2, \end{aligned}$$

which is the same as the linear combination portfolio. Thus, Claim 1 is proved.

Claim 2. *The quadratic coefficient in* $\sigma^2_{MBTE}(K)$ *is less than or equal to the quadratic*

coefficient in $\sigma^2_{LCR}(K)$*, i.e.,* $a_{MEAV} \leq a_{LCR}$*.*

Consider the following optimization problem:

$$(\mathrm{M}_0) \qquad \min_{\boldsymbol{e}^T\boldsymbol{x}=0,\, \boldsymbol{\mu}^T\boldsymbol{x}=K'} \ \frac{1}{2}\boldsymbol{x}^T\Sigma\boldsymbol{x}.$$

The system of first-order optimality conditions reads

$$\begin{cases} \Sigma\boldsymbol{x}^* - \lambda_1\boldsymbol{e} - \lambda_2\boldsymbol{\mu} = 0, \\ \boldsymbol{e}^T\boldsymbol{x}^* = 0, \\ \boldsymbol{\mu}^T\boldsymbol{x}^* = K', \end{cases}$$

where λ_1 and λ_2 are Lagrange multipliers. From the first equation, we get

$$\boldsymbol{x}^* = \lambda_1\Sigma^{-1}\boldsymbol{e} + \lambda_2\Sigma^{-1}\boldsymbol{\mu}.$$

Substituting the above expression of $\boldsymbol{x}$ into the last two equations of the optimality conditions, we have

$$\begin{cases} \boldsymbol{e}^T\boldsymbol{x}^* = \lambda_1\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e} + \lambda_2\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{\mu} = \lambda_1 C + \lambda_2 B = 0, \\ \boldsymbol{\mu}^T\boldsymbol{x}^* = \lambda_1\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{e} + \lambda_2\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} = \lambda_1 B + \lambda_2 A = K', \end{cases}$$

which yields

$$\lambda_1 = -\frac{K'B}{AC - B^2}, \text{ and } \lambda_2 = \frac{K'C}{AC - B^2}.$$

Therefore, the optimal solution is

$$\boldsymbol{x}^* = \frac{K'C}{AC - B^2}\Sigma^{-1}\boldsymbol{\mu} - \frac{K'B}{AC - B^2}\Sigma^{-1}\boldsymbol{e},$$

and the minimum objective value scaled by 2 is

$$\begin{aligned}
\boldsymbol{x}^{*T}\Sigma\boldsymbol{x}^{*} = & \left(\tfrac{K^{'}C}{AC-B^2}\right)^2 \boldsymbol{\mu}^T\Sigma^{-1}\Sigma\Sigma^{-1}\boldsymbol{\mu} + \left(\tfrac{K^{'}B}{AC-B^2}\right)^2 \boldsymbol{e}^T\Sigma^{-1}\Sigma\Sigma^{-1}\boldsymbol{e} \\
& -\tfrac{K^{'2}CB}{(AC-B^2)^2}\boldsymbol{\mu}^T\Sigma^{-1}\Sigma\Sigma^{-1}\boldsymbol{e} - \tfrac{K^{'2}BC}{(AC-B^2)^2}\boldsymbol{e}^T\Sigma^{-1}\Sigma\Sigma^{-1}\boldsymbol{\mu} \\
= & \left(\tfrac{K^{'}C}{AC-B^2}\right)^2 A + \left(\tfrac{K^{'}B}{AC-B^2}\right)^2 C - \tfrac{2K^{'2}CB^2}{(AC-B^2)^2} \\
= & \tfrac{K^{'2}C^2A-K^{'2}CB^2}{(AC-B^2)^2} \\
= & \tfrac{C}{AC-B^2}K^{'2}.
\end{aligned}$$

Observe that $(\boldsymbol{p}^1 - \boldsymbol{p}^2)$ is a feasible solution to Problem (M_0) with $K^{'} = \boldsymbol{\mu}^T(\boldsymbol{p}^1 - \boldsymbol{p}^2)$, then it must satisfy

$$\begin{aligned}
(\boldsymbol{p}^1 - \boldsymbol{p}^2)^T\Sigma(\boldsymbol{p}^1 - \boldsymbol{p}^2) \geq & \ \boldsymbol{x}^{*T}\Sigma\boldsymbol{x}^{*} \\
= & \ \tfrac{C}{AC-B^2}K^{'2} \\
= & \ \tfrac{C}{AC-B^2}\left[\boldsymbol{\mu}^T(\boldsymbol{p}^1 - \boldsymbol{p}^2)\right]^2.
\end{aligned}$$

Rearrange the terms, we get

$$\frac{1}{\left[\boldsymbol{\mu}^T(\boldsymbol{p}^1 - \boldsymbol{p}^2)\right]^2}(\boldsymbol{p}^1 - \boldsymbol{p}^2)^T\Sigma(\boldsymbol{p}^1 - \boldsymbol{p}^2) \geq \frac{C}{AC-B^2},$$

which is exactly $a_{LCR} \geq a_{MEAV}$. Therefore, we have proved Claim 2.

Now in order to prove the proposition, we only need to discuss two cases following Claim 2, $a_{LCR} = a_{MEAV}$ and $a_{LCR} > a_{MEAV}$.

Case 1. $a_{LCR} = a_{MEAV}$.

By the definition of mean-variance efficient frontier, $\sigma^2_{MEAV}(K) \leq \sigma^2_{LCR}(K)$, for any K, i.e.,

$$b_{MEAV}K + c_{MEAV} \leq b_{LCR}K + c_{LCR}, \forall K.$$

Then we must have $b_{MEAV} = b_{LCR}$. Otherwise, the above inequality will be violated

as $K \to +\infty$ if $b_{MEAV} > b_{LCR}$, or $K \to -\infty$ if $b_{MEAV} < b_{LCR}$. Consequently,

$$a_{MBTE} = a_{MEAV} = a_{LCR}, \text{ and } b_{MBTE} = b_{MEAV} = b_{LCR}.$$

Furthermore, since the multiple-benchmark tracking-error frontier has an intersection point with the linear combination rule frontier, it must be the case that

$$c_{MBTE} = c_{LCR},$$

which implies that the multiple-benchmark tracking-error frontier coincides with the linear combination rule frontier. i.e.,

$$\sigma^2_{LCR}(K) = \sigma^2_{MBTE}(K), \forall K.$$

Case 2. $a_{LCR} > a_{MEAV}$.

Recall that $a_{MBTE} = a_{MEAV}$. Then $a_{LCR} > a_{MBTE}$. In this case, it is obvious that there exists a constant $\bar{K}$ such that

$$\sigma^2_{LCR}(K) > \sigma^2_{MBTE}(K), \forall K \geq \bar{K}.$$

Combining Case 1 and Case 2, we complete the proof. □

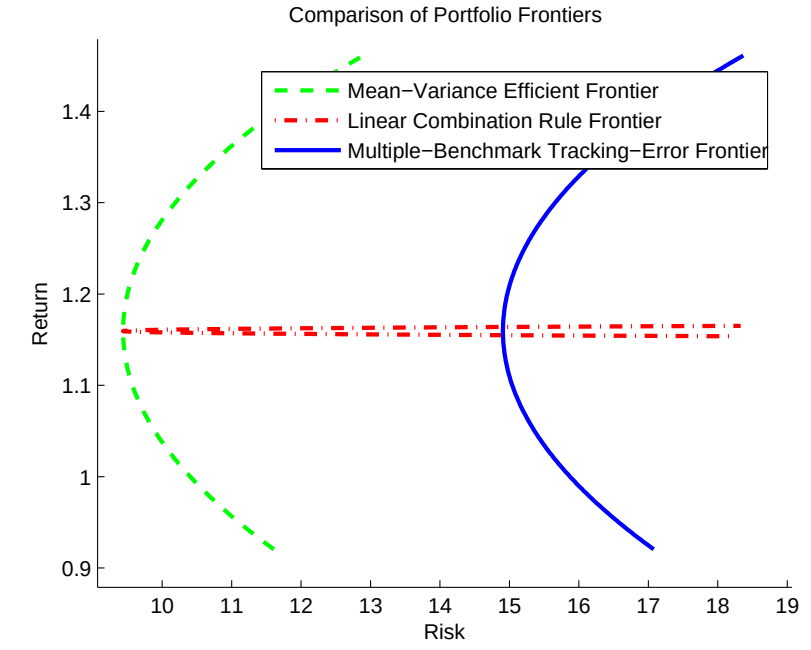
The above result shows that there exists a threshold such that once the target expected return exceeds this threshold, the linear combination rule would be less efficient than the multiple-benchmark tracking-error portfolio. In fact, from our numerical tests, such threshold value is usually very small, and the performance of the linear combination rule deteriorates significantly when the target expected return increases. We use a numerical experiment to illustrate this effect.

Following the same procedures as those in the previous subsection, we plot the frontiers of the mean-variance, linear combination, and multiple-benchmark tracking-error portfolios in Figure 3.2.4 using simulated data under the normality assumption. In Figure 3.2.4, the mean, variance and covariance of the monthly returns of the 10 Industry Portfolios from the Fama French online data library (in the period from 1981 to 2010) are used as the underlying distributional parameters[3]. Same as Tu & Zhou (2011), the $1/n$ portfolio and the global minimum-variance portfolio are used to construct the linear combination portfolios, and as benchmarks for our multiple-benchmark tracking-error portfolios. Figure 3.2.4(a) and 3.2.4(b) plot the same three frontiers with different ranges of risk and return to provide a complete picture of these curves. Figure 3.2.4(b) zooms in 3.2.4(a) vertically but zooms out horizontally at the same time.

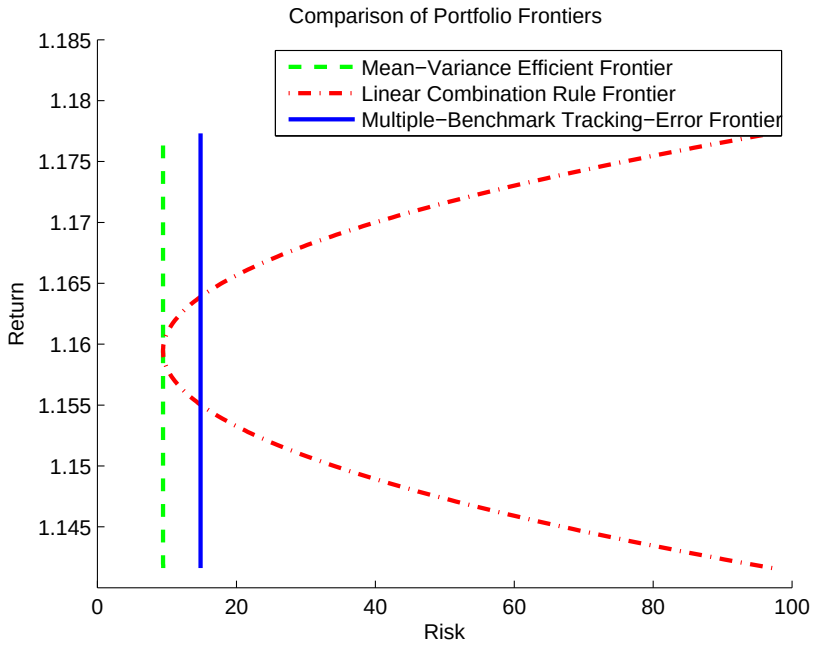
From the figures, although the linear combination rule does outperforms our model in a range of target expected return, the range is indeed very narrow. Once outside the range, the linear combination rule quickly loses its advantage. Note the values along the x-axis in Figure 3.2.4(a) and 3.2.4(b). The variance of the linear combination portfolio grows so fast as the target expected return increases that the other two frontiers appear to be almost straight lines in the picture. In fact, the performance of the linear combination rule is extremely sensitive to the expected return difference between the two source portfolios, i.e., $\left|\boldsymbol{\mu}^T(\boldsymbol{p}^1 - \boldsymbol{p}^2)\right|$. If the difference is very small, at a high target expected return level, the linear combination rule must take very extreme portfolio positions to achieve the desired target return, which usually involves a tremendous amount of risk.

Remark 3.13. We can also interpret the above result from the perspective of utility

[3]We have tried various scenarios, and the outcomes are very similar to this example. Hence, we only present the analysis for one example here.



(a)



(b)

Figure 3.2.4: Risk and return with known distributional parameters and simulated data

theory. Suppose that the portfolio manager has the mean-variance utility of the form,

$$\boldsymbol{\mu}^T \boldsymbol{x} - \frac{\gamma}{2} \boldsymbol{x}^T \Sigma \boldsymbol{x},$$

where γ is the risk aversion parameter. Then our multiple-benchmark tracking-error portfolio will be preferred over the linear combination rule for relatively small γ, i.e., less risk aversion. From the figures, the portfolio manager has to be extremely risk averse for the linear combination rule to be better than the multiple-benchmark tracking-error portfolio.

3.2.6 Transaction Cost

Transaction costs are often inevitable in real investment activities. As discussed before, we can explicitly include the buy-and-hold strategy as a benchmark portfolio to control the transaction volume. In this subsection, we show that our model is also capable of handling transaction cost in the conventional way by adding a penalty term into the objective function.

To facilitate the following exposition, we introduce some additional notation. Let W denote the wealth at the beginning of the current investment period, and define $\boldsymbol{x}^0$ as the starting portfolio, i.e., the initial weights of capital on each asset. As before, $\boldsymbol{x}$ represents the current investment decision. In this case, it can also be referred to as portfolio repositioning decision. The transaction volume is measured by $\sum_{i=1}^{n} W \left| x_i - x_i^0 \right| = W \boldsymbol{e}^T \left| \boldsymbol{x} - \boldsymbol{x}^0 \right|$. However, the problem becomes non-smooth if we directly work with the transaction cost that is linear in the transaction volume. To make the problem more tractable and emphasize on avoiding high transaction volume, we extend the basic model in Problem (T) by adding a penalty term on the quadratic

transaction volume. The problem is formulated as follows:

$$\text{(TC)} \qquad \min_{\boldsymbol{e}^T\boldsymbol{x}=1} \ \mathbf{E}\left[\left(Z_B(\tilde{\boldsymbol{r}}) - \sum_{i=1}^{n} \tilde{r}_i x_i\right)^2\right] + \nu(\boldsymbol{x}-\boldsymbol{x}^0)^T(\boldsymbol{x}-\boldsymbol{x}^0),$$

where $\nu \geq 0$ is a penalty parameter that captures the effect of the quadratic transaction volume, $W^2(\boldsymbol{x}-\boldsymbol{x}^0)^T(\boldsymbol{x}-\boldsymbol{x}^0)$. Since ν is a constant, we can absorb W^2 into ν. The new objective can be interpreted as an adjusted disutility function of the investor with a penalty on the transaction volume, where ν characterizes the investor's aversion to high transaction volume. With such change, Problem (TC) remains a convex quadratic programming problem, and we are able to establish its closed-form solution as shown in the following theorem.

Proposition 3.14. *Under Assumption (A), the closed-form solution to Problem (TC) is given by*

$$\frac{D\boldsymbol{e}}{\boldsymbol{e}^T D\boldsymbol{e}} + \left(I_n - \frac{DJ_n}{\boldsymbol{e}^T D\boldsymbol{e}}\right) D\left(\boldsymbol{\mu}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] + \Sigma P \mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \nu\boldsymbol{x}^0\right), \tag{3.2.9}$$

where $D = \left(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T + \nu I_n\right)^{-1}$*;* I_n *is the identity matrix; and* J_n *is the matrix in* $\mathbb{R}^{n\times n}$ *with all entries being 1.*

Proof. Since the convexity is preserved in Problem (TC) by adding the quadratic penalty term, the first-order optimality conditions are both necessary and sufficient to characterize the solution. The Lagrangian of Problem (TC) is given by

$$\mathcal{L}(\boldsymbol{x},\pi) = \mathbf{E}\left[\left(Z_B\left(\tilde{\boldsymbol{r}}\right) - \tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right] + \nu(\boldsymbol{x}-\boldsymbol{x}^0)^T(\boldsymbol{x}-\boldsymbol{x}^0) + 2\pi\left(1-\boldsymbol{e}^T\boldsymbol{x}\right).$$

The first-order conditions yield

$$2\mathbf{E}\left[\left(Z_B\left(\tilde{\boldsymbol{r}}\right)-\tilde{\boldsymbol{r}}^T\boldsymbol{x}^*\right)\left(-\tilde{\boldsymbol{r}}\right)\right]+2\nu(\boldsymbol{x}^*-\boldsymbol{x}^0)-2\pi\boldsymbol{e}=0$$

and

$$\sum_{i=1}^{n}x_i^*=1.$$

The first set of equalities can be rewritten as

$$\left(\begin{array}{cc}\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n & -\boldsymbol{e}\end{array}\right)\left(\begin{array}{c}\boldsymbol{x}^*\\ \pi\end{array}\right)=\mathbf{E}[Z_B\left(\tilde{\boldsymbol{r}}\right)\tilde{\boldsymbol{r}}]+\nu\boldsymbol{x}^0.$$

Applying Lemma 1.3, we have

$$\left(\begin{array}{cc}\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n & -\boldsymbol{e}\end{array}\right)\left(\begin{array}{c}\boldsymbol{x}^*\\ \pi\end{array}\right)=\boldsymbol{\mu}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]+\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]+\nu\boldsymbol{x}^0.$$

Since $\nu\geq 0$ and $(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T)$ is positive definite, $(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n)$ is also positive definite. In particular, it has an inverse. The we can multiple both sides of above equation by $(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n)^{-1}$ and obtain

$$\begin{aligned}&\left(\begin{array}{cc}I_n & -\left(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n\right)^{-1}\boldsymbol{e}\end{array}\right)\left(\begin{array}{c}\boldsymbol{x}^*\\ \pi\end{array}\right)\\ &\quad=\left(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n\right)^{-1}\left(\boldsymbol{\mu}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]+\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]+\nu\boldsymbol{x}^0\right).\end{aligned}\tag{3.2.10}$$

Multiplying $\boldsymbol{e}^T$ to the above equality and simplifying using the fact that $\boldsymbol{e}^T\boldsymbol{x}^*=1$, we get

$$\pi=\frac{1-\boldsymbol{e}^T\left(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n\right)^{-1}\left(\boldsymbol{\mu}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]+\Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right]+\nu\boldsymbol{x}^0\right)}{\boldsymbol{e}^T\left(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^T+\nu I_n\right)^{-1}\boldsymbol{e}}.$$

Substituting this expression of π into Equation (3.2.10), and simplifying the resulting expression with a substitution of $D = \left(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T + \nu I_n\right)^{-1}$, we have

$$\boldsymbol{x}^* = \frac{D\boldsymbol{e}}{\boldsymbol{e}^T D\boldsymbol{e}} + \left(I_n - \frac{DJ_n}{\boldsymbol{e}^T D\boldsymbol{e}}\right) D\left(\boldsymbol{\mu}\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] + \Sigma P\mathbf{E}\left[\boldsymbol{p}(\tilde{\boldsymbol{r}})\right] + \nu\boldsymbol{x}^0\right),$$

where J_n denotes the matrix in $\mathbb{R}^{n\times n}$ with all entries being 1. Thus, we obtain the closed-form solution to Problem (TC). □

3.3 Numerical Studies

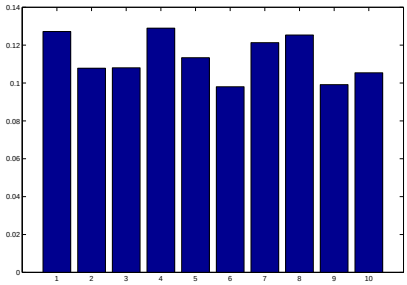
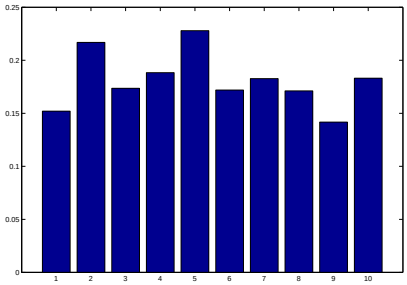
In this section, we will present some numerical studies based on real data to test the performance of the models proposed in this chapter. We start by describing the setup of these studies, including data sets, comparison portfolio strategies, methodology and performance measures.

3.3.1 Data Sets

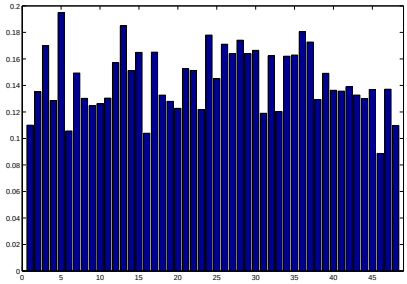
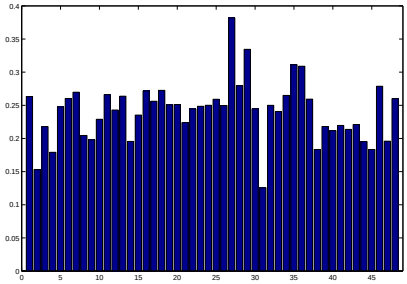
We evaluate the performance of our model in four empirical data sets listed in Table 3.3.1. The data sets we choose fall into two categories. The first three data sets are portfolios representing the U.S. stock market, and the last one is comprised of individual U.S. stocks. We present the analysis on the risk and return of these data sets in Figure 3.3.1. The first graph in each panel shows the annualized mean return and the second graph shows the annualized standard deviation of the returns. All the sample points are used in the calculation. From the figure, we observe that the four data sets demonstrate distinct risk and return characteristics, such as different spreads of mean returns within the same data set, and different risk levels, etc.

Annualized Mean Return | Annualized Standard Deviation

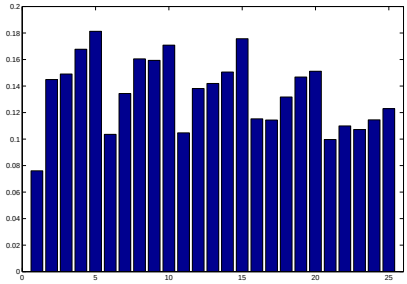
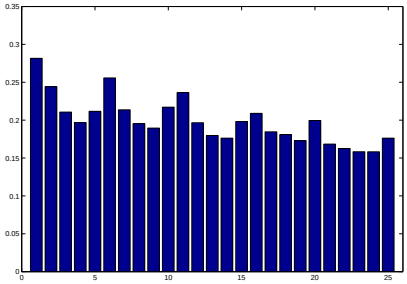
10 Industry Portfolios (10Ind)

48 Industry Portfolios (48Ind)

25 Size/BTM Portfolios (25FF)

8 U.S. Stocks (8Stock)

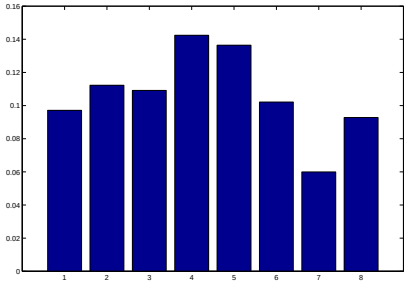
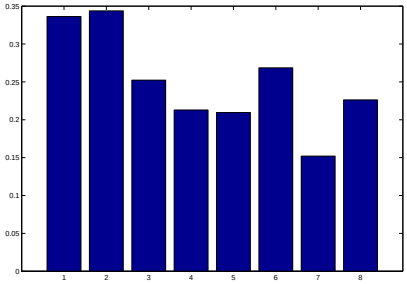



Figure 3.3.1: Risk and return characteristics of the data sets

Abbreviation	Data Set and Description	n	Time Period	Source
10Ind	Ten industry portfolios representing the U.S. stock market	10	07/1963 –12/2011	K. French
48Ind	Forty-eight industry portfolios representing the U.S. stock market	48	07/1963 –12/2011	K. French
25FF	Twenty-five Fama and French portfolios of firms sorted by size and book-to-market	25	07/1963 –12/2011	K. French
8Stock	Eight U.S. stocks (Crude Oil, J.P. Morgan Funds, General Electric Company, The Coca-Cola Company, Johnson & Johnson, International Business Machines Corp., Gold Ounce, AT & T Inc.)	8	08/1980 –01/2013	Bloomberg

Notes: "K. French" refers to the Kenneth R. French data library available online at *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*

Table 3.3.1: Data sets used in empirical experiments

3.3.2 Portfolio Models

The portfolio strategy developed in this chapter depends on the group of benchmark portfolios that are being tracked. To evaluate the performance of the multiple-benchmark tracking-error portfolio, we compare the performance of our portfolio against that of its benchmarks as well as that of the $1/n$ portfolio. We choose the buy-and-hold strategy in addition to another competitive portfolio as benchmarks. All the benchmark portfolios are listed in Table 3.3.2.

The multiple-benchmark tracking-error portfolio is listed as Portfolio 0a and the multiple-benchmark tracking-error portfolio with penalty on transaction volume is listed as Portfolio 0b in Table 3.3.2. We use the closed-from solutions in Equation (3.2.1)

No.	Portfolio Selection Model	Abbreviation
0a	Multiple-benchmark tracking-error portfolio	MBTE
0b	Multiple-benchmark tracking-error portfolio with penalty on transaction volume	MBTEP
1	Equally-weighted ($1/n$) portfolio (DeMiguel et al. (2007))	1/n
2	Markowitz mean-variance portfolio with target expected return (Markowitz (1952))	MEAV
3	Minimum-variance portfolio without short-sale constraints	MINU
4	Minimum-variance portfolio with covariance matrix being a weighted average of sample covariance matrix and the single-index covariance matrix (Ledoit & Wolf (2003))	M1FAC
5	Minimum-variance portfolio with covariance matrix being a weighted average of sample covariance matrix and the diagonal covariance matrix (Ledoit & Wolf (2003))	MIND
6	Minimum-variance portfolio with covariance matrix being a weighted average of sample covariance matrix and the identity matrix (Ledoit & Wolf (2003))	M1PAR
7	Minimum-variance portfolio with generalized constraints (DeMiguel et al. (2007))	GMC
8	On-line portfolio using multiplicative updates (Helmbold et al. (1998))	MUL
9	Minimum CVaR (Conditional Value-at-Risk) portfolio (Rockafellar & Uryasev (2000), Rockafellar & Uryasev (2004))	CVAR
10	Partial minimum-variance portfolio with k calibrated by maximizing the portfolio return in previous period (DeMiguel et al. (2009))	PARR
11	Buy-and-hold strategy	B-N-H

Table 3.3.2: List of portfolio strategies considered

and Equation (3.2.9) to compute our multiple-benchmark tracking-error portfolios.

Portfolios 1–11 listed in Table 3.3.2 serve two purposes. First, a subset of these portfolios are used as benchmark portfolios to compute the multiple-benchmark tracking-error portfolios. Second, all of these portfolios serve as comparison portfolios to gauge

the out-of-sample performance of our multiple-benchmark tracking-error portfolios.

The first comparison portfolio is the $1/n$ strategy, which simply assigns equal weights to all the assets in the data set. The Markowitz mean-variance portfolio (MEAV) relies on estimates of mean, variance and covariance of the returns, and is computed from Problem (M). The target expected return is set to be the expected return of the $1/n$ portfolio. Such target has been observed to be more appropriate for the Markowitz model (cf. Kirby & Ostdiek (2012)). Then we consider the class of minimum-variance portfolios, the first of which is the minimum-variance portfolio without short-sale constraints (MINU). The next three minimum-variance portfolios (M1FAC, MIND, and M1PAR) are formed using different covariance estimation techniques as described in Ledoit & Wolf (2003) and Ledoit & Wolf (2004). The last portfolio in this set (GMC) is adopted from DeMiguel et al. (2007). It is a combination of the $1/n$ portfolio and the constrained minimum-variance portfolio. The additional constraint is $\boldsymbol{x} \geq a\boldsymbol{e}$ with $a = 1/(2n)$. We also consider the on-line portfolio (MUL) using multiplicative updates as studied in Helmbold et al. (1998). The portfolio that minimizes the Conditional Value-at-Risk (CVAR), a coherent risk measure, is also included in our study. This portfolio is supposed to be very conservative and would refrain from much risk. For detailed discussion on CVaR, the reader is referred to Rockafellar & Uryasev (2000) and Rockafellar & Uryasev (2004). In addition, we consider the minimum-variance portfolio with a constraint on the portfolio norm developed in DeMiguel et al. (2009). In particular, PARR is the partial minimum-variance portfolio with k calibrated using cross-validation over portfolio variance, where k indicates which of the $n-1$ partial minimum-variance portfolios will yield the maximum last period portfolio return. Finally, we consider the buy-and-hold strategy, which makes no change in the allocation of capital in different assets. The initial portfolio

for the buy-and-hold strategy is set to be the $1/n$ portfolio in all the experiments[4].

3.3.3 Methodology

In each data set, we apply the rolling-horizon procedure to conduct the empirical analysis. Consider the benchmark portfolios chosen from one of the two groups listed in Table 3.3.2. The details of the methodology are summarized as follows:

1. Denote the total number of returns in the data set to be $\hat{\tau}$. We choose a history of length τ with $\tau < \hat{\tau}$, over which we conduct the estimation. In our experiments, $\tau = 120$, which corresponds to 10 years of monthly data.

2. Using the data in the estimation window, we estimate the parameters ($\boldsymbol{\mu}$, Σ, $\mathbf{E}[Z_B(\tilde{\boldsymbol{r}})]$, and $P\mathbf{E}[\boldsymbol{p}(\tilde{\boldsymbol{r}})]$), and compute the portfolios of investment strategies listed in Table 3.3.2.

3. The performance of the portfolios are then evaluated. The details of these measures are discussed in the next subsection.

4. Roll forward the time horizon by adding the next data point of the data set and dropping the first data point of the current estimation window.

5. By doing this repeatedly until the last time period, we obtain $\hat{\tau} - \tau$ portfolio-weight vectors for each strategy.

3.3.4 Performance Measures

Let t index the time periods with $t = 1, \ldots, \hat{\tau}$. We compute the portfolios at the beginning of period t for $t = \tau + 1, \ldots, \hat{\tau}$ using past information from the previous τ

[4]We have tested various initial portfolio positions, and found that the results for the buy-and-hold strategy are not sensitive to this initial condition.

periods. Let $\tilde{\boldsymbol{r}}^t$ denote the return for period t. For a portfolio strategy, we use $\boldsymbol{x}^t$ to represent the investment decision made for period t, and $\boldsymbol{x}^{t0}$ to represent the portfolio position at the beginning of period t before the repositioning decision $\boldsymbol{x}^t$ is made. The performance measures are listed as follows:

1. *In-sample tracking error*:

$$\frac{1}{\hat{\tau}-\tau}\sum_{t=\tau+1}^{\hat{\tau}}\sum_{t'=t-\tau}^{t-1}\left[\left(\tilde{\boldsymbol{r}}^{t'}\right)^T\boldsymbol{x}^t - Z_B\left(\tilde{\boldsymbol{r}}^{t'}\right)\right]^2.$$

2. *Turnover rate*:

$$\frac{1}{\hat{\tau}-\tau}\sum_{t=\tau+1}^{\hat{\tau}}\sum_{i=1}^{n}\left|x_i^t - x_i^{t0}\right|.$$

3. *Out-of-sample Sharpe ratio net of proportional transaction costs of 50 basis point (net Sharpe ratio)*:

$$\frac{\widehat{\mu}}{\widehat{\sigma}},$$

where

$$\widehat{\mu} = \frac{1}{\hat{\tau}-\tau}\sum_{t=\tau+1}^{\hat{\tau}}\left[\left(1+\left(\tilde{\boldsymbol{r}}^t\right)^T\boldsymbol{x}^t\right)\left(1-0.005\sum_{i=1}^{n}\left|x_i^t - x_i^{t0}\right|\right)-1\right],$$

and

$$\widehat{\sigma} = \sqrt{\frac{1}{\hat{\tau}-\tau-1}\sum_{t=\tau+1}^{\hat{\tau}}\left[\left(1+\left(\tilde{\boldsymbol{r}}^t\right)^T\boldsymbol{x}^t\right)\left(1-0.005\sum_{i=1}^{n}\left|x_i^t - x_i^{t0}\right|\right)-1-\widehat{\mu}\right]^2}.$$

4. *Out-of-Sample Net Aggregate Return*:

$$\begin{cases} \prod_{t=\tau+1}^{T} \left(1 + (\tilde{\boldsymbol{r}}^t)^T \boldsymbol{x}^t\right) \left(1 - 0.005 \sum_{i=1}^{n} |x_i^t - x_i^{t0}|\right), \\ \qquad \text{if } \left(1 + (\tilde{\boldsymbol{r}}^t)^T \boldsymbol{x}^t\right) \left(1 - 0.005 \sum_{i=1}^{n} |x_i^t - x_i^{t0}|\right) > 0, \ \forall t = \tau + 1, \\ 0, \\ \qquad \text{otherwise.} \end{cases}$$

The out-of-sample net aggregate return measures the long-term wealth growth of the portfolio strategies, where the second situation represents bankruptcy. It is possible since the model allows short sales.

3.3.5 Normality Assumption

The closed-form solutions in the previous section are established under the assumption that the return follows a multivariate normal distribution. We first check the validity of the normality assumption by drawing the Quantile-Quantile plot (QQ plot) of the Mahalanobis distance of the data from the first estimation window against that of a multivariate normal distribution for each data set. For comparison, we use the sample mean and sample covariance in place of their respective true values for the multivariate normal distribution. The plots are presented in Figure 3.3.2.

From these QQ plots, we observe that the sample Mahalanobis distance of the risky asset returns in all data sets demonstrate significant deviation from the normality assumption with heavy tails. However, as we will see later, the discrepancies shown in the QQ plots do not appear to be a major problem, and our multiple-benchmark tracking-error portfolio demonstrates greater superiority in the out-of-sample performance even though the normality assumption might not be completely satisfactory.

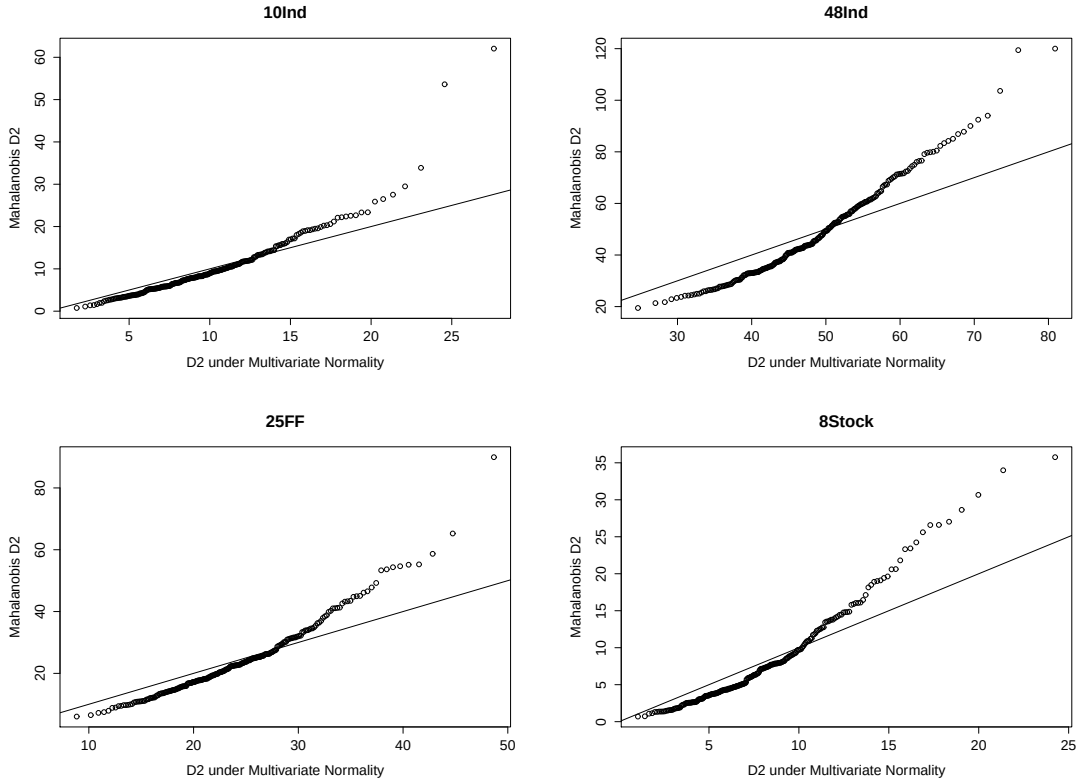


Figure 3.3.2: QQ plots of the distributions of asset returns against multivariate normal distribution

3.3.6 Results and Discussion

In this subsection, we first present results on the basic model, where the buy-and-hold strategy is included in the set of benchmark portfolios to control the transaction volume. Next, the results on the extended model, i.e., Problem (TC), are discussed.

3.3.6.1 Multiple-Benchmark Tracking-Error Portfolio with the Buy-and-Hold Strategy as a Benchmark

We first report results on the tracking error for each data set to show that our portfolio provides good tracking records of the highest benchmark return. Next, we use all the other performance measures discussed before to evaluate the performance of our

portfolio against its benchmarks. Finally, we discuss the results of the robustness tests on net aggregate returns, where we consider random starting times and random lengths of the investment horizon to gauge the out-of-sample performance of our portfolio strategy. Since all the performance measures except the tracking error are computed out-of-sample, we often drop the descriptive terms, "out-of-sample" and "in-sample", in the following discussion.

In all the experiments here, we use two benchmark portfolios to obtain our multiple-benchmark tracking-error portfolio, one from Portfolio 1–10 in Table 3.3.2, and the other is Portfolio 11 in Table 3.3.2, i.e., the buy-and-hold strategy. We report the results of the first benchmark portfolio and the corresponding multiple-benchmark tracking-error portfolio, as the buy-and-hold strategy for each single period only serves the purpose of controlling for transaction volume.

Tracking Error

As the portfolio is constructed to track a set of benchmark portfolios, the first step is to evaluate how closely our portfolio tracks the best return from the benchmark portfolios. The results on tracking errors are summarized in Table 3.3.3. Comparing the performance of the MBTE portfolio with that of its benchmark, we observe that the tracking errors of the MBTE portfolio are always smaller than those of its benchmarks. Note that the returns are not exactly normally distributed, so it is not guaranteed that the MBTE portfolio would be the best even in terms of in-sample mean squared tracking error. However, the results in Table 3.3.3 provide partial justification that the MBTE portfolio might still perform well even when the normality assumption is not completely satisfied.

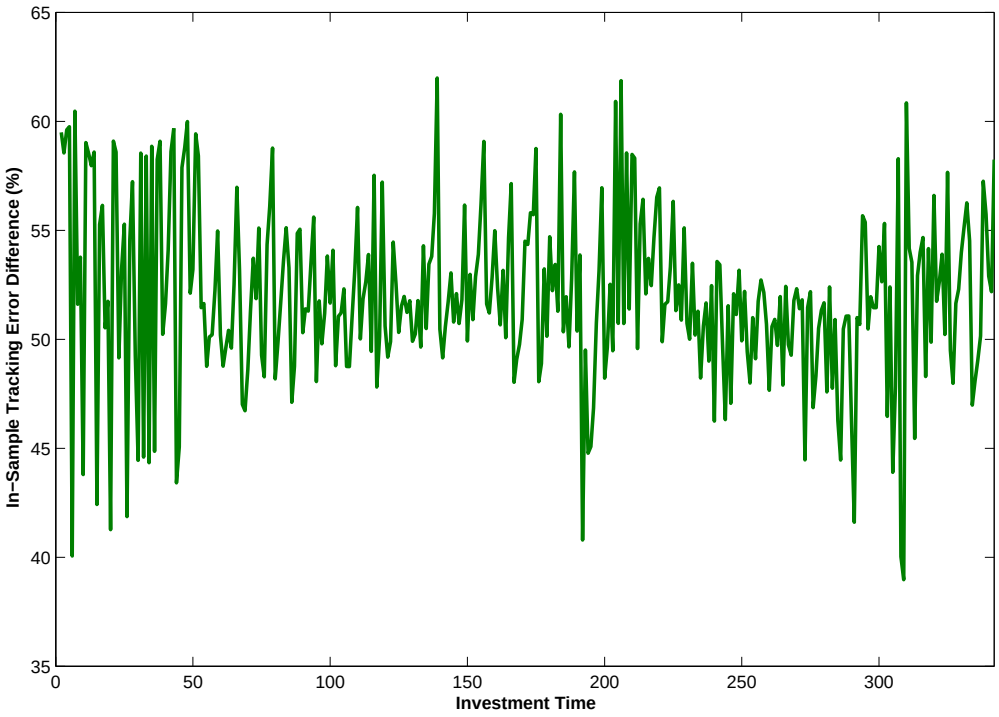
To visualize the tracking error, we plot the percentage decrease in in-sample mean squared tracking errors at every period from our portfolio compared to the PARR port-

10Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	3.772E-5	1.696E-5	2.527E-2	1.232E-2	6.447 E-4	2.939E-4	7.120E-4	2.962E-4	6.912E-4	2.863E-4
	(44.96%)		(48.74%)		(45.58%)		(41.60%)		(41.41%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	6.102E-4	2.520E-4	5.898E-4	2.433E-4	1.778E-4	7.778E-5	3.596E-5	1.608E-05	1.823E-3	8.205E-4
	(41.30%)		(41.25%)		(43.75%)		(44.72%)		(45.00%)	
48Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	1.319E-4	2.895E-5	1.270E-1	3.901E-2	4.850E-3	1.658E-3	5.328E-3	1.805E-3	4.668E-3	1.580E-3
	(21.95%)		(30.72%)		(34.18%)		(33.87%)		(33.84%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	4.294E-3	1.403E-3	4.224E-3	1.388E-3	6.319E-4	1.746E-4	1.243E-4	2.749E-5	5.528E-2	1.515E-2
	(32.68%)		(32.86%)		(27.63%)		(22.12%)		(27.40%)	
25FF	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	3.626E-5	1.023E-5	7.647E-2	2.502E-2	2.630E-3	9.705E-4	2.949E-3	9.991E-4	2.684E-3	9.067E-4
	(28.23%)		(32.72%)		(36.91%)		(33.88%)		(33.78%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	2.289E-3	7.833E-4	2.316E-4	7.915E-4	2.509E-4	8.129E-5	3.216E-5	9.357E-6	9.407E-3	3.304E-3
	(34.22%)		(34.17%)		(32.40%)		(29.09%)		(35.12%)	
8Stock	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	4.064E-4	1.945E-4	2.184E-2	1.082E-2	9.718E-4	4.655E-4	1.843E-3	9.255E-4	1.891E-3	9.527E-4
	(47.87%)		(49.53%)		(47.90%)		(50.22%)		(50.38%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	1.762E-3	8.873E-4	1.439E-3	7.218E-4	1.641E-3	8.227E-4	3.964E-4	1.891E-4	2.035E-3	1.004E-3
	(50.36%)		(50.16%)		(50.14%)		(47.71%)		(49.31%)	

Note: This table reports the in-sample tracking error and the corresponding percentage reduction in in-sample tracking error by the MBTE portfolio (in brackets).

Table 3.3.3: Comparison on in-sample tracking error

folio in Figure 3.3.3. Our multiple-benchmark tracking-error portfolio is constructed by using the buy-and-hold strategy and the PARR portfolio as benchmark portfolios. The figure shows a significant difference in tracking errors between the two portfolios. All the differences are positive, and on average, the in-sample tracking error of the MBTE portfolio demonstrates a 40% to 60% reduction from that of the PARR portfolio, which shows that the MBTE portfolio better tracks the highest return from the group of benchmark portfolios. Figures for the other scenarios are similar, so only one is reported here as an illustration.



Notes: This figure plots the percentage decrease in the in-sample mean squared tracking error at every period from the MBTE portfolio compared to the PARR portfolio.

Figure 3.3.3: Tracking-error difference between the PARR portfolio and the multiple-benchmark tracking-error portfolio using the buy-and-hold strategy and the PARR portfolio as benchmarks in the "10Ind" data set

Turnover Rate

Table 3.3.4 reports the turnover rate and the corresponding percentage deduction in turnover rate by the MBTE portfolio. From the table, we observe that the MBTE portfolio has lower turnover rates than its respective benchmark portfolio in all cases except for the GMC portfolio in the "48Ind" data set, where the GMC portfolio has a slightly smaller turnover rate than the corresponding MBTE portfolio. This is exactly the desired effect of introducing the buy-and-hold strategy as one of the benchmark portfolios. In particular, the turnover rates of its corresponding MBTE portfolio are only half as large as the turnover rates of the PARR portfolio across all the data sets. The MBTE portfolio demonstrates a decrease in turnover rate of over 40% from the MUL portfolio across all the data sets. In addition, the turnover rates of the respective MBTE portfolio are at least 30% less than those of the $1/n$ and MEAV portfolios, and the turnover rates of the respective MBTE portfolio are over 20% less than those of the MIND, M1PAR, MINU, M1FAC, and CVAR portfolios.

Out-of-Sample Net Sharpe Ratio

Table 3.3.5 reports the out-of-sample net Sharpe ratio and the corresponding p-value that the net Sharpe ratio of each benchmark strategy is smaller than that of the respective multiple-benchmark tracking-error portfolio. The one-sided p-values are computed based on the studentized circular block bootstrapping method used in Ledoit & Wolf (2008). From the table, we observe that the MBTE portfolio consistently dominates the benchmark portfolio used in its construction. In particular, the MBTE portfolio has higher net Sharpe ratios than the $1/n$ portfolio and all the differences are significant at 0.005 level for all except the "8Stock" data set. Moreover, the MBTE portfolio has higher net Sharpe ratios than the MUL portfolio across all the data sets, and all the differences are significant at 0.05 level. The MBTE portfolio also

10Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.0238	0.0137	0.9075	0.4094	0.1281	0.0820	0.0962	0.0627	0.0900	0.0593
	(42.44%)		(54.89%)		(35.99%)		(34.82%)		(34.11%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.0801	0.0531	0.0773	0.0514	0.0262	0.0184	0.0238	0.0133	0.1852	0.1292
	(33.71%)		(33.51%)		(29.77%)		(45.72%)		(37.3%)	
48Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.0306	0.0170	2.4015	0.9243	0.4632	0.2729	0.4459	0.2591	0.3026	0.1835
	(44.44%)		(61.51%)		(41.08%)		(41.89%)		(39.36%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.3004	0.1809	0.2815	0.1705	0.0266	0.0267	0.0304	0.0165	1.4778	0.9257
	(39.78%)		(39.43%)		(−0.38%)		(45.72%)		(37.36%)	
25FF	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.0174	0.0120	3.1537	1.2923	0.4368	0.2699	0.4278	0.2633	0.3520	0.2193
	(31.03%)		(59.02%)		(38.21%)		(38.45%)		(37.70%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.2569	0.1644	0.2532	0.1628	0.0347	0.0286	0.0173	0.0099	1.0400	0.6760
	(36.00%)		(35.70%)		(17.58%)		(42.77%)		(35.00%)	
8Stock	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.0406	0.0222	0.4034	0.1714	0.0540	0.0360	0.0381	0.0289	0.0373	0.0285
	(45.32%)		(57.51%)		(33.33%)		(24.15%)		(23.59%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.0374	0.0282	0.0387	0.0278	0.0371	0.0274	0.0407	0.0217	0.0511	0.0371
	(24.60%)		(28.17%)		(26.15%)		(46.68%)		(27.40%)	

Note: This table reports the turnover rate and the corresponding percentage deduction in turnover rate by the MBTE portfolio (in brackets).

Table 3.3.4: Comparison on turnover rate

outperforms the PARR portfolio across all the data sets, and the differences are significant at 0.005 level in all but the "8Stock" data set. Additionally, the MBTE portfolio shows significant difference from the MEAV, GMC, M1FAC, and CVAR portfolios in the "25FF" data set. It is worth noting that except the case when CVAR is used as one benchmark, all the other MBTE portfolios have much higher net Sharpe ration than the $1/n$ portfolio, independent of the choice on the benchmark portfolio, in all the data sets we consider.

Net Aggregate Return

Table 3.3.6 reports the out-of-sample net aggregate return and the corresponding percentage increment in net aggregate return by the MBTE portfolio. From the table, we observe that the MBTE portfolios almost always dominate their respective benchmark portfolios over the whole investment period (07/1983–12/2011 for the "10Ind", "48Ind", "25FF" data sets, and 08/2000–01/2013 for the "8Stock" data set). In particular, the MBTE portfolio shows great superiority over the PARR portfolio. The net aggregate return of the MBTE portfolio is twice as large as that of the PARR portfolio in the "10Ind" data set, nearly 10 times larger in the "48Ind" portfolio, and 37 times larger in the "25FF" data set. Furthermore, the MBTE portfolio outperforms the CVAR portfolio by 88.6% in the "25FF" data set and the MINU portfolio by 38.6% in the same data set. However, the net aggregate return of the MBTE portfolio does not always outperform the GMC portfolio or the MUL portfolio, though the difference is small (of order 0.1%) in these two cases.

After all, these aggregate returns only provide partial information as we only consider one investment horizon. To demonstrate the robustness of the findings, we consider random starting times and random lengths of the investment horizon. Some results for the "48Ind" data set are provided in Figure 3.3.4 and 3.3.5[5].

[5]We have done similar tests on all the other data sets, and tried many other random starting points

10Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.2260	0.2275	0.1941	0.2582	0.2809	0.2858	0.2915	0.2957	0.2929	0.2966
	(0.0050***)		(0.0005***)		(0.4785)		(0.2018)		(0.2483)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.2919	0.2950	0.2919	0.2948	0.2616	0.2626	0.2274	0.2285	0.2536	0.2593
	(0.3322)		(0.3222)		(0.1923)		(0.0290*)		(0.1818)	
48Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.1805	0.1857	0.0759	0.2383	0.2560	0.2747	0.2277	0.2447	0.2474	0.2553
	(0.0005***)		(0.0015***)		(0.0790)		(0.3617)		(0.2922)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.2608	0.2678	0.2588	0.2640	0.2693	0.2704	0.1832	0.1877	0.1009	0.1450
	(0.3237)		(0.3477)		(0.4231)		(0.0025***)		(0.0015***)	
25FF	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.1999	0.2030	-0.02499	0.2530	0.3803	0.4046	0.3720	0.3965	0.3590	0.3791
	(0.0010**)		(0.0005***)		(0.0005***)		(0.0005***)		(0.0020***)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.3481	0.3615	0.3587	0.3716	0.2038	0.2064	0.2016	0.2039	0.3776	0.4115
	(0.0345*)		(0.0375*)		(0.0065*)		(0.0005***)		(0.0195*)	
8Stock	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	0.1807	0.1836	0.1982	0.2074	0.1773	0.1814	0.2551	0.2637	0.2567	0.2653
	(0.2253)		(0.3515)		(0.1514)		(0.1893)		(0.1963)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	0.2562	0.2649	0.2491	0.2574	0.2681	0.2756	0.1853	0.1854	0.2348	0.2443
	(0.1714)		(0.1504)		(0.1753)		(0.036*)		(0.0924)	

Note: This table reports the monthly out-of-sample Sharpe ratio net of proportional transaction costs of 50 basis points and the corresponding one-sided p-value (in brackets) that the net Sharpe ratio of each benchmark strategy is smaller than that of the respective MBTE portfolio. Star symbols are included for p-values: (.*) for significance at 0.05 level, (.**) for 0.01, and (.***) for 0.005.

Table 3.3.5: Comparison on net Sharpe ratio

10Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	19.2912	19.7050	8.8949	18.5599	26.7493	28.9076	31.5124	32.8691	31.4552	32.6069
	(2.15%)		(108.66%)		(8.07%)		(4.31%)		(3.66%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	30.5331	31.5403	30.2300	31.1623	22.8364	23.2083	19.5630	19.8028	21.9975	22.8334
	(3.30%		(3.08%)		(1.63%)		(1.23%)		(3.80%)	
48Ind	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	15.0473	16.5383	2.0602	22.4944	22.6779	29.4095	18.5407	19.2705	17.7096	18.6572
	(9.91%)		(991.86%)		(29.68%)		(3.94%)		(5.35%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	19.3608	20.3869	18.5407	19.2705	24.9412	24.8306	15.6788	16.8206	16.1426	19.2452
	(5.30%)		(3.94%)		(-0.44%)		(7.28%)		(19.22%)	
25FF	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	20.5105	21.8385	0.5317	20.2806	97.0287	134.3689	87.0998	120.7095	75.695	98.1391
	(6.47%)		(3714.29%)		(38.48%)		(38.59%)		(29.65%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	65.7026	79.5398	73.8475	89.0208	17.4811	18.1652	20.9800	21.7274	234.5823	442.5228
	(21.06%)		(20.55%)		(3.91%)		(3.56%)		(88.64%)	
8Stock	$1/n$	MBTE	PARR	MBTE	MEAV	MBTE	MINU	MBTE	M1FAC	MBTE
	2.0784	2.0834	2.0900	2.1405	1.9623	1.9896	2.6711	2.7580	2.6878	2.7754
	(0.24%)		(2.42%)		(1.39%)		(3.25%)		(3.26%)	
	MIND	MBTE	M1PAR	MBTE	GMC	MBTE	MUL	MBTE	CVAR	MBTE
	2.6622	2.7484	2.5553	2.6308	2.7881	2.8705	2.1102	2.0995	2.5372	2.6398
	(3.24%)		(2.95%)		(2.96%)		(-0.51%)		(4.04%)	

Note: This table reports the out-of-sample net aggregate return and the corresponding percentage increment in net aggregate return by the MBTE portfolio (in brackets).

Table 3.3.6: Comparison on net aggregate return

In both Figure 3.3.4 and 3.3.5, we observe that the net aggregate returns of our multiple-benchmark tracking-error portfolios are constantly higher than or comparable to those of the $1/n$ portfolio. Figure 3.3.4 shows that the shorter the investment periods, the less difference in net aggregate returns between the MBTE portfolio and the $1/n$ portfolio. When the investment activity is conducted for the whole 28 years, we observe a clear difference between the net aggregate returns from these two portfolios.

Note that the MBTE portfolios in Figure 3.3.5 are obtained using the PARR and buy-and-hold portfolios as benchmarks, which do not contain the $1/n$ portfolio, but the performance of the $1/n$ portfolio at the same time periods are included for comparison. It is interesting to observe that although the PARR portfolio dominates the $1/n$ portfolio in terms of Sharpe ratio, when the transaction costs are considered, the resulting performance of the PARR portfolio is usually worse than the $1/n$ portfolio. However, if we put the PARR portfolio together with the buy-and-hold strategy in the set of benchmark portfolios, our model yields a new portfolio that combines the strength of both portfolios. The resulting portfolio provides a high level of return while requiring much less transaction, and the net aggregate returns clearly outperform both the $1/n$ portfolio and the PARR portfolio.

Remark 3.15. It is not so intuitive that why our model, which is a single-period model, performs very well in terms of long-run wealth growth. Interestingly, the approach of minimizing the tracking-error volatility is closely related to a substream of on-line portfolio selection models based on aggregate return maximization and the concept of the universal portfolio, which is introduced by Cover (1991) and discussed further in Helmbold et al. (1998). These portfolio selection models apply the logarithm transform to the aggregate return and use the Taylor series expansion to approximate the

with random investment horizons. As the findings are similar, we only report part of the results here.

objective function, provided that the portfolio return is sufficiently small, i.e.,

$$\mathbf{E}\left[\log\left(1+\tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)\right] \cong \mathbf{E}\left[\tilde{\boldsymbol{r}}^T\boldsymbol{x} - \frac{\left(\tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2}{2}\right].$$

Then maximizing the approximate aggregate return is equivalent to minimizing

$$\mathbf{E}\left[\left(1-\tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right],$$

i.e., using 1 as the benchmark when the returns are sufficiently small. This connection partially explains why our model usually yields very high aggregate return over an investment horizon even though it is a single-period model.

3.3.6.2 Multiple-Benchmark Tracking-Error portfolio with Penalty on Transaction Volume

We dedicate this part to evaluate the alternative way to control transaction volume as proposed in Section 3.2.6, in which we penalize the transaction volume directly in the objective function. In particular, we solve Problem (TC) to obtain the multiple-benchmark tracking-error portfolio with a penalty on transaction volume (MBTEP). We choose the $1/n$, M1FAC, and CVAR portfolios as benchmarks. In choosing the penalty parameter, ν, we use an in-sample calibration approach, where the turnover of the multiple-benchmark tracking-error portfolio is restricted to be less than that of the $1/n$ portfolio in the last period of estimation window. We fix $W = 1$ throughout the investment horizon to facilitate the search of ν.

Table 3.3.7 presents a comparison between our multiple-benchmark tracking-error portfolio (MBTEP) and the $1/n$ portfolio. From the table, we observe that the in-sample tracking errors of the MBTEP portfolio are consistently smaller than that of the

$1/n$ portfolio across all the data sets. The net Sharpe ratios of our MBTEP portfolio outperform the $1/n$ portfolio across all the data sets, and the difference is statistically significant in all but the "8Stock" data set. Furthermore, over the whole investment period of the data sets, the MBTEP portfolio tends to yield higher net aggregate returns than the $1/n$ portfolio[6]. However, the turnover rates of the MBTEP portfolio are always higher than those of the $1/n$ portfolio. This is expected as the in-sample calibration of the penalty parameter might induce large out-of-sample turnovers.

Overall, the MBTEP portfolio provides better results than the $1/n$ portfolio. The flexibility in choosing a value for the penalty parameter could be either a bonus or a burden as determining the value is a judgment call of the portfolio manager. The desired performance can only be induced by appropriately chosen penalty values. Adding the buy-and-hold strategy into the set of benchmarks seems to be more natural and effective in controlling the transaction volume.

3.4 Conclusion

We propose a new multiple-benchmark tracking-error model for portfolio selection. The target return being tracked is the highest return from a set of given benchmark portfolios. Our model differs from existing literature by directly capturing the concerns arising from real investment activities. By resorting to Stein's identity, we obtain the closed-form expression for the optimal portfolio weights under the assumption of normal return distribution. The closed-form solution reveals that persistency is the basic component of the optimal portfolio, which partially explains the robustness of our portfolio against the estimation errors, as the probability of one benchmark

[6]We have also conducted a robustness test in this case. As the results are similar as before, we do not report it here.

Data Set	Portfolio Model	Tracking Error	Net Sharpe Ratio	Turnover Rate	Net Aggregate Return
10Ind	$1/n$	0.1130	0.2260	0.0238	19.2912
	MBTEP	0.09597	0.2364	0.0370	19.3359
		(15.03%)	(0.0460^{*})	(-35.68%)	(0.23%)
48Ind	$1/n$	0.7323	0.1805	0.0306	15.0473
	MBTEP	0.6160	0.2224	0.0723	18.5023
		(15.89%)	(0.0050^{***})	(-57.68%)	(22.96%)
25FF	$1/n$	0.4768	0.1999	0.0174	20.5105
	MBTEP	0.4149	0.2317	0.0365	29.8953
		(12.97%)	(0.0040^{***})	(-52.33%)	(45.76%)
8Stock	$1/n$	0.02817	0.1807	0.0406	2.0784
	MBTEP	0.01712	0.2126	0.0493	2.2427
		(39.22%)	(0.1169)	(-17.65%)	(7.91%)

Note: The number inside the brackets under "Tracking Error" column is the corresponding percentage decrease in in-sample tracking error by the MBTE portfolio from the $1/n$ portfolio. The number inside the brackets under "Net Sharpe Ratio" column is the corresponding one-sided p-value that the net Sharpe ratio of the $1/n$ portfolio is smaller than that of the respective MBTE portfolio. Star symbols are included for p-values: $(.^{*})$ for significance at 0.05 level, $(.^{**})$ for 0.01, and $(.^{***})$ for 0.005. The number inside the brackets under "Turnover Rate" column is the corresponding percentage deduction in turnover rate by the MBTE portfolio from the $1/n$ portfolio. Negative numbers indicate increased turnover rate. The number inside the brackets under "Net Aggregate Return" column is the corresponding percentage increment in net aggregate return by the MBTE portfolio from the $1/n$ portfolio.

Table 3.3.7: Comparison on the performance of the $1/n$ portfolio and the multiple-benchmark tracking-error portfolio with penalty on transaction volume (MBTEP)

outperforming the rest are less prone to estimation errors on the expected returns of the financial assets.

The closed-form solution allows us to conduct more in-depth analysis of our model, especially, the comparison with the Markowitz mean-variance portfolio and the linear combination rule proposed by Tu & Zhou (2011). In particular, we show that the linear combination rule would be inferior to our portfolio if the portfolio manager has a mean-variance utility with low risk aversion. This further strengthens the motivation

of our multiple-benchmark tracking-error model. In addition, we prove that the entire mean-variance efficient frontier can be generated from our model when two distinct mean-variance portfolios are used as the benchmark portfolios, a result similar to the well-known Two-Fund Theorem.

To address the common problem of whether to reposition the portfolio, our modeling framework allows a natural solution by including the buy-and-hold strategy as one of the benchmark portfolios. Our numerical analysis shows that adding the buy-and-hold strategy as a benchmark can significantly reduce the turnover rate, which might be attractive to investors when transaction costs are considerable. When combining the buy-and-hold strategy with other benchmarks, we demonstrate using the real data sets that our portfolio has consistently provided higher net Sharpe ratio, higher net aggregate return, and lower turnover rate compared to the benchmark portfolios, in particular, the $1/n$ portfolio, a well-known tough benchmark to beat.

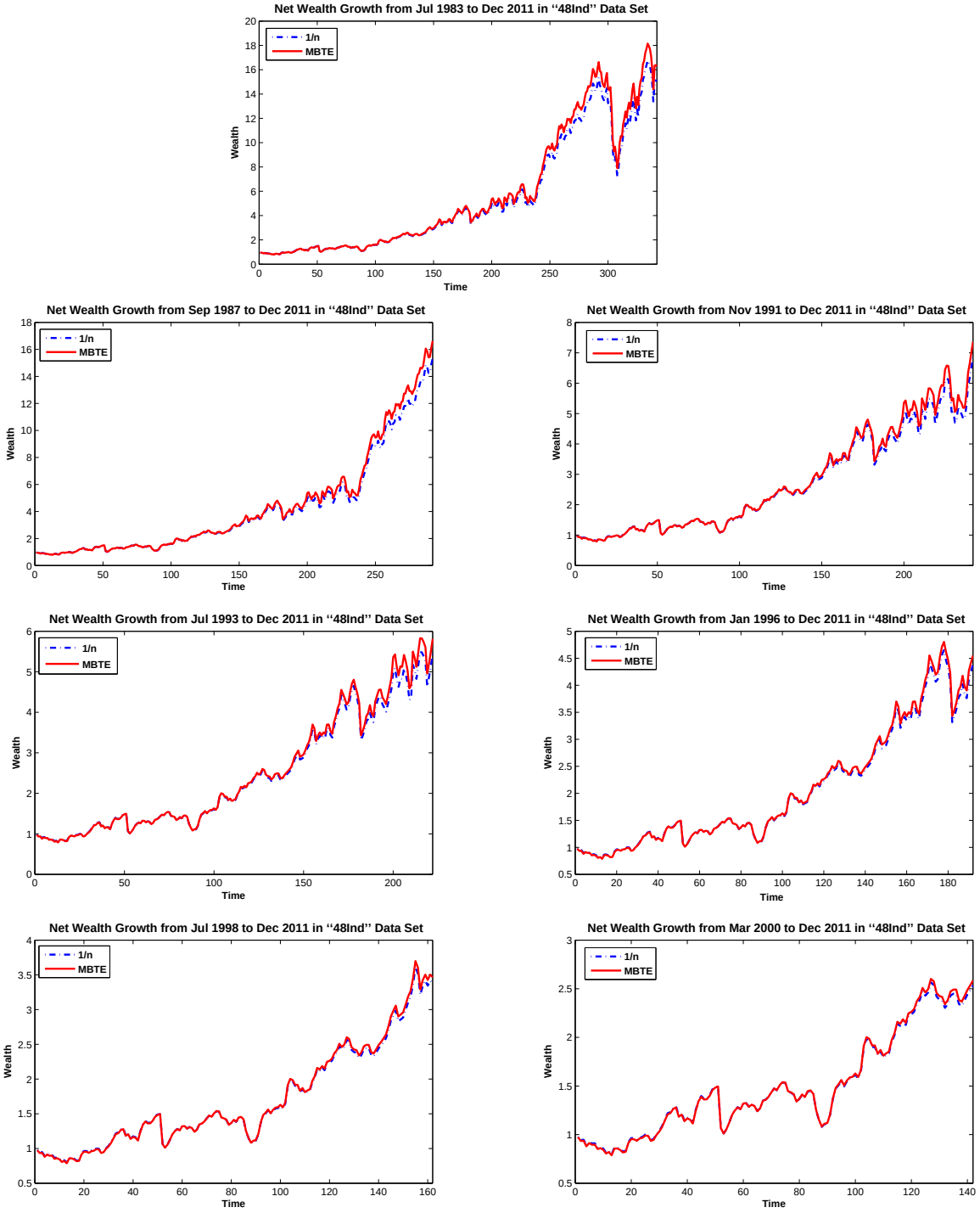


Figure 3.3.4: Wealth growth of the multiple-benchmark tracking-error (MBTE) portfolio using the $1/n$ and buy-and-hold portfolios as benchmarks, and the $1/n$ portfolio with random starting times and evaluation periods in the "48Ind" data set

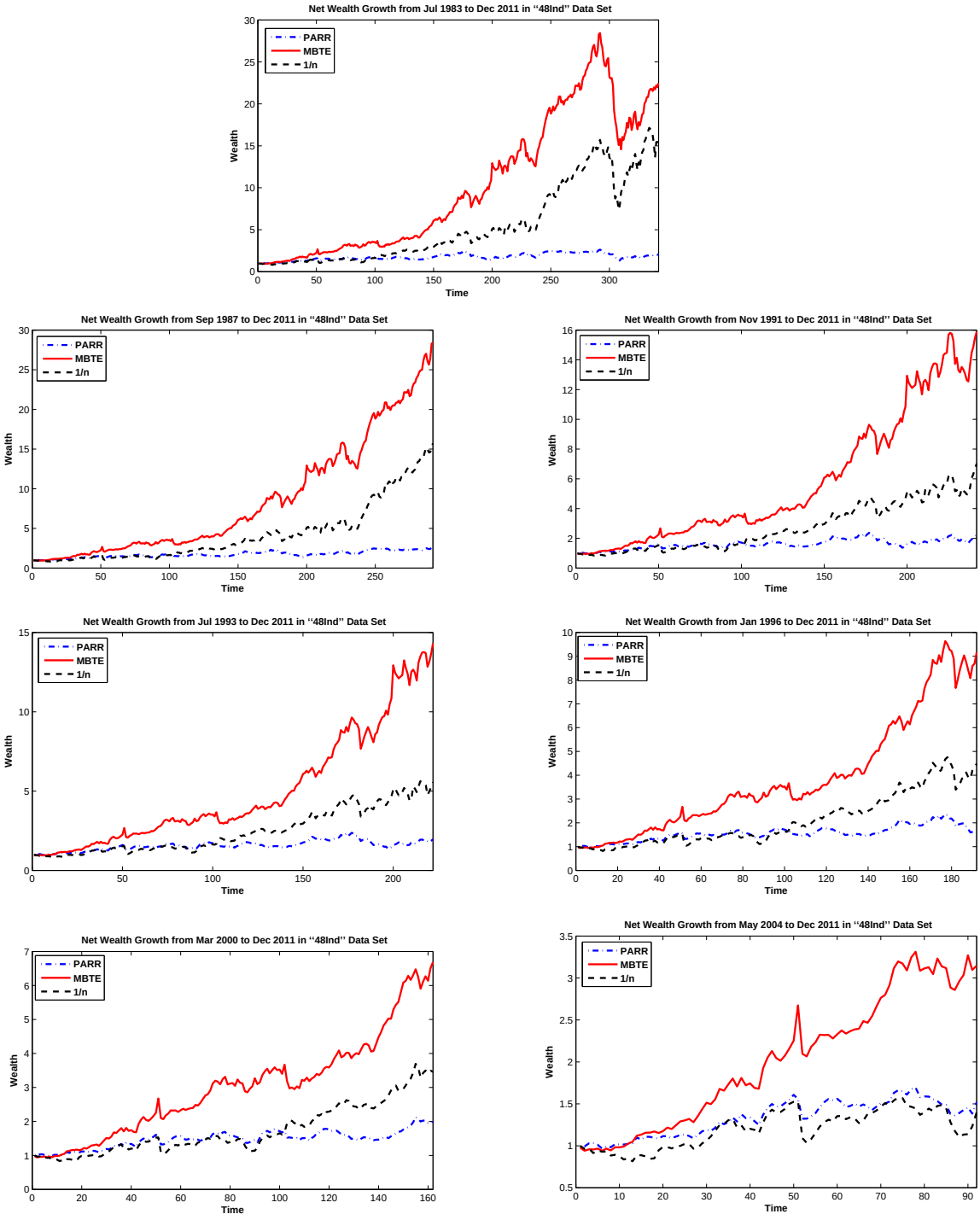


Figure 3.3.5: Wealth growth of the multiple-benchmark tracking-error (MBTE) portfolio using the PARR and buy-and-hold portfolios as benchmarks, the $1/n$ portfolio, and the PARR portfolio with random starting time for evaluation period in the "48Ind" data set

Chapter 4

Summary and Discussions

In previous chapters, we have shown how to transform two classes of problems into the corresponding persistency estimation problems. In this chapter, we first review the main results from the previous two chapters. Next, we will discuss several directions for potential applications and future research. Last but not least, we will briefly discuss how to improve the persistency estimation to wrap up the thesis.

4.1 Review and Discussions

4.1.1 Least Squares Distribution Approximation

In Chapter 2, we discuss the problem to estimate the distribution of the outcome from a stochastic optimization system. When such optimization system can be represented as a mixed zero-one linear programming problem with uncertainty in the objective coefficients, i.e.,

$$Z\left(\tilde{\boldsymbol{c}}\right) := \max_{\boldsymbol{x}\in\mathcal{P}} \sum_{j=1}^{n} \tilde{c}_j x_j,$$

where

$$\mathcal{P} := \{\boldsymbol{x} \in \mathbb{R}^n \,:\, \boldsymbol{a}_i^T \boldsymbol{x} = b_i,\, \forall i = 1, \ldots, m;\; x_j \in \{0,1\},\, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\};\; \boldsymbol{x} \geq \boldsymbol{0}\},$$

we developed a least squares framework to approximate the distribution of its optimal objective value. The approximating distribution could be a linear or a quadratic function of the stochastic objective coefficients, i.e.,

$$W\left(\tilde{\boldsymbol{c}}\right) = \alpha + \sum_{j=1}^{n} \beta_j \left(\tilde{c}_j - \mu_j\right), \text{ or}$$

$$Q(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n} \beta_j (\tilde{c}_j - \mu_j) + \sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma_{j_1,j_2} (\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2}).$$

The motivation to extend to a quadratic function is to capture the skewness in the optimal objective value. We illustrated the importance and effectiveness of the quadratic approximation by several case studies. The parameters in the approximating functions are chosen to minimize the expected squared deviation of the approximating distribution from the true distribution of the optimal objective value, i.e.,

$$\min_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n} \mathbf{E}\left[\left(Z\left(\tilde{\boldsymbol{c}}\right) - \alpha - \sum_{j=1}^{n} \beta_j \left(\tilde{c}_j - \mu_j\right)\right)^2\right], \text{ or}$$

$$\min_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \Gamma \in \mathbb{R}^{n \times n}} \mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha - \sum_{j=1}^{n} \beta_j (\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma_{j_1,j_2} (\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\right)^2\right].$$

When the objective coefficients follow a nondegenerate multivariate normal distribution, we derived the closed-form expression of the approximating distribution by resorting to Stein's Identity. It turned out that persistency plays an important role in

the optimal approximating distributions. In particular, for both linear and quadratic approximations, the optimal linear parameters are just the persistency values, i.e.,

$$\beta_j^* = \mathbf{E}\left[x_j(\tilde{\boldsymbol{c}})\right], \; j = 1, \ldots, n.$$

We also extended our analysis to other distributions of the random objective coefficients. In particular, we obtained the closed-form expression of the linear approximation when the random objective coefficients follow a multivariate skew-normal distribution by exploiting the extension of the Stein's Identity under such distribution.

We presented the application of these results in three problems: maximum partial sum, project management, and statistical timing analysis. Exploiting the existing results from literature, we derived the closed-form expression to approximate the distribution of the maximum partial sum for several special cases. In project management problem, persistency is well-known by another name, criticality index. Our results on the closed-form solutions tell that the criticality index can be also used to estimate the variability in the project completion time. We demonstrated by numerical analysis that the criticality index is very effective in capturing the variance of the project completion time, and also that the overall approximation accuracy of our approach is significantly higher than most existing methods. Moreover, our model differs from the existing literature by its ability to capture correlations among activity completion times, which can severely impact the project completion time. Finally, we illustrated the application of our results in statistical timing analysis and tested the robustness of our approximation method through a case with extreme distributional parameters.

4.1.2 Quadratic Regret Strategy

Next, in Chapter 3, we discussed the problem of decision making under uncertainty, and proposed a new decision criterion, named quadratic regret. In particular, we sought to make a decision whose outcome has the minimum expected squared deviation from the best possible alternative, when the outcome depends linearly on the decision and the linear coefficients are stochastic. Mathematically, we considered the following problem:

$$\min_{\boldsymbol{x}\in\mathcal{P}} \mathbf{E}\left[\left(Z\left(\tilde{\boldsymbol{c}}\right)-\tilde{\boldsymbol{c}}^T\boldsymbol{x}\right)^2\right],$$

where $\mathcal{P}$ is the feasible space of the decision as defined before, and $Z\left(\tilde{\boldsymbol{c}}\right)$ denotes the best possible outcome, i.e.,

$$Z\left(\tilde{\boldsymbol{c}}\right) := \max_{\boldsymbol{y}\in\mathcal{P}} \tilde{\boldsymbol{c}}^T\boldsymbol{y}.$$

We illustrated this decision criterion using the example of portfolio management problem, where the portfolio manager needs to divide the capital into a pool of financial assets before knowing the rates of return of all the assets. The objective is to maximize the investment return while keeping the risk as low as possible. In the portfolio management problem, a popular strategy in practice is to track the outcome of a certain benchmark portfolio. The deviation of the portfolio return from the benchmark return is called tracking error, and the portfolio selection models that minimize the volatility of the deviation are referred to as benchmark tracking-error models. We proposed a new tracking-error model for portfolio selection that tracks multiple benchmarks. We defined the tracking error under multiple benchmarks as the return deviation from the highest possible return among the benchmark portfolios. In particular, we solved the following problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathbf{E}\left[\left(Z_B(\tilde{\boldsymbol{r}})-\tilde{\boldsymbol{r}}^T\boldsymbol{x}\right)^2\right].$$

where $\mathcal{X}$ is the set of feasible portfolios, and $\tilde{\boldsymbol{r}}$ denotes the random asset return, and $Z_B(\tilde{\boldsymbol{r}})$ represents the highest return from a set of m benchmark portfolios, $\{\boldsymbol{p}^1, \ldots, \boldsymbol{p}^m\}$, i.e.,

$$Z_B(\tilde{\boldsymbol{r}}) := \max_{j \in \{1,\ldots,m\}} \tilde{\boldsymbol{r}}^T \boldsymbol{p}^j.$$

Under the assumption that the return follows a nondegenerate multivariate normal distribution, we derived the closed-form expression of the optimal multiple-benchmark tracking-error portfolio when $\mathcal{X} = \left\{\boldsymbol{x} \in \mathbb{R}^n \,:\, \boldsymbol{e}^T\boldsymbol{x} = 1\right\}$ as follows:

$$\sum_{j=1}^{m} \mathbf{E}\left[p_j\left(\tilde{\boldsymbol{r}}\right)\right] \boldsymbol{p}^j + \Sigma^{-1}\left(\frac{C\boldsymbol{\mu} - B\boldsymbol{e}}{(A+1)C - B^2}\right)\left(\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] - \sum_{j=1}^{m} \mathbf{E}\left[p_j\left(\tilde{\boldsymbol{r}}\right)\right] \boldsymbol{\mu}^T \boldsymbol{p}^j\right),$$

where $A = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$, $B = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{e}$, and $C = \boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}$. Similar to the distribution approximation problem, this was achieved by applying the Stein's Identity. The closed-form expression shows that the optimal multiple-benchmark tracking-error portfolio is formed by the persistency weighted benchmark portfolio with an adjustment term. In this case, the persistency of a benchmark portfolio, $\mathbf{E}\left[p_j\left(\tilde{\boldsymbol{r}}\right)\right]$, is the probability that this benchmark portfolio yields the highest return among all the benchmark portfolios.

The closed-form solution suggests an interesting connection of our model to a well-known behavioural abnormality, probability matching. It is evident from numerous experiments that when making repeated decisions under uncertainty, rather than choosing the strategy that maximizes the expected utility, human subjects tend to adopt mixed strategies where the mixing probabilities converge to the probabilities of the best possible option, i.e., persistency. In our portfolio selection problem setting, the objective is to track the highest return from a pool of benchmarks. If the problem is repeatedly presented to the portfolio manager, and the manager can only form the portfolio exactly as one of the benchmarks, the probability matching theory would

predict the manager to mix his or her choices over the periods, and the choice probability of a particular benchmark would match the probability that this benchmark gives the highest return, which is exactly its persistency. However, in our case, the investment decision is made only once, but allowed to be continuous. A natural extension from the probability matching theory would yield the decision that combines the mixed strategies based on their mixing probabilities, which gives exactly the first term in our closed-form solution. Moreover, as reported in many probability matching experiments, the matching probabilities are usually not exactly the same as the persistency values, and they are affected by several factors, like the amount of incentive, etc. The adjustment term in the closed-form solution surprisingly describes such effect to a certain extent. To see this, note that the term, $\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right] - \sum_{j=1}^{m} \mathbf{E}\left[p_j\left(\tilde{\boldsymbol{r}}\right)\right] \boldsymbol{\mu}^T \boldsymbol{p}^j$, represents the shortfall of the expected return of the persistency weighted portfolio, $\sum_{j=1}^{m} \mathbf{E}\left[p_j\left(\tilde{\boldsymbol{r}}\right)\right] \boldsymbol{p}^j$, from the expected return of our target, $\mathbf{E}\left[Z_B(\tilde{\boldsymbol{r}})\right]$. Therefore, if the i^{th} asset has relatively higher expected return, i.e., μ_i is relatively larger, our optimal portfolio will have a positive adjustment in the weight of the i^{th} asset due to the effect from the term, $C\boldsymbol{\mu} - B\boldsymbol{e}$. For our model with target expected return constraint ($\boldsymbol{\mu}^T\boldsymbol{x} = K$), the optimal solution reads

$$\sum_{j=1}^{m} \mathbf{E}\left[p_j\left(\tilde{\boldsymbol{r}}\right)\right] \boldsymbol{p}^j + \Sigma^{-1}\left(\frac{C\boldsymbol{\mu} - B\boldsymbol{e}}{AC - B^2}\right)\left(K - \sum_{j=1}^{m} \mathbf{E}\left[p_j\left(\tilde{\boldsymbol{r}}\right)\right] \boldsymbol{\mu}^T \boldsymbol{p}^j\right),$$

which is also composed of the persistency weighted benchmark portfolio plus an adjustment term. Note that this adjustment term is exactly the optimal solution to Problem (M_0) on page 111 with the target expected return $K' = K - \sum_{j=1}^{m} \mathbf{E}[p_j(\tilde{\boldsymbol{r}})]\boldsymbol{\mu}^T\boldsymbol{p}^j$. Problem ($\mathrm{M}_0$) essentially finds the minimum-variance adjustment portfolio ($\boldsymbol{e}^T\boldsymbol{x} = 0$) with expected return K' ($\boldsymbol{\mu}^T\boldsymbol{x} = K'$). Therefore, the optimal multiple-benchmark tracking-error portfolio at target expected return K is the persistency weighted bench-

mark portfolio plus the minimum-variance adjustment to bring the expected return to the level of K.

Overall, the closed-from expressions of our portfolios turn out to have an almost perfect description from the perspective of probability matching behaviour. On the other hand, this result also sheds some light on the emergence of the probability matching behaviour. Our findings tell that if the subject makes the decision based on minimizing the expected quadratic regret, his or her decision would exhibit the pattern of probability matching. This partially coincides with some existing results and conjectures that suggests the probability matching behaviour is closely related to the mindset of regret minimization (cf. Simon (1956)).

Furthermore, the closed-form solution allows us to conduct more in-depth analysis of our model, especially, the comparison with the Markowitz mean-variance portfolio and the linear combination rule proposed by Tu & Zhou (2011). In particular, we showed that the linear combination rule would be inferior to our portfolio if the portfolio manager has a mean-variance utility with low risk aversion. This further strengthens the motivation of our multiple-benchmark tracking-error model. In addition, we prove that the entire mean-variance efficient frontier can be generated from our model when two distinct mean-variance portfolios are used as the benchmark portfolios, a result similar to the well-known Two-Fund Theorem.

To address the common problem of whether to reposition the portfolio, our modeling framework allows a natural solution by including the buy-and-hold strategy as one of the benchmark portfolios. Our numerical analysis showed that adding the buy-and-hold strategy as a benchmark can significantly reduce the turnover rate, which might be attractive to investors when transaction costs are considerable. When combining the buy-and-hold strategy with other benchmarks, we demonstrated using the real

data sets that our portfolio has consistently provided higher net Sharpe ratio, higher net aggregate return, and lower turnover rate compared to the benchmark portfolios, in particular, the $1/n$ portfolio, a well-known tough benchmark to beat.

4.2 Future Research

Besides the applications demonstrated in Chapter 2 and 3, the methodology developed in this thesis could be applied to many other problem settings. In what follows, we will discuss three areas of potential application.

4.2.1 Structural Calibration and Prediction

The first promising direction for future research will probe into the area of business analytics. What was done in Chapter 2 is closely related to the regression analysis. If we view $Z(\tilde{\boldsymbol{c}})$ as the dependent variable, and $\tilde{\boldsymbol{c}}$ as the independent variable, the problem we tried to solve is to predict the dependent variable using a linear or a quadratic function of the independent variable. The key differences with the traditional multivariate linear regression are that in our case, we know that the dependent variable is an optimization outcome of the independent variables, and that the independent variables are allowed to be correlated. Thus, one potential research question is how to calibrate the optimization model using the observed independent and dependent variables without knowing the details of the optimization process, i.e., the functional form of $Z(\tilde{\boldsymbol{c}})$. A promising application is in the predictive modeling for consumer choice, where only consumer's past choices are observed but not the utility function of the consumer. The problem is to calibrate the utility model in order to predict consumer's future choices.

In another case, we can exploit the results in Chapter 2 to help improve the regression analysis. Consider the example of predicting the service level in an clinic or a hospital, where the patients have to wait in a queue for the service provided by the doctors. In particular, we are interested in how long the patients will stay in the system, i.e., the patient's sojourn time in the system. Data available for such analysis usually include patient's characteristics (like demographic data, clinical conditions, etc.) and the clinic's or the hospital's operational records (like queue length, arrival time, departure time, etc.). Typical approach is to apply the multivariate linear regression, and try to predict the patient's sojourn time using the available data. Then the conclusion is drawn based on the statistical significance of related parameters. One of the most important assumptions in the typical linear regression analysis is uncorrelated noise terms, which unfortunately fails in this problem setting. The sojourn times of the patients served by the same doctor are correlated through the service sequence. In particular, the sojourn time of a patient will be affected by the sojourn time of the previous patient, similarly for the noises in the sojourn times. Indeed, the relationship between these two sojourn times can be expressed through a certain function involving the maximum operator, which can be represented as an optimization process. Therefore, the estimation method developed in Chapter 2 can be applied to simplify such relationship by writing the sojourn time noise of a patient as a linear function of the sojourn time noise of the previous patient and the idiosyncratic noise of this patient. The parameters of the linear function can be easily estimated from data. With such modification, we can refine the linear regression model, and hopefully improve the prediction.

4.2.2 Two-Stage Stochastic Programming

The methodology developed in the distribution approximation problem can be also applied to many two-stage stochastic optimization problems. In particular, if the second-stage problem can be formulated as a mixed zero-one linear optimization problem with uncertainty in the objective, we can use the results in Chapter 2 to approximate the distribution of the second-stage costs. Armed with the distributional knowledge on the second-stage cost, we can then conduct more in-depth risk analysis or parameter calibration for the first-stage problem. Potential applications include various two-stage problems in supply chain management as well as principal-agent models in game theory.

For example, in many supply chain coordination models under the principal-agent framework, the principal needs to design the contract in the first stage while some parameters remain unknown to the principal, such as the demand information. In the second stage, the uncertain parameters realize their values, and the agent learns these values and responses optimally to the contract designed by the principal in the first stage. The payoffs to the principal and the agent are then realized. The challenge for the principal is to design the contract in the first stage such that his or her payoff is maximized. Conventional objective adopted in literature is to maximize the principal's expected payoff. Such objective can be justified when the problem is repeated in many periods and the principal is not allowed to redesign the contract in different periods. However, the risk involved in the principal's payoff makes the objective based on expectation solely inefficient if the problem happens only once or a few times. The difficulty of adopting different risk measures on the principal's payoff partially comes from the difficulty in assessing the distribution of the principal's payoff, which depends on the agent's optimal response. If we can apply the results in Chapter 2 and construct

some approximating distributions for the principal's payoff, we can certainly conduct more risk analysis on the principal's contract design. Consequently it will help the principal design a better contract that tailors to his or her needs.

However, it might still be difficult to optimize the first-stage decision because of the challenge to predict the change in the approximating distribution as the first-stage decision changes. Nonetheless, the potential applications of the method in various two-stage stochastic optimization problems are definitely worth future research.

4.2.3 Quadratic Regret Solution

Besides the portfolio selection problem, the concept of quadratic regret solution can be brought into many other stochastic optimization problems. The connection between this decision criterion with the probability matching behaviour through persistency makes it worthwhile for further investigation. To provide a glimpse into the feature of the quadratic regret solution in other problem settings, we present a preliminary study on a stochastic shortest path problem in the next example.

Example 4.1. Consider the shortest path problem in the transportation network as shown in Figure 4.2.1, where Node 1 is the origin and Node 3 is the destination. There are three possible paths to reach Node 3 from Node 1. Denote the path using Arc $(1, 3)$ as Path A. Denote the two paths going through Node 2 as Path B and Path C. The transportation times on Arc $(1, 2)$ and either arc connecting Node 2 and 3 are normally distributed with mean 5 and standard deviation 2, while the transportation time on Arc $(1, 3)$ is normally distributed with mean $(10 - \epsilon)$ and standard deviation $2\sqrt{2}$, where $\epsilon > 0$. The transportation times on all the arcs are uncorrelated. Means and variances of different paths are summarized in Table 4.2.1 for the case of $\epsilon = 0.1$.

Thus, for any $\epsilon > 0$, the path with the shortest expected transportation time is

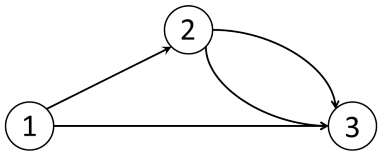


Figure 4.2.1: Transportation network in Example 4.1

Path	Mean	Variance	Prob(Shortest)	Prob(Longest)	ESD
A $(1 \to 3)$	9.9	8	40%	38%	11.10
B $(1 \to 2 \to 3)$	10	8	30%	31%	10.61
C $(1 \to 2 \to 3)$	10	8	30%	31%	10.61

Table 4.2.1: Stochastic parameters in Example 4.1 with $\epsilon = 0.1$

Path A. Furthermore, since all the transportation times are normally distributed, the transportation time of Path A is less than any other path in the usual stochastic order. However, if we solve for the quadratic regret solution where the regret is measured as the difference in transportation time between the path chosen and the shortest path, then either Path B or Path C would be optimal for small ϵ. In addition, as $\epsilon \to 0$, the fractional solution will converge to the persistency values of the arcs, i.e., the probabilities that the arcs are on the shortest path.

To better understand the quadratic regret solution, Table 4.2.1 reports several stochastic parameters of the problem when $\epsilon = 0.1$. First of all, the quadratic regret solution indeed has a smaller expected squared deviation from the shortest path. Although Path A has larger probability of being the shortest, the collective probability that either Path B or Path C is shortest is larger, i.e., 60%. Due to the positive correlation between Path B and Path C as a result of a common arc, their individual probability of being the longest is smaller than Path A. This shows that the quadratic regret solution tends to be less aggressive on achieving the best result, but tries to avoid the downside risk of being the worst decision.

The above example shows that the quadratic regret solution departs significantly

from the expected utility solution and permits interesting probabilistic interpretations. The quadratic regret model suits well in those decision making situations, where the objective is to achieve reasonable good outcome and avoid the severe downside risk as much as possible. The results from this preliminary study show that it is a promising direction to pursue further analysis on the quadratic regret solution in both stochastic combinatorial and continuous optimization problems.

4.3 Improving Persistency Estimation

One of the key messages in this thesis is problem transformation. Rather than solving the two classes of problems directly, we transformed them uniformly to the persistency estimation problems by appealing to Stein's Identity. Then the solutions to the persistency estimation problems can be used to build the solutions to the original problems. Therefore, the quality of the solutions to the original problems now hinges on the quality of the persistency estimates.

There are various ways to estimate persistency values. When we have enough historical data, for example, in the case of the portfolio management problem, we can directly estimate the persistency values from the data. When we only have distributional information on the uncertainties, we need to either perform a Monte Carlo simulation or resort to some persistency estimation models. In Section 1.1, we reviewed several generic SDP based models for persistency estimation. However, all those models are distributionally robust models, which assume that the exact distributional information of the uncertainty is unavailable and adopts the worst case analysis. This is not true in our situation, where we know that the uncertainty follows a multivariate normal distribution. There is certainly some accuracy lost when we use such distributionally robust models to estimate the persistency under a normal distribution. Such accuracy

lost will result in quality issues of the solution obtained for the original problems. As observed in our numerical studies, the larger the problem size, the worse the solution quality. Hence, there is a need to refine the persistency estimation models to capture more distributional information. In this section, we will discuss a way to refine one of the latest persistency estimation models in this category, CPCMM, developed by Natarajan et al. (2011). The reason why we choose to focus on CPCMM is its ability to capture correlations while maintaining a compact formulation. We have demonstrated the importance to consider correlations in project management problems in Chapter 2. Let us first recall the result of CPCMM.

4.3.1 CPCMM Revisit

As reviewed in Section 1.1, CPCMM's main result is using a deterministic conic optimization model to find the tight upper bound on the expected optimal value of a mixed zero-one linear programming problem when the cost coefficient vector is stochastic and described by the first two moments and nonnegative support, i.e.,

$$
\begin{aligned}
Z_P := \quad & \sup_{\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] \\
= \quad Z_C := \quad & \max \quad \sum_{j=1}^{n} Y_{j,j} \\
& s.t. \quad \boldsymbol{a}_i^T X \boldsymbol{a}_i - 2 b_i \boldsymbol{a}_i^T \boldsymbol{x} + b_i^2 = 0, \, \forall i = 1, \ldots, m \\
& \qquad X_{j,j} = x_j, \, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\} \\
& \qquad \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{x}^T \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T & Y^T \\ \boldsymbol{x} & Y & X \end{pmatrix} \succeq_{cp} 0
\end{aligned}
$$

where

$$\begin{aligned} Z(\tilde{\boldsymbol{c}}) \quad := \quad \max \quad & \tilde{\boldsymbol{c}}^T \boldsymbol{x} \\ s.t. \quad & \boldsymbol{a}_i^T \boldsymbol{x} = b_i, \, \forall i = 1, \ldots, m \\ & x_j \in \{0,1\}, \, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

The optimal variables of the decision variables, $\boldsymbol{x}$, Y and X, will provide estimates of the stochastic parameters, $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$ and $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, respectively. These estimates are exact under some worst case distribution.

4.3.2 Relationship to Scenario Planning

In CPCMM, it is assumed that the moments and covariance parameters of the random variables are known, and the only support information that is captured is nonnegativity. Then CPCMM constructs a set of scenarios, associated probability functions, and solutions that attain the worst case performance objective under this set of scenarios. The set of scenarios is completely determined by the optimization process. However, in practice there are usually some typical scenarios that we know will happen with relatively accurate probabilities. Next, we demonstrate how to extend CPCMM to capture such specific scenarios when describing the uncertainty set for the random variables. More specifically, suppose that there are N scenarios $\boldsymbol{c}^s$ with probability p_s, $s = 1, \ldots, N$, such that $\sum_{s=1}^{N} p_s = p \leq 1$. Furthermore, the conditional first and second moments for the remaining scenarios are denoted by $(\boldsymbol{\mu}, \Sigma)^+$. Then our problem reduces to

$$Z_P^S := (1-p) \sup_{\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] + \sum_{s=1}^{N} p_s Z_s(\boldsymbol{c}^s),$$

where $Z(\tilde{\boldsymbol{c}})$ is defined as before, and

$$\begin{aligned} Z_s\left(\boldsymbol{c}^s\right) := \quad \max \quad & \left(\boldsymbol{c}^s\right)^T \boldsymbol{x}^s \\ s.t. \quad & \boldsymbol{a}_i^T \boldsymbol{x}^s = b_i, \, \forall i = 1, \ldots, m \\ & x_j^s \in \{0,1\}, \, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\} \\ & \boldsymbol{x}^s \geq \mathbf{0} \end{aligned}$$

In this way, we can use a small set of scenarios with high probabilities to ensure that the optimal solution constructed will not perform too badly for these typical scenarios, and hence will not be overly conservative. Note that together with the original result of CPCMM, we can easily reformulate Z_P^S into a conic optimization problem. In particular, applying CPCMM to the remaining scenarios, we can obtain

$$\begin{aligned} Z_P^S \quad = \quad Z_C^S := \quad \max \quad & (1-p)\sum_{j=1}^{n} Y_{j,j} + \textstyle\sum_{s=1}^{N} p_s \left(\boldsymbol{c}^s\right)^T \boldsymbol{x}^s \\ s.t. \quad & \boldsymbol{a}_i^T X \boldsymbol{a}_i - 2 b_i \boldsymbol{a}_i^T \boldsymbol{x} + b_i^2 = 0, \, \forall i = 1, \ldots, m \\ & X_{j,j} = x_j, \, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\} \\ & \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{x}^T \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T & Y^T \\ \boldsymbol{x} & Y & X \end{pmatrix} \succeq_{cp} 0 \\ & \boldsymbol{a}_i^T \boldsymbol{x}^s = b_i, \, \forall i = 1, \ldots, m, \, \forall s = 1, \ldots, N \\ & x_j^s \in \{0,1\}, \, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\}, \, \forall s = 1, \ldots, N \\ & \boldsymbol{x}^s \geq \mathbf{0}, \, \forall s = 1, \ldots, N \end{aligned}$$

When $p = 1$ and $p_s = 1/N$ for all s, Z_P^S reduces to the conventional stochastic optimization problem solved via the sample average approximation method. Hence, this framework can be viewed as a bridge between the traditional stochastic optimization and modern robust optimization.

4.3.3 Capturing Normal Uncertainty

Now we will discuss how to utilize the above result to better describe the uncertainty following a multivariate normal distribution. The idea is based on discretizing the distribution and capturing different components of the sample points using different approaches. We summarize the main ideas in the following steps:

1. Discretize the random variable by generating a set of samples from the multivariate normal distribution;

2. Determine a region around the mean vector with high density and partition the region into N small grids;

3. For each grid s, $s = 1, \ldots, N$, estimate its probability (p_s) and conditional mean ($\boldsymbol{c}^s$), and treat (p_s, $\boldsymbol{c}^s$) as a specific scenario;

4. Remove the sample points inside the region from the set of samples;

5. Compute the probability ($1 - \sum_{s=1}^{N} p_s$) and conditional moments ($\boldsymbol{\mu}$, Σ) for the rest sample points;

6. Use the results from Sections 4.3.2 to reformulate the following problem into a conic optimization problem:

$$\left(1 - \sum_{s=1}^{N} p_s\right) \sup_{\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right] + \sum_{s=1}^{N} p_s Z_s\left(\boldsymbol{c}^s\right);$$

7. Solve the conic optimization problem and compute the persistency estimates from its optimal solution.

It is obvious that two extreme cases of the above approach are sample average approximation method and CPCMM. There are several advantages of this intermediate

method. Firstly, it captures much richer distributional information than the original CPCMM so that the persistency estimates will be more accurate. Secondly, the formulation can be maintained in a moderate size compared to the traditional sample average approximation method. The method focuses on the most probable scenarios around the mean for the multivariate normal distribution, and aggregates the less probable events for the worst case analysis. In other words, it transforms the difficulty from the large stochastic programming formulation into the conic constraint. Observe that the optimal solution to $Z(\tilde{\boldsymbol{c}})$ will not change if there is only a little perturbation in $\tilde{\boldsymbol{c}}$. Therefore, if the grid size is chosen properly, the optimal values of $Z_s(\boldsymbol{c}^s)$ from those specific scenarios are just the conditional expectations of $Z(\tilde{\boldsymbol{c}})$. Last but not least, the computational effort will not increase too much compared to the original CPCMM if N is not too large, as the conic constraint is the bottleneck when solving the problem. Since improving the persistency estimation is not the focus of this thesis, we leave these numerical studies and other issues to future research.

Bibliography

Adcock, C. J. (2007) *Extensions of Stein's Lemma for the skew-normal distribution*, Communications in Statistics–Theory and Methods, **36**, pp. 1661–1671.

Agrawal, S., Y. Ding, A. Saberi, Y. Ye (2012) *Price of correlations in stochastic optimization*, Operations Research, **60**, pp. 150–162.

Agrawal, S., D. Blaauw, V. Zolotov (2003) *Statistical timing analysis for intra-die process variations with spatial correlations*, Proceedings of the 2003 International Conference on Computer Aided Design, pp. 900–907.

Aissi, H., C. Bazgan, D. Vanderpooten (2009) *Min-max and min-max regret versions of combinatorial optimization problems: A survey*, European Journal of Operational Research, **197**, pp. 427–438.

Aldous, D., M. Steele (2003) *The objective method: Probabilistic combinatorial optimization and local weak convergence*, in Probability on Discrete Structures, H. Kesten (ed), Springer, Berlin, **110**, pp. 1–72.

Alexander, G. J., A. M. Baptista (2008) *Active portfolio management with benchmarking: Adding a value-at-risk constraint*, Journal of Economic Dynamics and Control, **32**, pp. 779–820.

Andersen, E. P. (1953) *On the fluctuations of sums of random variables*, Mathematica Scandinavica, **1**, pp. 263–285.

Azzalini, A. (2005) *The skew-normal distribution and related multivariate families*, Scadinavian Journal of Statistics, **32**, pp. 159–188.

Banerjee, A., Paul, A. (2008) *On path correlation and PERT bias*, European Journal of Operational Research, **189**, pp. 1208–1216.

Barbour, A. D., L. H. Y. Chen, W. L. Loh (1992) *Compound Poisson approximation for nonnegative random variables via Stein's method*, The Annals of Probability, **20**, pp. 1843–1866.

Bereanu, B. (1963) *On stochastic linear programming. I: Distribution problems: A single random variable*. Romanian Journal of Pure and Applied Mathematics, **8**, pp. 683–697.

Berman, A., N. Shaked-Monderer (2003) *Completely Positive Matrices*, World Scientific, Singapore.

Bertsimas, D., X. V. Doan, K. Natarajan, C. P. Teo (2010) *Models for minimax stochastic linear optimization problems with risk aversion*, Mathematics of Operations Research, **35**, pp. 580–602.

Bertsimas, D., K. Natarajan, C. P. Teo (2004) *Probabilistic combinatorial optimization: moments, semidefinite programming and asymptotic bounds*, SIAM Journal of Optimization, **15**, pp. 185–209.

Bertsimas, D., K. Natarajan, C. P. Teo (2006) *Persistence in discrete optimization under data uncertainty*, Mathematical Programming, **108**, pp. 251–274.

Best, M. J., R. R. Grauer (1991) *On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results*, The Review of Financial Studies, **4**, pp. 315–342.

Blaauw, D., K. Chopra, A. Srivastava, L. Scheffer (2008) *Statistical timing analysis: From basic principles to state of the art*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, **27**, pp. 589–607.

Bomze, I. M., M. Dür, E. D. Klerk, C. Roos, A. J. Quist, T. Terlaky, (2000) *On copositive programming and standard quadratic optimization problems,* Journal of Global Optimization, **18**, pp. 301–320.

Bowman, R. A. (1995) *Efficient estimation of arc criticalities in stochastic activity networks,* Management Science, **41**, pp. 58–67.

Brodie, J., I. Daubechies, C. D. Mol, D. Giannone, I. Loris (2009) *Sparse and stable Markowitz portfolios*, Proceedings of the National Academy of Sciences, **106**, pp. 12267–12272.

Boyd, S., L. Vandenberghe (2004) *Convex Optimizatioin*, Cambridge University Press.

Borkar, V. (1995) *Probability Theory: An Advanced Course*, S. Axler, F. W. Gehring, P. R. Halmos (eds), Springer, New York.

Brown, G. G., R. F. Dell, R. K. Wood (1997) *Optimization and persistence, I*nterfaces, **27**, pp. 15–37.

Bullock, E. R., M. E. Bitterman (1961) *Probability-matching in the fish*, The American Journal of Psychology, **74**, pp. 542–551.

Bullock, E. R., M. E. Bitterman (1962) *Probability-matching in the pigeon*, The American Journal of Psychology, **75**, pp. 634–639.

Burer, S. (2009) *On the copositive representation of binary and continuous nonconvex quadratic programs,* Mathematical Programming, **120**, pp. 479–495.

Cacoullos, T. (1982) *On upper and lower bounds for the variance of a function of a random variable*, The Annals of Probability, **10**, pp. 799–809.

Chang, H., S. S. Sapatnekar (2005) *Statistical timing analysis under spatial correlations*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, **24**, pp. 1467–1482.

Chen H. H., H. T. Tsai, D. Lin (2011) *Optimal mean-variance portfolio selection using Cauchy-Schwarz maximization*, Applied Economics, **43**, pp. 2795–2801.

Clark, E. C. (1961) *The greatest of a finite set of random variables*, Operations Research, **9**, pp. 145–162.

Conniffe, D., J. E. Spencer (2000) *Approximating the distribution of the maximum partial sum of normal deviates*, Journal of Statistical Planning and Inference, **88**, pp. 19–27.

Cornell, B., R. Roll (2005) *A delegated-agent asset-pricing model*, Financial Analysts Journal, **61**, pp. 57–69.

Cover, T. M. (1991) *Universal portfolios*, Mathematical Finance, **1**, pp. 1–29.

Cox, M. A. (1995) *Simple normal approximation to the completion time distribution for a PERT network,* International Journal of Project Management, **13**, pp. 265–270.

DeMiguel, V., L. Garlappi, F. J. Nogales, R. Uppal (2009) *A Generalized approach to portfolio optimization: Improving performance by constraining portfolio norms*, Management Science, **55**, pp. 798–812.

DeMiguel, V., L. Garlappi, R. Uppal (2007) *Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?*, The Review of Financial Studies, **22**, pp. 1915–1953.

Dodin, B. M. (1984) *Determining the K most critical paths in PERT networks*, Operations Research, **32**, pp. 859–877.

Dodin, B. M. (1985) *Bounding the project completion time distribution in PERT networks*, Operations Research, **33**, pp. 862–881.

Dodin, B. M., S. E. Elmaghraby (1985) *Approximating the criticality indices of the activities in Pert networks*, Management Science, **31**, pp. 207–223.

Dür, M. (2009) *Copositive programming: A survey*, In: M. Diehl, F. Glineur, E. Jarlebring, W. Michiels (Eds.), Recent Advances in Optimization and its Applications in Engineering, Springer, pp. 3–20.

El-Hassan, N., P. Kofman (2003) *Tracking error and active portfolio management*, Australian Journal of Management, **28**, pp. 183–207.

Elmaghraby, S. E. (2000) *On criticality and sensitivity in project networks*, European Journal of Operational Research, **127**, pp. 220–238.

Ewbank, J. B., B. L. Foote, H. J. Kumin (1974), *A method for the solution of the distribution problem of stochastic linear programming*, SIAM Journal on Applied Mathematics, **26**, pp. 225–238.

Fiorina, M. P. (1971) *A note on probability matching and rational choice*, behavioural Science, **16**, pp. 158–166.

Friedman, D., D. W. Massaro, S. N. Kitzis, M. M. Cohen (1995) *A comparison of learning models*, Journal of Mathematical Psychology, **39**, pp. 164–178.

Fulkerson, D. R. (1962) *Expected critical path lengths in PERT networks*, Operations Research, **10**, pp. 808–817.

Guttel, E., A. Harel (2005) *Matching probabilities: The behavioural law and economics of repeated behaviour*, The University of Chicago Law Review, **72**, pp. 1197–1236.

Helmbold, D. P., R. E. Schapire, Y. Singer, M. K. Warmuth (1998) *On-line portfolio selection using multiplicative updates*, Mathematical Finance, **8**, pp. 325–347.

Herbranson, T., J. Schroeder (2010) *Are birds smarter than mathematicians? Pigeons (Columba livia) perform optimally on a version of the Monty Hall Dilemma*, Journal of Comparative Psychology, **124**, pp. 1–13.

Hagstrom, J. N. (1988) *Computational complexity of PERT problems*, Networks, **18**, pp. 139–147.

Hertog, D. den, E. de Klerk, J. Roos (2002) *On convex quadratic approximation*, Statistica Neerlandica, **56**, pp. 376–385.

Hickson, R. H. (1961) Response probability in a two-choice learning situation with varying probability of reinforcement, Journal of Experimental Psychology, **62**, pp. 138–141.

Hurst, H. E. (1951) *Long term storage capacity of reservoirs*, Transactions of the American Society of Civil Engineers, **56**, pp. 376–385.

Jagannathan, R., T. Ma (2003) *Risk reduction in large portfolios: Why imposing the wrong constraints helps?*, Journal of Finance, **58**, pp. 1651–1684.

James, B., K. L. James, D. Siegmund (1987) *Test for a change-point*, Biometrika, **74**, pp. 71–83.

Jorion, P. (2003) *Portfolio optimization with constraints on tracking error*, Financial Analysts Journal, **59**, pp. 70–82.

Isserlis, L. (1918) *On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables*, Biometrika, **12**, pp. 134–139.

Kamburowski, J. (1985) *A note on the stochastic shortest route problem*, Operations Research, **33**, pp. 696–698.

Khandewal, V., A. Srivastava (2007) *A quadratic modeling-based framework for accurate statistical timing analysis considering correlations*, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, **15**, pp. 206–215.

Kirby, C., B. Ostdiek (2012) *It's all in the timing: Simple active portfolio strategies that outperform naive diversification*, Journal of Financial and Quantitative Analysis, **47**, pp. 437–467.

Kirk, K. L., M. E. Bitterman (1965) *Probability-learning by the turtle*, Science, New Series, **148**, pp. 1484–1485.

Klerk, E. de, D. V. Pasechnik (2002) *Approximation of the stability number of a graph via copositive programming*, SIAM Journal on Optimization, **12**, pp. 875–892.

Kleindorfer, G. B. (1971) *Bounding distributions for a stochastic acyclic network*, Operations Research, **19**, pp. 1586–1601.

Koehler, D. J., G. James (2009) *Probability matching in choice under uncertainty: Intuition versus deliberation*, Cognition, **113**, pp. 123–127.

Kong, Q., C. Y. Lee, C. P. Teo, Z. Zheng (2013) *Scheduling arrivals to a stochastic service delivery system using copositive cones*, to appear in Operations Research.

Lasserre, J. B. (2010) *A "joint+marginal" approach to parametric polynomial optimization*, SIAM Journal of Optimization, **20**, pp. 1995–2022.

Le, J., X. Li, L. T. Pileggi (2004) *STAC: Statistical timing analysis with correlation*, Proceedings of the 2004 Design Automation Conference, pp. 83–88.

Ledoit, O., M. Wolf (2003) *Improved estimation of the covariance matrix of stock returns with an application to portfolio selection*, Journal of Empirical Finance, **10**, pp. 603–621.

Ledoit, O., M. Wolf (2004) *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis, **88**, pp. 365–411.

Ledoit, O., M.Wolf (2008) *Robust performance hypothesis testing with the Sharpe ratio*, Journal of Empirical Finance, **15**, pp. 850–859.

Li, H., C. K. Koh, V. Balakrishnan, Y. Chen (2007) *Statistical timing analysis considering spatial correlations*, Proceedings of the 2007 International Symposium on Quality Electronic Design, pp. 102–107.

Liu, J. S. (1994) *Siegel's formula via Stein's identities*, Statistics & Probability Letters, **21**, pp. 247–251.

Longo, N. (1964) *Probability-learning and habit-reversal in the cockroach*, The American Journal of Psychology, **77**, pp. 29–41.

Lindsey, J. H. (1972) *An estimate of expected critical-path length in PERT networks*, Operations Research, **20**, pp. 800–812.

Löfberg, J. (2004) *YALMIP: A Toolbox for Modeling and Optimization in MATLAB*, In Proceedings of the CACSD Conference, Taipei, Taiwan, available online at: http://control.ee.ethz.ch/ ∼joloef/yalmip.php.

MacCrimmon, K. R., C. A. Ryavec (1964) *An analytical study of the PERT assumptions*, Operations Research, **12**, pp. 16–37.

Markowitz, H. M. (1952) *Portfolio selection*, Journal of Finance, **7**, pp. 77–91.

Michaud, R. O. (1989) *The Markowitz optimization enigma: Is 'optimized' optimal?*, Financial Analysts Journal, **45**, pp. 31–42.

Mishra, V. K., K. Natarajan, H. Tao, C. P. Teo (2012a) *Choice Prediction with Semidefinite Optimization when utilities are correlated*, IEEE Automatic Control, **57**. pp. 2450–2463.

Mishra, V. K., K. Natarajan, D. Padmanabhan, C. P. Teo (2013) *On theoretical and empirical aspects of marginal distribution choice models*, working paper.

Moritz, B. B., A. V. Hill, K. L. Donohue (2013) *Individual differences in the newsvendor problem: Behavior and cognitive reflection*, Journal of Operations Management, **31**, pp. 72–85.

Natarajan, K., M. Sim, J. Uichanco (2010) *Tractable robust expected utility and risk models for portfolio optimization*, Mathematical Finance, **20**, pp. 695–731.

Natarajan, K., M. Song, C. P. Teo (2009) *Persistency model and its applications in choice modeling*, Management Science, **55**, pp. 453–469.

Natarajan, K., C. P. Teo, Z. Zheng (2011) *Mixed zero-one linear programs under objective uncertainty: a completely positive representation*, Operations Research, **59**, pp. 713–728.

Ord, J. K. (1991) *A simple approximation to the completion time distribution for a PERT network*, The Journal of the Operational Research Society, **42**, pp. 1011–1017.

Rockafellar, R. T., S. Uryasev (2000) *Optimization of conditional value-at-risk*, Journal of Risk, **2**, pp. 21–42.

Rockafellar, R. T., S. Uryasev (2004) *Conditional value-at-risk for general loss distributions*, Journal of Banking and Finance, **26**, pp. 1443–1471.

Roll, R. (1992) *A mean/variance analysis of tracking error*, Journal of Portfolio Management, **18**, pp. 13–22.

Rudolf, M., H.-J. Wolter, H. Zimmermann (1999) *A linear model for tracking error minimization*, Journal of Banking & Finance, **23**, pp. 85–103.

Rustem, B., M. Howe (2002) *Algorithms for Worst-Case Design and Applications to Risk Management*, Princeton University Press, pp. 261–271.

Papadatos, N., V. Papathanasiou (2003) *Multivariate covariance identities with an application to order statistics*, Sankhya: The Indian Journal of Statistics, **65**, pp. 307–316.

Parrilo, P. A. (2000) *Structured semidefinite programs and semi-algebraic geometry*

methods in robustness and optimization, P.D. Dissertation, California Institute of Technology.

Prekopa, A. (1966) *On the probability distribution of the optimum of a random linear program*, SIAM Journal on Control and Optimization, **4**, pp. 211–222.

Schweitzer, M. E., G. P. Cachon (2000) *Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence*, Management Science, **46**, pp. 404–420.

Shanks, D. R., R J. Tunney, J. D. McCarthy (2002) *A re-examination of probability matching and rational choice*, Journal of behavioural Decision Making, **15**, pp. 233–250.

Shapiro, A. (1985) *Extremal problems on the set of nonnegative definite matrices*, Linear Algebra and its Applications, **67**, pp. 7–18.

Simon, H. A. (1956) *A comparison of game theory and learning theory*, Psychometrika, **21**, pp. 267–272.

Sculli, D. (1983) *The completion time of PERT networks*, The Journal of the Operational Research Society, **34**, pp. 155–158.

Siegel, A. F. (1993) *A surprising covariance involving the minimum of multivariate normal variables*, Journal of the American Statistical Association, **88**, pp. 77–80.

Spitzer, F. (1956) *A combinatorial lemma and Its application to probability theory*, Transactions of the American Mathematical Society, **82**, pp. 323–339.

Stein, C. M. (1972) *A bound for the error in the normal approximation to the distri-*

bution of a sum of dependent random variables, Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, **2**, pp. 583–602.

Stein, C. M. (1981) *Estimation of the mean of a multivariate normal distribution*, The Annals of Statistics, **9**, pp. 1135–1151.

Steinbach, M. C. (2001) *Markowitz revisited: Mean-variance models in financial portfolio analysis*, SIAM Review, **43**, pp. 31–85.

Stoltz, G., G. Lugosi (2005) *Internal regret in on-line portfolio selection*, Machine Learning, **59**, pp. 125–159.

Tang, Q., A. Zjajo, N. van der Meijs (2012) *Transistor-level gate model based statistical timing analysis considering correlations*, Proceedings of the 2012 Design, Automation & Test in Europe Conference & Exhibition, pp. 917–922.

Toh, K. C., M. J. Todd, R. H. Tutuncu (1999) *SDPT3 — a Matlab software package for semidefinite programming*, Optimization Methods and Software, **11**, pp. 545–581.

Tsukiyama, S., M. Tanaka, M. Fukui (2001) *A statistical static timing analysis considering correlations between delays*, Proceedings of the 2001 Asia and South Pacific Design Automation Conference, pp. 353–358.

Tu, J., G. Zhou (2011) *Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies*, Financial Analysts Journal, **55**, pp. 63–72.

Tutuncu, R. H., K. C. Toh, M. J. Todd (2003) *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming, **95**, pp. 189–217.

Vandenberghe, L., S. Boyd, K. Comanor (1972) *Generalized Chebyshev bounds via semidefinite programming*, SIAM Review, **49**, pp. 52–64.

von Briesen Raz, J. (1983) *Probability matching behaviour, association, and rational choice*, behavioural Science, **28**, pp. 35–52.

Vulkan (2000) *An economist's perspective on probability matching*, Journal of Economic Surveys, **14**, 101–118.

Wang, M. Y. (1999) *Multiple-Benchmark and Multiple-Portfolio Optimization*, Journal of Financial Economics, **99**, pp. 204–215.

Wilson, W. A., M. Oscar Jr., M. E. Bitterman (1964) *Probability-learning in the monkey*, Quarterly Journal of Experimental Psychology, 16, pp. 163–165.

Yao, M. J., W. M. Chu (2007) *A new approximation algorithm for obtaining the probability distribution function for project completion time*, Computers & Mathematics with Applications, **54**, pp. 282–295.

Zhan, Y., A. J. Strojwas, X. Li, L. T. Pileggi, D. Newmark, m. Sharma (2005) *Correlation-aware statistical timing analysis with non-Gaussian delay distributions*, Proceedings of the 2005 Design Automation Conference, pp. 77–82.

Zhang, L., W. Chen, Y. Hu, J. A. Gubner, C. C. P. Chen (2005) *Correlation-preserved non-Gaussian statistical timing analysis with quadratic timing model*, Proceedings of the 2005 Design Automation Conference, pp. 83–88.