Nondominated Sets and Surfaces in Multiple Criteria Optimization and Portfolio Selection Theory in Finance

by

Yue Qi

(Under the direction of Ralph E. Steuer)

Abstract

Because of the degree to which the effectiveness of an economy rests upon its ability to most carefully allocate its capital, and because of the degree to which the investment strategy of a portfolio manager or individual investor can effect the success of his or her organization or personal welfare, no stone should be left unturned in trying to better understand the investment process. In this dissertation, the view is taken that one of the reasons for complexity in the investment process is that criteria beyond variance and expected return often come into play, such as dividends, social responsibility, the number of securities in a portfolio, and so forth. One of the consequences of admitting criteria beyond the two is that the efficient (nondominated) frontier becomes a nondominated surface, thus rendering much of traditional investment analysis a projection onto two dimensions of the often much more complicated problem of portfolio selection in higher dimensional space. In this dissertation, the types of nondominated surfaces that can result in multiple criteria optimization, particularly in multiple criteria portfolio optimization, are studied along with the development of methods for computing. Also, procedures are explored for locating the point on a given nondominated surface that represents the investor's multiple criteria optimal portfolio.

Index words: Multiple criteria decision making, Multiple criteria optimization, Portfolio selection, Efficient frontier, Nondominated surface

Nondominated Sets and Surfaces in Multiple Criteria Optimization and Portfolio Selection Theory in Finance

by

Yue Qi

B.S., Nankai University China, 1992

M.Eco., Nankai University China, 1996

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2004

Nondominated Sets and Surfaces in Multiple Criteria Optimization and Portfolio Selection Theory in Finance

by

Yue Qi

Approved:

Major Professor: Ralph E. Steuer

Committee: K. Roscoe Davis
Chris T. Stivers
Elliot C. Gootman
Qing Zhang

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2004

Acknowledgments

I would like to thank the members of my dissertation committee, Dr. Ralph E. Steuer, Dr. K. Roscoe Davis, Dr. Chris T. Stivers, Dr.Elliot C. Gootman, and Dr. Qing Zhang for their support. I am especially grateful to my advisor Dr. Steuer who originally proposed this research. I would like to express my gratitude for his enthusiasm, direction, and encouragement. Finally I deeply appreciate all the patience and help from my wife throughout these years.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In standard portfolio theory, it is presumed that investors' major goal is to minimize variance and maximize expected return. In this dissertation as an extension, the existence of another type of investors, *suitable-portfolio investors* is proposed. The goal of suitable-portfolio investors is to build a *suitable portfolio*. This means that in addition to standard deviation and expected return, such investors may wish to monitor their portfolios with regard to dividend, amount invested in R & D, the maximum amount invested in any security, tracking error, number of securities in a portfolio, short selling and so forth. To accommodate such investors, a multiple criteria portfolio selection model is developed. From this, the extension of an efficient (nondominated) *frontier* to a nondominated *surface* is observed, and the idea arises that what is the "modern portfolio theory" of today might be interpreted as a projection onto two-space of the multiple criteria portfolio selection problem from higher dimensional space. In this dissertation, the term nondominated frontier is used instead of efficient frontier in order to reserve the term efficiency for points in decision space according to the custom in multiple criteria optimization. From this point on, the term "suitable-portfolio investor" will often be applied to an investor who has criterion concerns (such as described above) beyond standard deviation and expected return.

1.1 Standard vs. Suitable-portfolio investors

The degree to which standard portfolio theory has been developed is impressive. But knowing that all kinds of investors exist in the real world, one difficulty with existing theory is that it only assumes *standard* investors, investors whose goals revolve around minimizing variance and maximizing expected return. This leads to the question whether these two goals are sufficient representatives of investors' criteria as also considered by Markowitz [65]

> "This single period utility function may depend on portfolio return and perhaps other state variables ... the modern portfolio theorist is able to trace out mean-variance frontiers for large universes of securities. But is this the right thing to do for the investor? In particular, are mean and variance proper and sufficient criteria for portfolio choice?" (1991, p. 471).

In this dissertation, for side-by-side existence with standard investors, another type of investor, called a *suitable-portfolio investor* is considered. Suitable-portfolio investors strive for the construction of a "suitable portfolio" by including more criteria beyond standard deviation and expected return. For instance, consider a common case of a parent who wishes to create a portfolio for his or her child to interest the child in investing. Would the parent purchase stock in tobacco, alcohol, or assault weapons manufacturers? Probably not. Most likely the parent would seek stocks like Merck, Johnson & Johnson, and Disney to set up a sense of social responsibility and interest by keeping the child from losing focus. In addition, the parent might wish to overweight on stocks that pay dividends so as to generate a warm reminder of the project for the child to enjoy every three months. The parent may also try to lower the number of securities in a portfolio to minimize the time, headache and distraction involved in monitoring and managing a portfolio, and restrict short selling to avoid being over-nervous about margin calls. Virtually a lot, if not most, families will testify that these considerations are not irrational.

Consequently, this means that many investors may well wish to seek balance in their portfolios among various competing concerns such as dividends, liquidity, the number of securities involved, social responsibility, and so forth. Suitable-portfolio investors can have the following advantages.

1. Not only seeing the (standard deviation, expected return) space, they also add dimensions to it with one extra dimension for each criterion. Their portfolios is in a high

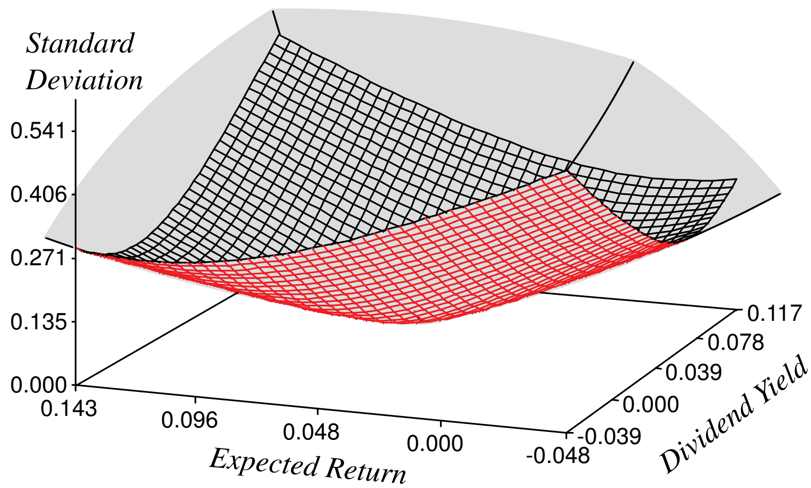

Figure 1.1: Suitable-portfolio investors' portfolios in (standard deviation, expected return, and dividend yield) space

dimensional space and brings flexibility. For example, for a given expected return, suitable-portfolio investors still have freedom in choosing from other criteria, while a standard investor' portfolio is fixed after determining expected return. In order to visualize, suppose suitable-portfolio investors add only dividend yield objective. Then their portfolios constitute a set in (standard deviation, expected return, and dividend yield) space in Figure 1.1. The mesh surface is the minimum-variance surface as an extension of standard investors' minimum-variance frontier. The derivation of this surface of unlimited short-sales allowed model (with the only constraint $\mathbf{1}^T\mathbf{x} = 1$) is in Chapter 7. Suitable-portfolio investors' portfolios are the shaded area above and including the surface.

2. Suitable-portfolio investors see a nondominated surface in high dimensional space as an extension of nondominated frontier in (standard deviation, expected return) space. Suppose suitable-portfolio investors add only dividend yield objective. Their nondominated surface in (standard deviation, expected return, and dividend yield) space is represented by the mesh surface in Figure 1.2 *top*. One connection between suitable-portfolio investors and standard investors is to project suitable-portfolio investors' nondominated surface onto (standard deviation, expected return) space. Such projection is performed based on the nondominated surface in Figure 1.2 *top*. The projection of the nondominated surface is in 1.2 *bottom*. One boundary of this projection is marked by the thick line in graph. This boundary can be exactly the nondominated frontier of standard investors of some models, for example unlimited short-sales allowed models in Chapter 7.

3. Standard investors usually obtain a portfolio by fixing expected return and minimizing variance. However, this approach loses its appeal for more than two objectives. Furthermore, it can locate dominated portfolios and induce imprecise definitions of nondominated frontier. Detailed discussions are in Section 6.4. Suitable-portfolio investors can avoid this difficulty by applying definitions and methods of multiple criteria optimization.

Suitable-portfolio investors will typically encounter the following difficulties.

1. There are little of tools to study portfolio selection in this multiple criteria way. Suitable-portfolio investors pursue a two-stage procedure to select a portfolio. They compute a nondominated surface in the first stage and locate an optimal portfolio from this surface in the second stage. Research on the first stage is being conducted by related scholars mentioned in Section 1.3 and Steuer, Hirschberger, and Qi. More research need to be done as discussed in Chapter 9 to provide suitable-portfolio investors with more tools.

2. In practice it is difficult to work with nondominated surfaces. In this dissertation, as an extension of Merton [70], the nondominated surface and other results in closed-form formulae have been developed of unlimited short-sales allowed model (with the only constraint $\mathbf{1}^T\mathbf{x} = 1$) with three objectives (variance, expected return, and one other linear objective). Also, Hirschberger, Qi and Steuer [39] have proposed an

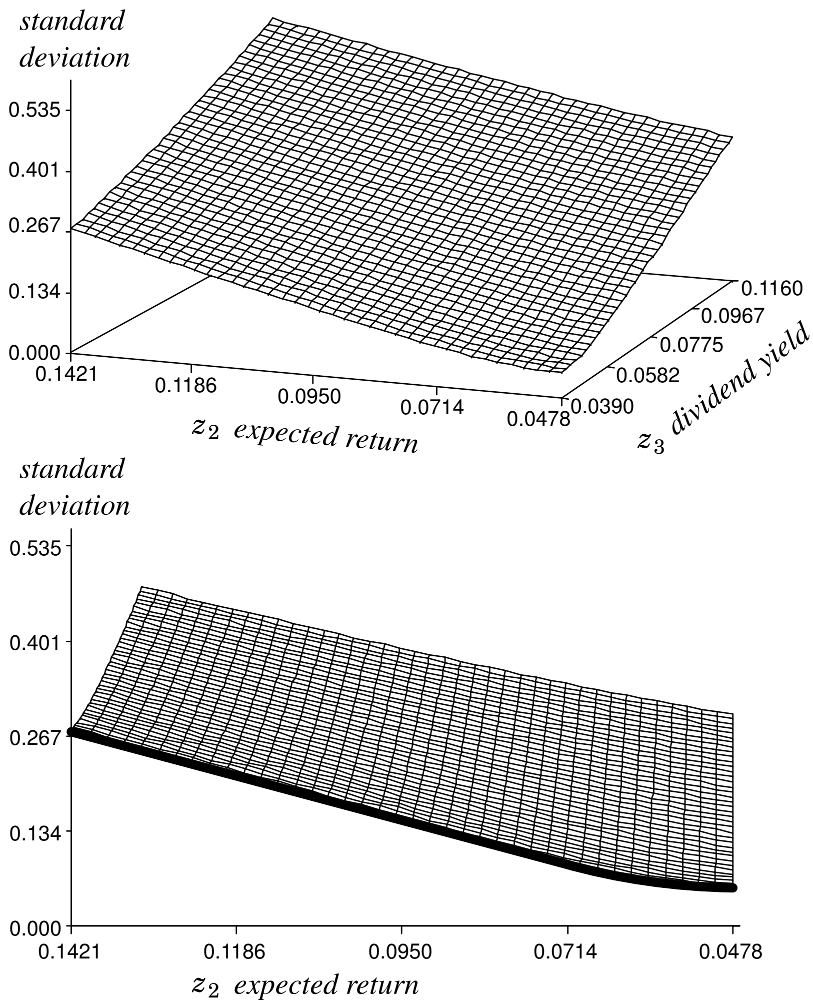

Figure 1.2: Suitable-portfolio investors' nondominated surface and the projection onto (standard deviation, expected return) space

algorithm to compute the nondominated surface of linearly constrained models with three objectives (variance, expected return, and one other linear objective). However, beyond this there is only uncharted territory where suitable-portfolio investors in an effort to be more realistic, can easily find themselves confronted with nonlinear constraints, integer variables, and other nonsmooth conditions. In such situations, suitable-portfolio investors may have to rely on multiple criteria evolutionary algorithms (see for example Deb [19]) and and then in the end be able to only compute a discretized representation of the nondominated set.

3. In contrast to the nondominated frontier of a standard investor, the nondominated set of a suitable-portfolio investor can be quite large and difficult to deal with, particularly in the case when it is described by large number of discrete points. With finding the best point on a nondominated surface being a nontrivial task, interactive procedures appear to be the only way to proceed in most situations.

1.2 Buried Market Portfolio

In addition to seeming reasonable, might there be any real-world evidence consistent with suitable-portfolio investors' multiple criteria approach? Consider the market portfolio, the portfolio containing all securities in proportion to their market capitalization (see for example Bodie, Kane and Marcus [10], pp. 264-265). By assumptions of the Capital Asset Pricing Model, the market portfolio is on the nondominated frontier and is investors' optimal portfolio of risky assets. However with many thousands securities existing in the capital markets, observing the true market portfolio is difficult. Therefore, well-known indices such as the Standard & Poor's 500 Index are often used as surrogates. Empirically for 1997-2002, the Standard & Poor's 500 Index was relatively deep below the nondominated frontier constructed from only its constituents. Moreover, indices of other major industrialized countries, Nikkei 225 Index, FTSE 100 Index, CAC 40 Index (France), DAX 30 Index (Germany) followed approximately the same pattern for this time period. Descriptions and graphs are in

Appendix. This universally buried market portfolio surrogate phenomenon may be difficult to explain by standard finance theory (However, it is recognized that this empirical study is far from being exhaustive, but the trend of results is interesting nevertheless).

Suitable-portfolio investors' approach can be used to explain this phenomenon. A simple experiment is conducted as follows. Assume that the criterion vectors (portfolios) of a standard investor form an elliptical area (the set bounded by an ellipse) in Figure 1.3. Here, the nondominated frontier of the standard investor is the portion of the ellipse indicated by the thick line. Further assume that the criterion vectors (portfolios) of a suitable-portfolio investor is an ellipsoidal volume (the set bounded by an ellipsoid) in k-dimensional space, with the same major and minor axes in (standard deviation, expected return) space, where k is the number of the suitable-portfolio investor's criteria. Then, the nondominated surface of the suitable-portfolio investor is the portion of the surface of the ellipsoid that projects onto the upper left quarter of the standard investor's portfolios. This detail is proved in the Appendix. The projection is marked by the shaded area in Figure 1.3.

If the optimal portfolio of the standard investor is in the "middle" (according to some definition) of the nondominated frontier like $\mathbf{z}^2$, then it can be reasonable to find that the optimal portfolio of the suitable-portfolio investor in the middle of a nondominated surface in k-space. In this sense, a portfolio in the middle of a nondominated surface in 3-space would project onto (standard deviation, expected return) space as $\mathbf{z}^3$, a portfolio in the middle of a nondominated surface in 4-space would project onto (standard deviation, expected return) space as $\mathbf{z}^4$, and so forth. As seen in Figure 1.3, the $\mathbf{z}^3$, $\mathbf{z}^4$ and $\mathbf{z}^5$ become buried more deeply below the nondominated frontier as the number of criteria increases. Therefore, if an index or market portfolio is a suitable-portfolio investor's optimal portfolio, being found below the nondominated frontier is not a contradiction to theory, but consistent with exactly what one would expect with suitable-portfolio theory.

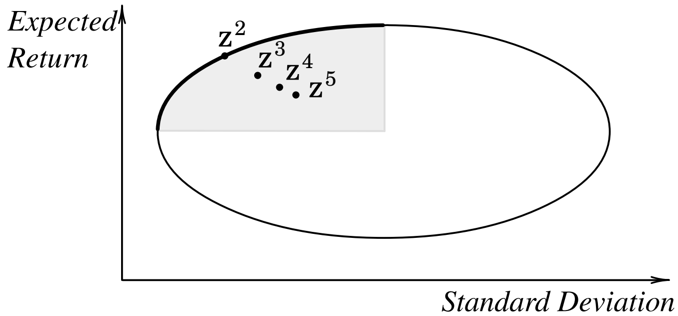

Figure 1.3: The projection of optimal portfolios become deeper below the nondominated frontier as the number of criteria increases

1.3 Recent Papers on Multiple Criteria Portfolio Selection

Existing portfolio theory is based upon approaches developed by Von Neumann and Morgenstern [94] and Markowitz [61] over 50 years ago. The world has certainly become more complex in the interim. New tools such as those from multiple criteria decision analysis have become available in the meantime, with evidence mounting, it now appears timely to see if existing portfolio theory can't be extended to better meet the modelling needs of investors other than of the standard variety.

To place previous multiple criteria oriented portfolio analysis research into perspective, six categories are employed: (1) overall framework, (2) the modelling of individual investor objectives, (3) portfolio ranking, (4) skewness inclusion, (5) alternative measures of risk, and (6) decision support systems.

In the first category, there are overview pieces such as by Spronk and Hallerbach [86] in 1997, Bana e Costa and Soares [6] in 2001, and Hallerbach and Spronk [34, 35] in 2002 in which the benefits of employing multiple criteria concepts in financial decision making are outlined. In the second category, Chang, Meade, Beasley and Sharaiha [14] in 2000 added integer variables to control the number of securities in a portfolio and utilized a genetic algorithm, Tabu search, and simulated annealing to solve the portfolio selection problems.

Bouri, Martel and Chabchoub [11] in 2002 deployed a five objective (expected return, β, liquidity, size and price-earning ratio) portfolio selection model. Ehrgott, Klamroth and Schwehm [24] also set up a five objective (12-month performance, 3-year performance, annual dividend, Standard and Poor's star ranking, and 12-month volatility) portfolio selection model and constructed value functions to solve the problem.

Employing tools from multiple criteria decision analysis (see for example Belton and Stewart [7]) for portfolio ranking, there are the papers by Hurson and Zopounidis [45] , Yu [96], Jog, and Kaliszewski and Michalowski [47], and Hurson and Ricci [44]. Noteworthy in the category of skewness inclusion, there is the early paper by Stone [92] and, more recently, the papers by Konno, Shirakawa and Yamazaki [51] and Konno and Suzuki [52]. With regard to alternative measures of risk, there are the efforts by Zeleny [97], Konno and Yamazaki [53], Feinstein and Thapa [29], Doumpos, Spanos and Zopounidis [21], and Michalowski and Ogryczak [71]. With regard to decision support systems employing mathematical programming techniques, there are the approaches of Colson and DeBruyn [17], Ballestero and Romero [5], Tamiz, Hasham, Jones, Hesni and Fargher [93], Dominiak [20], Ogryczak [73], Arenas Parra, Bilbao Terol and Rodríguez Uría [77], Ballestero and Pla-Santamaria [4], Mansini, Ogryczak and Speranza [60], and Zopounidis and Doumpos [99].

Multiple criteria in portfolio selection have sometimes been alluded by researchers in finance. Markowitz [65] in 1991 pointed out that investors can add practical measures as criteria and nondominated frontiers can be expanded into three dimensional or four dimensional surfaces by taking skewness or kurtosis into account. Sharpe [82] in his online lecture briefly introduces additional linear objectives into an expected utility function framework so that investors can determine parameters in the function, maximize the function, and obtain a maximizing portfolio. Chow [16] in 1995 set up a three objective (variance, expected return, and expected tracking error) portfolio selection model. Not only utilizing an expected utility function to get a maximizing portfolio similar to Sharpe's approach, he also got a set of parameters for the expected utility function, obtained a set of maximizing portfolios (one maximizing portfolio for each element in the set of parameters), and formed a discretized representation of the nondominated surface from these maximizing portfolios in (variance, expected return, and expected tracking error) space. Lo, Petrovz and Wierzbicki [58] in 2003 set up a three objective (variance, expected return, and liquidity) portfolio selection model. They used three methods ("liquidity filtering, liquidity constraints, and a mean-variance-liquidity function") to get a nondominated surface ("mean-variance-liquidity frontier") in (variance, expected return, and liquidity) space. Their second method is an e-constraint approach (e-constraint approach is described in Chapter 2). Their third method is an expected utility function approach, similar to Chow's method. They also observed that "portfolios close to each other on the traditional mean-variance efficient frontier can differ substantially in their liquidity characteristics". This can support the advantage of suitable-portfolio investors. That is, suitable-portfolio investors can have an portfolio with approximately the same variance and expected return as standard investors' portfolio and have the freedom to choose high liquidity.

It is mostly in the first two categories, overall framework and the modelling of individual investor objectives that this dissertation is concerned. The contributions of all the research are recognized. However, the research in the first two categories and in finance can be improved in the following aspects.

1. The researchers set up objectives from practical considerations. They may typically need theoretical arguments for the modelling, for example properties of an extended utility function and the relationship between this function and the nondominated surface.

2. These research is relatively specialized and the solving methods are relatively heuristic. For example, it can be subjective to determine a set of parameters for the expected utility function for one model, but this set of parameters may not work well for another model.

3. It can be very time-consuming to get a set of parameters for the expected utility function, obtain a set of maximizing portfolios, and form a discretized representation of the nondominated surface from these maximizing portfolios. Moreover, this method loses its appeal if a suitable-portfolio investor has more than three objectives.
4. The distribution of the maximizing portfolios can be uneven, although the distribution of the set of parameters is even. Therefore, the discretized representation of the nondominated surface can be inappropriate. That is, some parts of the nondominated surface have a lot of maximizing portfolios, while some other parts have few maximizing portfolios.

Based on the research above, this dissertation makes the following contributions.

1. proposing the existence of general multiple criteria in portfolio selection.
2. arguing for this proposal by using an stochastic programming approach and an extended utility function.
3. modelling these multiple criteria.
4. deriving the closed-form formulae for unlimited short-sales allowed (with the only constraint $\mathbf{1}^T\mathbf{x} = 1$) general k objective (variance, expected return, and other linear objectives) portfolio selection models.
5. proposing an algorithm to compute the nondominated surface of linearly constrained models with three objectives (variance, expected return, and one other linear objective). This algorithm can get the exact nondominated surface and is more effective than the repetitively maximizing expected utility function approach (for example Chow [16]).

1.4 Outline of Dissertation

The following chapters of this dissertation are arranged as follows. Chapter 2 introduces multiple criteria optimization, gives readers a feel for how nondominated surfaces can grow with the problem size, and discusses the consequence that can result from ignoring criteria. Chapter 3 reviews standard portfolio theory through the view of stochastic programming. Chapter 4 outlines Markowitz's critical line algorithm and Hirschberger's procedure for computing nondominated frontiers of general mean-variance portfolio selection problems. Whereas initial comparisons show that Markowitz's critical line algorithm becomes faster than Hirschberger's procedure beyond about 200 securities, one advantage of Hirschberger's procedure is that it can be extended to additional linear criteria.

Chapter 5 outlines a procedure developed in Hirschberger, Qi and Steuer [41] to generate covariance matrices and expected returns. The procedure allows the entries in the matrices to have distributional characteristics closely matching those of empirical data. This is achieved by relatively easily setting some parameters. This procedure can also provide approximately unlimited reasonable data for benchmarking portfolio selection optimization algorithms and comparing these algorithms, while historical data can be severely limiting in this region.

Chapter 6 provides the theoretical arguments for multiple criteria portfolio selection by introducing multiple criteria stochastic programming and an extended utility function. It also points out the difficulty of standard portfolio selection theory in treating multiple criteria as constraints.

In Chapter 7, a three objective (variance, expected return, and one other linear objective) portfolio selection model is analyzed, as an extension of Merton [70]'s work to derive the nondominated frontier of a standard unlimited short-sales allowed model. With the only constraint $\mathbf{1}^T\mathbf{x} = 1$, approximately all results are in closed-form formulae to bring convenience in teaching and research and offer insight. The standard minimum-variance frontier, a parabola, is extended into a minimum-variance surface, a paraboloid. The portfolio weights (inverse image) of a standard investor's minimum-variance frontier is a subset of a suitable-portfolio investor's minimum-variance surface. By the same token, the efficient set of a standard investor is a subset of the efficient set of a suitable-portfolio investor. Properties of the minimum-variance surface and nondominated surface are delineated. A general k objective (variance, expected return, and other linear objectives) portfolio selection model is analyzed. The minimum-variance surface is a paraboloid in k-dimensional space. Furthermore, the previously nondominated portfolios will still stay nondominated as suitable-portfolio investors gradually add objectives.

Chapter 8 outlines the author's current research to approximate the nondominated surface of a two quadratic and two linear objective model. Chapter 9 concludes the dissertation and future research is proposed.

In this dissertation, matrices are typically notated by bold upper-case letters, for example $\mathbf{A}$ and $\mathbf{\Sigma}$. Vectors are typically notated by bold lower-case letters, for example $\mathbf{x}$ and $\boldsymbol{\mu}$. The elements of matrices and vectors are typically notated by italic lower-case letters with subscripts, for example $a_{11}, a_{12}, \ldots, a_{1n}$ and $\sigma_{11}, \sigma_{12}, \ldots, \sigma_{1n}$ and $x_1, x_2, \ldots, x_n$ and $\mu_1, \mu_2, \ldots, \mu_n$.

Chapter 2

Multiple Criteria Optimization

Because the requirements of portfolio selection with multiple criteria exceed those of standard mathematical programming, multiple criteria optimization (see Steuer [89], Miettinen [72] and Ehrgott [22]), a specialized area of mathematical programming devoted to problems with three or more objective functions, is now overviewed.

A multiple criteria optimization problem is written as follows

$$\begin{aligned} &\max\{f_1(\mathbf{x}) = z_1\} \\ &\max\{f_2(\mathbf{x}) = z_2\} \\ &\qquad\vdots \\ &\max\{f_k(\mathbf{x}) = z_k\} \\ &\text{s.t.} \quad \mathbf{x} \in S \end{aligned} \tag{2.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is a *decision vector* in *decision space*, S is the *feasible region* in *decision space*, and k is the number of objectives. The $f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})$ are *objective functions*, $\mathbf{z} \in \mathbb{R}^k$ is a *criterion vector* with *criterion values* $z_1, \ldots, z_k$, and $Z = \{(z_1, \ldots, z_k) \mid z_1 = f_1(\mathbf{x}), \ldots, z_k = f_k(\mathbf{x}), \mathbf{x} \in S\}$ is the feasible region in *criterion space*. If minimization on objective i is encountered, it can be remodelled as

$$\min\{f_i(\mathbf{x}) = z_i\} = -\max\{-f_i(\mathbf{x}) = -z_i\} \tag{2.2}$$

Definition 2.1. *Assume* (2.1). *A criterion vector* $\bar{\mathbf{z}} \in Z$ dominates $\mathbf{z} \in Z$ *if only if* $\bar{z}_i \geq z_i$ *for all* $i \in K$ *and* $\bar{z}_i > z_i$ *for at least one* $i \in K$*, where* $K = \{1, 2, \ldots, k\}$.

Definition 2.2. *Assume* (2.1)*. A criterion vector* $\bar{\mathbf{z}} \in Z$ *is* nondominated *if and only if there does not exist* $\mathbf{z} \in Z$ *such that* $\mathbf{z}$ *dominates* $\bar{\mathbf{z}}$*. Otherwise,* $\bar{\mathbf{z}}$ *is* dominated.

Definition 2.3. *A point* $\bar{\mathbf{x}} \in S$ *is* efficient *if and only if its criterion vector* $\bar{\mathbf{z}} = (f_1(\bar{\mathbf{x}}), \ldots, f_k(\bar{\mathbf{x}}))$ *is nondominated. Otherwise,* $\bar{\mathbf{x}}$ *is* inefficient.

The set of nondominated criterion vectors is designated N, and called the *nondominated set* (or *nondominated frontier* or *nondominated surface* in the financial parts in this dissertation). The set of all efficient points is designated E and called the *efficient set*. It is the endeavor of multiple criteria optimization to compute both E and N for display to the decision maker. Solving (2.1) usually means computing E and N. Sets E and N are crucial because if the decision maker's value function (defined later in Section 2.1) is coordinatewise increasing ("the more, the better" in each objective, all others held constant), all optimal vectors in criterion space are in N, and all optimal solutions in decision space are in E.

2.1 Three General Approaches

Early approaches for computing E, N and solving (2.1) include the e-constraint approach, weighted-sums approach, and value function approach.

In the e-constraint approach, only one objective function (here assume $f_1(\mathbf{x})$) is retained, while all the others are transformed into $\geq$ or $=$ constraints with the corresponding right-hand-side (parameters) e_j as follows

$$\begin{aligned} \max\, & \{f_1(\mathbf{x}) = z_1\} \\ \text{s.t.}\quad & f_2(\mathbf{x}) = e_2 \\ & f_3(\mathbf{x}) = e_3 \\ & \vdots \\ & f_k(\mathbf{x}) = e_k \\ & \mathbf{x} \in S \end{aligned} \tag{2.3}$$

An advantage of this approach is that it converts a multiple objective program to an ordinary single criterion one. The e-constraint approach can be effective for computing the nondominated frontier when $k = 2$, i.e., when there are only two original objectives. This can be accomplished by incrementing the right-hand-side, e_j of the objective that has been converted to an e-constraint, and then solving (2.3) each time. While effective when $k = 2$, this method loses its appeal when $k \geq 3$ because there is no systematic way to vary the right-hand-sides in such situations.

In the weighted-sums approach, a positive weighting vector $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ where

$$\boldsymbol{\Lambda} = \{\, \boldsymbol{\lambda} \in \mathbb{R}^k \mid \lambda_1 > 0, \lambda_2 > 0, \ldots, \lambda_k > 0, \sum_{i=1}^{k} \lambda_i = 1\} \tag{2.4}$$

is used to form the weighted-sums (composite) function as follows

$$\begin{aligned} \max \{\lambda_1 f_1(\mathbf{x}) + \ldots + \lambda_k f_k(\mathbf{x}) = z_w\} \\ \text{s.t.} \qquad\qquad \mathbf{x} \in S \end{aligned} \tag{2.5}$$

where $\boldsymbol{\Lambda}$ is *weighting space* of (2.1). The requirement $\sum_{i=1}^{k} \lambda_i = 1$ can be relaxed, because a positive weighting vector $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ can be normalized as $\frac{1}{\lambda_1+\lambda_2+\ldots,+\lambda_k}(\lambda_1, \lambda_2, \ldots, \lambda_k)$. To discover E and N, the idea is to vary the λ_i. While this approach can be made to be effective when $k = 2$, when $k \geq 3$ the method loses its appeal for roughly the same reason as the e-constraint approach. As noted by several authors (e.g., Geoffrion [31] and Ehrgott [22]) the method has some theoretical appeal. For example,

Theorem 2.1. *For any given* $\bar{\boldsymbol{\lambda}} \in \Lambda$, *if* $\bar{\mathbf{x}} \in S$ *maximizes* (2.5), *then* $\bar{\mathbf{x}}$ *is efficient.*

Theorem 2.2. *In* (2.1), *assume all* $f_i(\mathbf{x})$ *are concave objective functions and* S *is a convex set. Then* $\bar{\mathbf{x}} \in S$ *is properly efficient if and only if there exists an* $\bar{\boldsymbol{\lambda}} \in \Lambda$ *such that* $\bar{\mathbf{x}}$ *is the optimal solution of* (2.5) *for this* $\bar{\boldsymbol{\lambda}}$.

An efficient solution $\bar{\mathbf{x}} \in E$ is properly efficient if there exists $M > 0$ such that, for each i, the relationship $\frac{f_i(\mathbf{x})-f_i(\bar{\mathbf{x}})}{f_j(\bar{\mathbf{x}})-f_j(\mathbf{x})} < M$ holds for some j such that $f_j(\mathbf{x}) < f_j(\bar{\mathbf{x}})$ whenever $\mathbf{x} \in S$

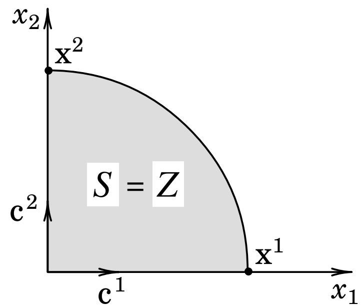

Figure 2.1: Example of improperly efficient points

and $f_i(\mathbf{x}) > f_i(\bar{\mathbf{x}})$. Otherwise, the $\bar{\mathbf{x}} \in E$ is improperly efficient (See Steuer [89], p. 438). The interpretation can be that the gain-loss ratio at a properly efficient point is bounded. Properly efficient is stronger than efficient, that is, properly efficient implies efficient, but not vice versa. Properly efficient points only appear in solving nonlinear multiple criteria optimization problems. For example, by taking $\mathbf{c}^1 = (1,0)^T$ and $\mathbf{c}^2 = (0,1)^T$, feasible regions S and Z are identical and represented by the shaded region in Figure 2.1. The efficient set is the boundary (curve) from $\mathbf{x}^1$ to $\mathbf{x}^2$ including $\mathbf{x}^1$ and $\mathbf{x}^2$. But $\mathbf{x}^1$ and $\mathbf{x}^2$ are the only efficient points that are improperly efficient.

In the value function approach, a value function $V\colon \mathbb{R}^k \to \mathbb{R}$ is used to assign a value z_v along the real line to each criterion vector using

$$\begin{aligned} &\max \{V(z_1, z_2, \ldots, z_k) = z_v\} \\ &\text{s.t.} \qquad\qquad \mathbf{z} \in Z \end{aligned} \tag{2.6}$$

Basically designed for small scale discrete problems, this approach is rarely used in discrete problems when S consists of more than about 20 points, or in problems in which S is continuous. Nevertheless, two interesting theoretical insights can be drawn (see Steuer [89]).

Theorem 2.3. *Let V be coordinatewise increasing. Then, if $\mathbf{z}^{\circ} \in Z$ is optimal, $\mathbf{z}^{\circ}$ is non-dominated.*

Theorem 2.4. *Let $\bar{\mathbf{z}} \in Z$ be nondominated. Then, there exists a coordinatewise increasing value function V such that $\bar{\mathbf{z}}$ is optimal.*

2.2 Multiple Objective Linear Programming

If all objective functions and constraints of a multiple criteria optimization problem are linear, then it is a *multiple objective linear programming* (MOLP) problem that can be written as.

$$\begin{aligned} \max \{ \mathbf{c}^{1^T}\mathbf{x} &= z_1 \} \\ \max \{ \mathbf{c}^{2^T}\mathbf{x} &= z_2 \} \\ &\vdots \\ \max \{ \mathbf{c}^{k^T}\mathbf{x} &= z_k \} \\ \text{s.t.} \quad \mathbf{A}\mathbf{x} &\leq \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \tag{2.7}$$

where:

$\mathbf{c}^i$	is the *gradient* ($n \times 1$ vector of objective function coefficients) of the ith objective
$\mathbf{A}$	is the $m \times n$ *constraint matrix*
$\mathbf{b}$	is the $m \times 1$ *right-hand-side vector*

Note that with $S = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, S is convex and polyhedral (created from linear constraints) and thus possessing extreme points and facets up to possibly dimensionality of n. Extreme points and facets are defined as follows (see Steuer [89] and Sayin [80]).

Definition 2.4. *Let set $S \subset \mathbb{R}^n$ be convex. A point $\bar{\mathbf{x}} \in S$ is an* extreme point *(sometimes called a* vertex*) of S if and only if points $\mathbf{x}^1, \mathbf{x}^2 \in S$, with $\mathbf{x}^1 \neq \mathbf{x}^2$, do not exist such that $\bar{\mathbf{x}} = \lambda\mathbf{x}^1 + (1-\lambda)\mathbf{x}^2$ for some $\lambda \in (0, 1)$.*

That is an extreme point can not be expressed as a strictly convex combination of any two different points in the set in question.

Definition 2.5. *Let F be a subset of S. F is a facet of S if every line segment in S with a relative interior point in F has both end-points in F.*

Because in multiple criteria optimization, the computation is typically done in decision space (space of the $\mathbf{x}$ vectors) and the display of information is done in criterion space (space of the $\mathbf{z}$ vectors), it is necessary to understand these two spaces. To illustrate, consider the multiple objective linear program (MOLP)

$$
\begin{aligned}
\max \{ \ \ 4x_1 - \ \ x_2 &= z_1 \} \\
\max \{ -x_1 + 2x_2 &= z_2 \} \\
\text{s.t.} \qquad x_1 + 2x_2 &\leq 14 \\
-x_1 + \ \ x_2 &\leq 4 \\
x_1 \qquad &\leq 8 \\
x_1, x_2 &\geq 0
\end{aligned}
$$

portrayed in Figure 2.2 *top* in which S is the feasible region in decision space. Feasible region S has five extreme points, $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^5$. In this problem, the efficient set E is the portion of the boundary of S from $\mathbf{x}^1$ to $\mathbf{x}^2$ to $\mathbf{x}^3$ represented by the thick line in Figure 2.2 *top*. Note that in E there are three 0-dimensional facets or extreme points ($\mathbf{x}^1, \mathbf{x}^2$ and $\mathbf{x}^3$) and two 1-dimensional facets ($\mathbf{x}^1$ to $\mathbf{x}^2$, and $\mathbf{x}^2$ to $\mathbf{x}^3$). The criterion space representation of the MOLP is given in Figure 2.2 *bottom* in which Z is the feasible region in criterion space. Observe the distortion that takes place whenever the gradients of the objectives in decision space are not of equal length and do not lie along the axes. Here the nondominated set N (or as one might say in finance when there are only two criterion space dimensions, nondominated frontier) is the portion of the boundary of Z from $\mathbf{z}^1$ to $\mathbf{z}^2$ to $\mathbf{z}^3$ represented by the thick line in Figure 2.2 *bottom*. Note that in N are three 0-dimensional facets or extreme points ($\mathbf{z}^1, \mathbf{z}^2$ and $\mathbf{z}^3$) and two 1-dimensional facets ($\mathbf{z}^1$ to $\mathbf{z}^2$, and $\mathbf{z}^2$ to $\mathbf{z}^3$).

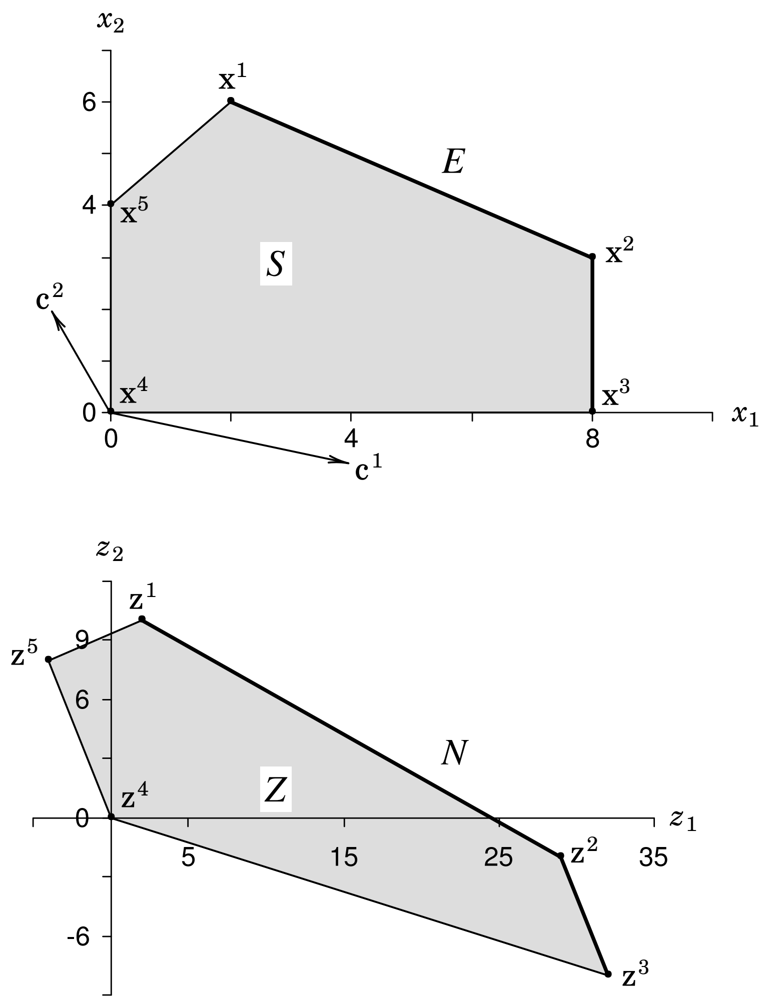

Figure 2.2: Feasible regions S and Z in decision space and criterion space

If some objective functions or constraints are nonlinear, then the S or Z can be nonlinear in decision space or criterion space, for example, as in portfolio selection, where a nonlinear objective function for variance is involved. While the S can still be polyhedral, the Z is not. Because of the nonlinearity, the facet structure is not suitable anymore. The descriptions of S, E, Z and N of a 3-asset portfolio selection problem are given in Section 3.2.

2.3 Size of Nondominated Set

The question is: what happens to the size of the efficient and nondominated sets as problem size increases? Typically, they increase in complexity as the number of variables and constraints increase, and grow rapidly in size as k, the number of objectives, increases.

To illustrate, ADBASE, a solver for computing all efficient extreme points of an MOLP by Steuer [88], is employed. Involved is the random problem generator of ADBASE to generate and solve MOLPs of various size. Sample results are in Table 2.1. Whereas the complexity of the efficient and nondominated sets grows with the number of constraints m, and the number of variables n, the size of the efficient set and nondominated set grows with the number of objectives k. While the effects are mild with the number of constraints and variables, they are more dramatic with the number of objectives.

Since, with m and n fixed, the size of the efficient and nondominated sets grows dramatically as k increases, one can see the possible loss in information, if objectives, for whatever reason, are omitted.

2.4 Consequences of Ignoring Criteria

To further illustrate the loss of information that can happen in a general multiple criteria optimization context, consider the following experiment.

1. A 7-objective, 16-constraint, 16-variable MOLP is randomly generated by ADBASE.

2. The feasible region S has $172,005$ extreme points. These points are projected into the criterion space of the first two objectives (with the first objective in minimization form

Table 2.1: Results for computing efficient extreme points of different sizes of MOLP

Objectives	Constraints	Efficient extreme points n, variables			CPU times (seconds) n, variables		
k	m	50	100	150	50	100	150
	20	33	33	42	.1	.2	.3
2	40	71	89	102	.2	.4	.7
	60	97	169	208	.3	.9	1.7
	20	302	588	669	1	3	7
3	40	1,682	2,641	4,149	3	13	41
	60	3,019	7,391	10,306	6	37	110
	20	1,635	3,137	4,814	3	19	59
4	40	9,469	21,232	41,877	20	138	592
	60	41,184	101,747	233,571	90	727	3,525

and the second in maximization form as in finance). They show as the mass of gray dots in Figure 2.3 (although their true color is red in color prints). Considering only the first two objectives, only 11 of the $172,005$ extreme points produce nondominated vertices in criterion space. These nondominated vertices show as the black dots in the upper left corner of Figure 2.3.

3. Suppose we continue to insist to look at this problem in the criterion space of the first two objectives. However, suppose that unknown to us, the problem actually contains also the third objective. Then the problem would in reality have the 101 black dots in Figure 2.4 as nondominated vertices. But viewing in the criterion space of only the first two objectives, we would inadvertently conclude that 90 of them were dominated (when in fact any of them in theory could be optimal when all three objectives are considered).

4. Suppose we continue to insist to look at this problem in the criterion space of the first two objectives. However, suppose that unknown to us, the problem actually contains

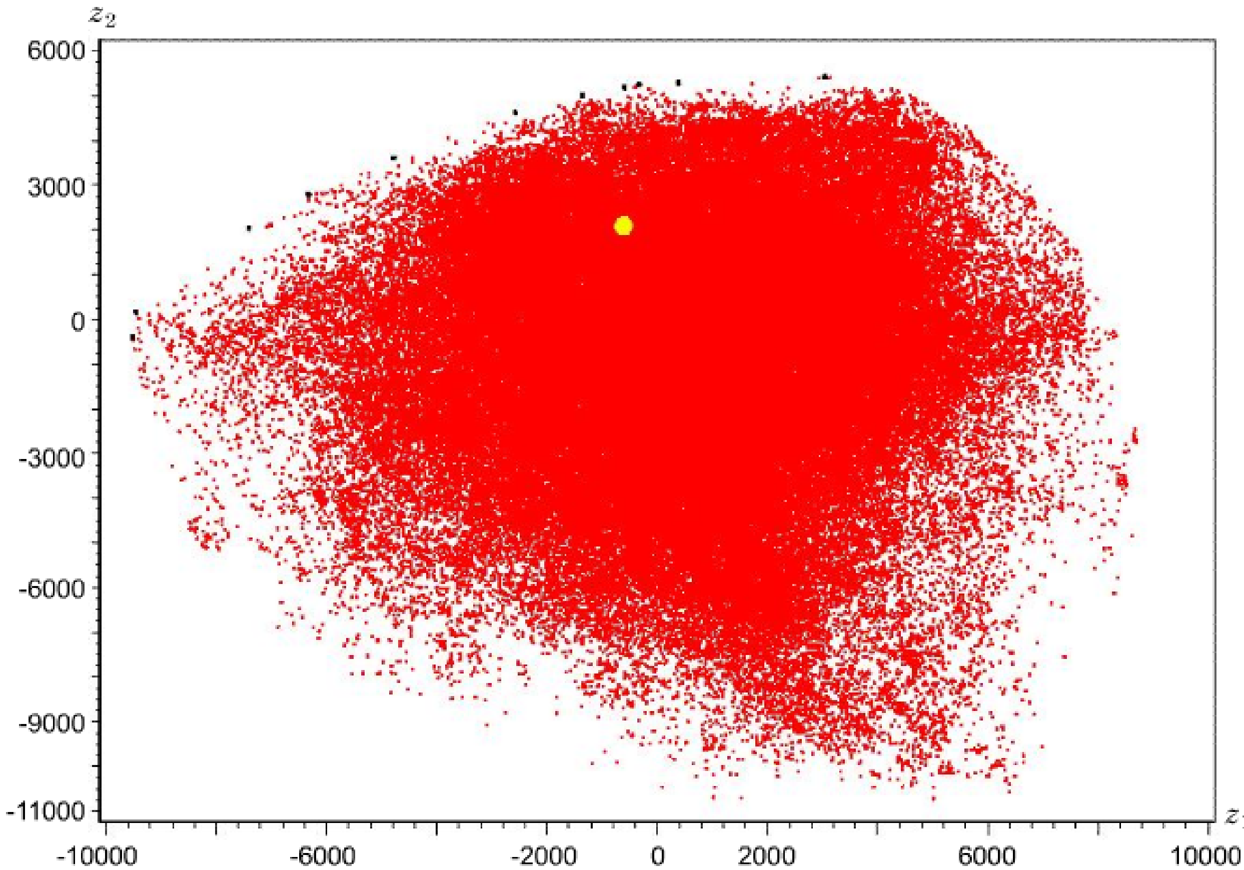

Figure 2.3: Shows the 11 nondominated vertices w.r.t. the first two objectives

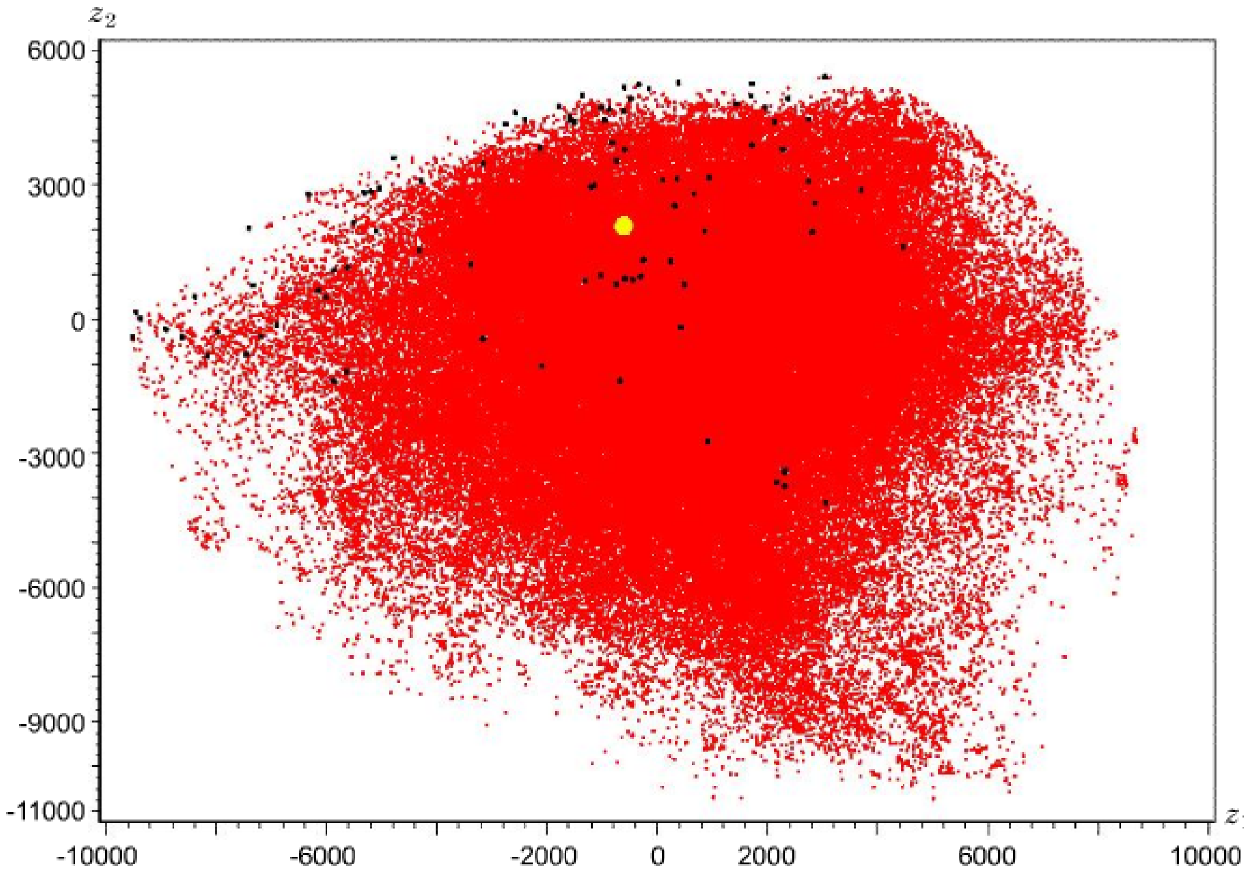

Figure 2.4: Shows the 101 nondominated vertices w.r.t. the first three objectives

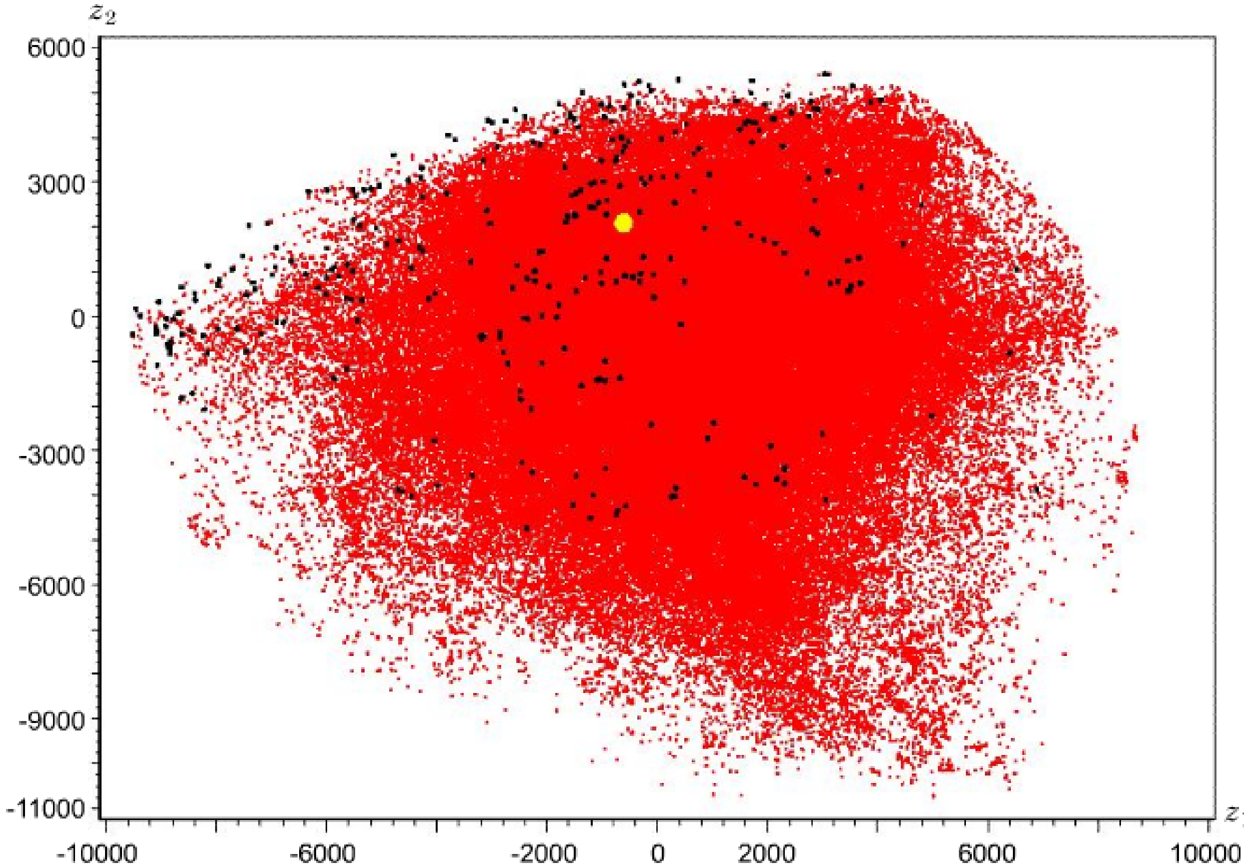

Figure 2.5: Shows the 312 nondominated vertices w.r.t. the first four objectives

also the third and fourth objectives. Then the problem would in reality have the 312 black dots in Figure 2.5 as nondominated vertices. But viewing in the criterion space of only the first two objectives, we would inadvertently conclude that 301 of them were dominated (when in fact any of them in theory could be optimal when all four objectives are considered).

5. Suppose we continue to insist to look at this problem in the criterion space of the first two objectives. However, suppose that unknown to us, the problem actually contains

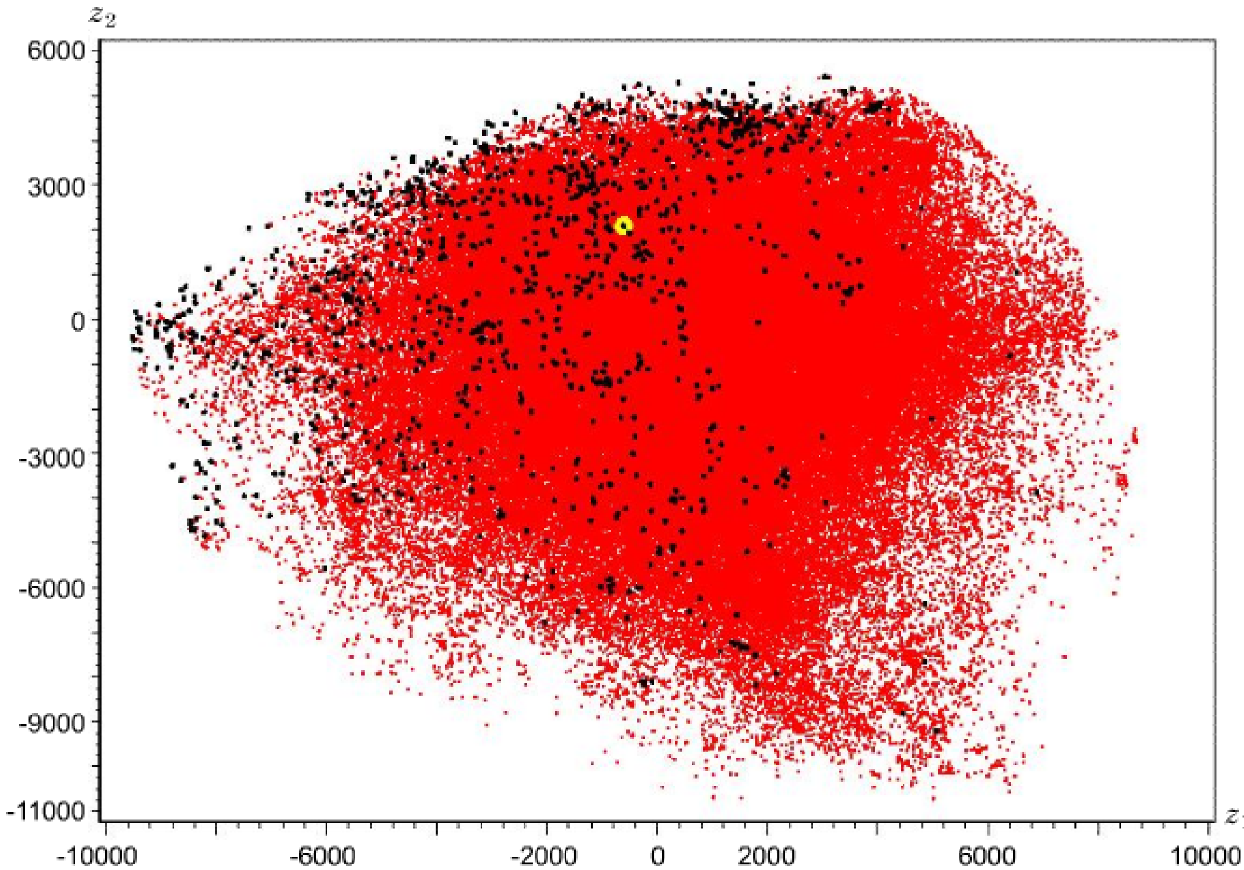

Figure 2.6: Shows the 1, 129 nondominated vertices w.r.t. the first five objectives

also the third, fourth, and fifth objectives. Then the problem would in reality have the 1, 129 black dots in Figure 2.6 as nondominated vertices. But viewing in the criterion space of only the first two objectives, we would inadvertently conclude that 1, 118 of them were dominated (when in fact any of them in theory could be optimal when all five objectives are considered).

6. Suppose we continue to insist to look at this problem in the criterion space of the first two objectives. However, suppose that unknown to us, the problem actually contains

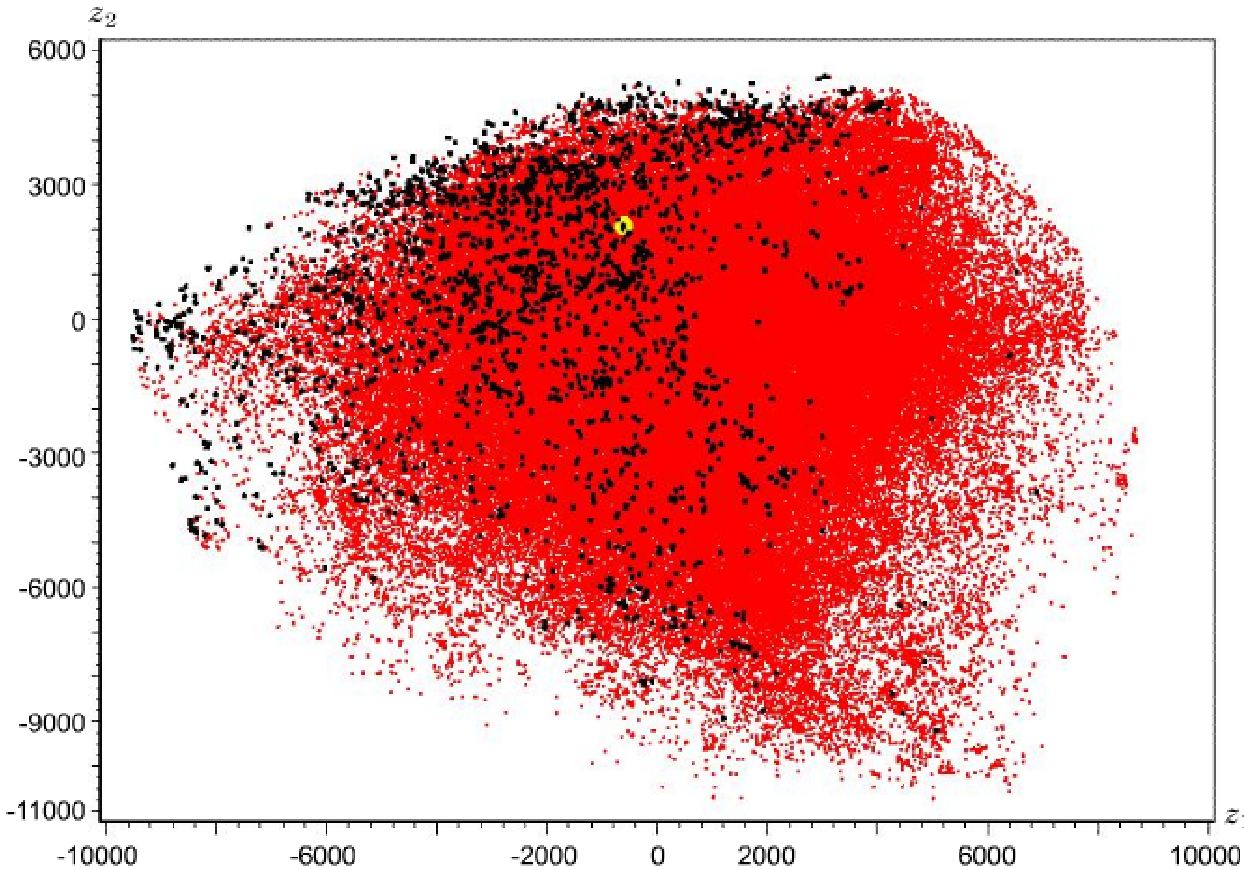

Figure 2.7: Shows the 2,053 nondominated vertices w.r.t. the first six objectives

also the third, fourth, fifth, and sixth objectives. Then the problem would in reality have the 2,053 black dots in Figure 2.7 as nondominated vertices. But viewing in the criterion space of only the first two objectives, we would inadvertently conclude that 2,042 of them were dominated (when in fact any of them in theory could be optimal when all six objectives are considered).

7. Suppose we continue to insist to look at this problem in the criterion space of the first two objectives. However, suppose that unknown to us, the problem actually contains

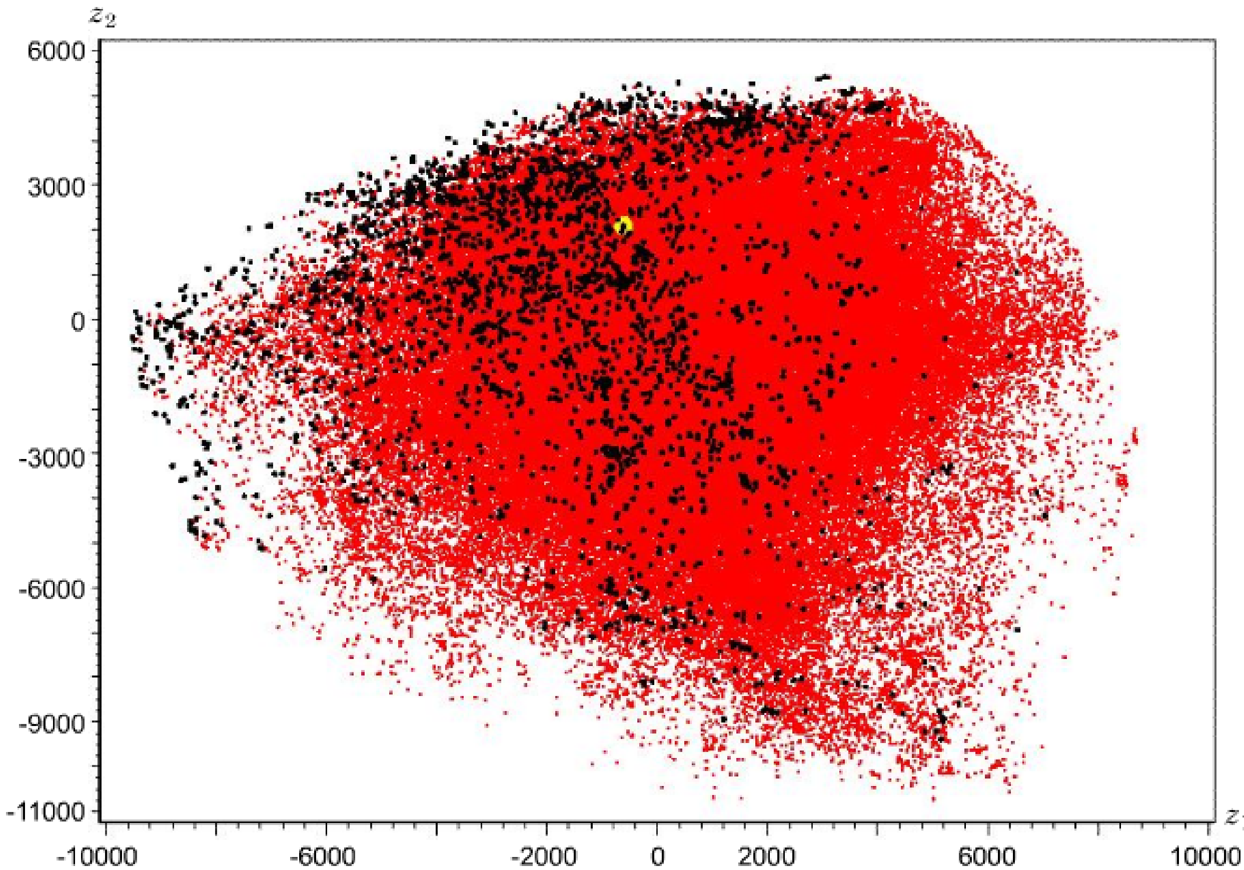

Figure 2.8: Shows the 2, 726 nondominated vertices w.r.t. the first seven objectives

also the third, fourth, fifth, sixth, and seventh objectives. Then the problem would in reality have the 2, 726 black dots in Figure 2.8 as nondominated vertices. But viewing in the criterion space of only the first two objectives, we would inadvertently conclude that 2, 715 of them were dominated (when in fact any of them in theory could be optimal when all seven objectives are considered).

8. About the white dot (although its true color is yellow in color prints) as the biggest dot in all the figures, assume that the real problem had all the seven objectives.

> Randomly select one of the $2,726$ nondominated vertices as the optimal solution to the 7-objective MOLP. The white dot is the optimal solution projected in the first two objectives. With high probability $\frac{2726-11}{2726}$, this dot will lie buried below the nondominated set of the first two objectives formed by the 11 nondominated vertices. Also note that it would not appear as a black dot until the first five objectives are included. This means that this true optimal solution will not appear as nondominated of the first two objective model, the first three objective model, or the first four objective model.

The projection and plotting of this experiment is performed by dots.sas programmed by the author and illustrated in Appendix. This experiment based on linear models can serve as an educated guess of the nondominated sets of suitable-portfolio investors and standard investors and the potential loss of information of standard investors.

With the amount of information potentially lost in general multiple criteria optimization when objectives are ignored, the concern of this dissertation is about the loss of information if indeed portfolio selection involves for some investors objectives beyond standard variance and expected return, and how to possibly avoid such situations.

Chapter 3

Standard Portfolio Selection Theory

This chapter reviews standard portfolio selection theory from the point of view of stochastic programming. Reasons for utilizing stochastic programming are as follows

1. Stochastic programming can show how a standard investor is different from a suitable-portfolio investor from the very outset.

2. Stochastic programming nicely shows the rationale for Markowitz's mean-variance approach.

3. Stochastic programming can describe an investor' objectives more practically than maximizing expected utility approach.

The problem of portfolio selection starts as follows. Assume

(a) n securities

(b) an initial sum of money to be invested

(c) the beginning of a holding period

(d) the end of the holding period.

Let $x_1, x_2, \ldots, x_n$ be the proportions of the initial sum to be invested in the n securities at the beginning of the holding period that define a portfolio to be held fixed until the end of the holding period. Let $r_1, r_2, \ldots, r_n$ denote the returns of the n securities to be realized on over the holding period. Note that $r_1, r_2, \ldots, r_n$ are random variables. A portfolio is defined by its vector of portfolio weight $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. A portfolio return is given by

$$R(\mathbf{x},\mathbf{r}) = \sum_{i=1}^{n} r_i x_i = \mathbf{r}^T \mathbf{x} \tag{3.1}$$

The reason to use symbol $R(\mathbf{x{,}r})$ is to denote that the end-of-holding-period portfolio return is dependent upon both $\mathbf{x}$ and $\mathbf{r}$ and is thus a random variable. Let the feasible region S in decision space be defined as follows

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{1}^T\mathbf{x} = 1,\ \mathbf{A}\mathbf{x} \leq \mathbf{b},\ \boldsymbol{\ell} \leq \mathbf{x} \leq \boldsymbol{\omega}\} \tag{3.2}$$

where $\boldsymbol{\ell}$ is a lower bound vector on $\mathbf{x}$, and $\boldsymbol{\omega}$ is an upper bound vector on $\mathbf{x}$. However, standard portfolio selection theory (see Huang and Litzenberger [43]) starts with maximizing expected utility approach

$$\begin{aligned} &\max \{E[U(R(\mathbf{x}, \mathbf{r}))] = z\} \\ &\quad \text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{3.3}$$

where $U(.)$ is a utility function, typically assumed to be increasing and concave. If $U(.)$ is quadratic or the asset universe's returns $r_1, r_2, \ldots, r_n$ are multivariate normally distributed, the mean-variance approach proposed by Markowitz [61] is consistent with (3.3). That is, an maximizing solution of (3.3) is an efficient solution of the mean-variance approach

$$\begin{aligned} &\min \{V[R(\mathbf{x}, \mathbf{r})] = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}\} \\ &\max \{E[R(\mathbf{x}, \mathbf{r})] = \boldsymbol{\mu}^T\mathbf{x}\} \\ &\quad \text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{3.4}$$

where $V[\,]$ is variance, $\boldsymbol{\Sigma}$ is a covariance matrix, and $\boldsymbol{\mu}$ is an expected return vector of $r_1, r_2, \ldots, r_n$. In many finance texts, the mean-variance approach is modelled as

$$\begin{aligned} &\min \{V[R(\mathbf{x}, \mathbf{r})] = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}\} \\ &\quad \text{s.t.} \quad \boldsymbol{\mu}^T\mathbf{x} = e \\ &\qquad\qquad \mathbf{x} \in S \end{aligned} \tag{3.5}$$

where e is some target expected portfolio return. Model (3.5) is an e-constraint approach of model (3.4). Model (3.5) has the advantage of being a single criterion optimization problem. However in this dissertation, model (3.4) is primarily used because it demonstrates that Markowitz's mean-variance approach is a two criterion model.

"Two objectives, however, are common to all investors for which the techniques of this monograph are designed:

1. They want "return" to be high ...

2. Then want this return to be dependable, stable, not subject to uncertainty.

" Markowitz [63] (1959, p. 6).

3.1 Portfolio Selection through Stochastic Programming

Stochastic programming is proposed to formulate uncertainty. Defined by Prékopa [79] as

> "... stochastic programming handles mathematical programming problems, where some of the parameters are random variables; either we study the statistical properties of the random optimum value or other random variables that come up with the problem, or we reformulate it into a decision type problem by taking into account the joint probability distribution of the random parameters" (1995, p. viii).

Stochastic programming problems can be written as

$$\max \{f(\mathbf{x}, \mathbf{u}) = z\} \quad (3.6)$$
$$\text{s.t. } \mathbf{x} \in S$$

where as usual S is the feasible region. The uncertainty is modelled by $\mathbf{u}$, and $\mathbf{u}$ is a random vector. Note that $f(\mathbf{x}, \mathbf{u})$ is still a random variable. In contrast to deterministic optimization, stochastic programming can contain random variables as parameters in objective functions to model uncertainty.

It must be pointed out that a stochastic programming problem is not well-defined. What it means to maximize a random variable is not clear. A substitute problem (deterministic model) is needed to operationalize a stochastic programming problem. Prékopa [79] and

Caballero, Cerdá, Munõz, Rey, and Stancu-Minasian [13] have proposed and compared several substitute problems. One of these substitute problems is

$$\min \{V[f(\mathbf{x}, \mathbf{u})] = z_1\} \tag{3.7}$$
$$\max \{E[f(\mathbf{x}, \mathbf{u})] = z_2\}$$
$$\text{s.t.} \qquad \mathbf{x} \in S$$

It is proposed that the overall objective of an investor at the very beginning is to maximize her or his portfolio return $R(\mathbf{x}, \mathbf{r})$.

> "A key measure of investors' success is the rate at which their funds have grown during the investment period" in Bodie and Kane and Marcus [9] (2004, p. 132).

> "Why invest? Stated in simplest terms, investors wish to earn a return on their money" in Jones [49] (2000, p. 9).

Note that at the very beginning an investor considers her or his "portfolio return" $R(\mathbf{x}, \mathbf{r})$ instead of the "expected portfolio return" $E[R(\mathbf{x}, \mathbf{r})]$, although this $E[R(\mathbf{x}, \mathbf{r})]$ only first appears in a substitute problem. Also note that the portfolio weight $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ must be determined at the beginning of the holding period, but the realized return is not known until the end of the holding period. In other words, the investors has no other choice but to make decisions under uncertainty.

> "Any investment involves some degree of uncertainty about future holding period returns, and in most cases that uncertainty is considerable" Bodie and Kane and Marcus [9] (2004, p. 135).

> "Uncertainty is a salient feature of security investment ... all can affect the capital gains or dividends of one or many securities" Markowitz [63] (1959, p. 4).

Thus, to reflect the fact that an investor's overall objective at the beginning is to maximize her or his portfolio return $R(\mathbf{x}, \mathbf{r})$, the following stochastic programming problem is proposed

$$\begin{aligned} &\max \{R(\mathbf{x}, \mathbf{r}) = z\} \\ &\text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{3.8}$$

One substitute problem for (3.8) as shown by Caballero, Cerdá, Munõz, Rey, and Stancu-Minasian [13] is the mean-variance approach (3.4) (as proposed and used by Markowitz). Prékopa [79] proposed maximizing expected utility approach (3.3) as a substitute problem of (3.8) as well.

Therefore with mean-variance approach and maximizing expected utility approach as substitute problems, (3.8) results in approximately the same formulation as standard portfolio selection theory. However, the reasons to utilize stochastic programming are as follows

1. Stochastic programming can show how a standard investor is different from a suitable-portfolio investor from the very outset. A standard investor, with portfolio return in mind follows (3.8), while a suitable-portfolio investor has additional objectives and follows a multiple criteria stochastic programming approach such as

$$\begin{aligned} &\max \{R(\mathbf{x}, \mathbf{r}) = \text{return}\} \\ &\max \{f_2(\mathbf{x}, \boldsymbol{\delta}) = \text{dividend yield}\} \\ &\max \{f_3(\mathbf{x}, \mathbf{s}) = \text{social responsibility}\} \\ &\max \{f_4(\mathbf{x}, \mathbf{h}) = \text{R\&D}\} \\ &\text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{3.9}$$

where the random vectors $\boldsymbol{\delta}, \mathbf{s}$, and $\mathbf{h}$ represent the uncertainty of dividend yield, social responsibility, and R&D, respectively. This multiple criteria stochastic model is discussed further in Chapter 6.

2. In standard theory, stochastic programming problem (3.8) ultimately results in a mean-variance approach as the most widely accepted substitute problem. In suitable-portfolio theory, following similar rationale, (3.9) results in a substitute problem such as the following

$$
\begin{aligned}
\min \{V[R(\mathbf{x},\mathbf{r})] &= \text{variance}\} \\
\max \{E[R(\mathbf{x},\mathbf{r})] &= \text{expected return}\} \\
\max \{f_2(\mathbf{x}) &= \text{dividend yield}\} \\
\max \{f_3(\mathbf{x}) &= \text{social responsibility}\} \\
\max \{f_4(\mathbf{x}) &= \text{R\&D}\} \\
\text{s.t.} \quad \mathbf{x} &\in S
\end{aligned}
$$

Note that a suitable-portfolio investor will need a simplifying assumption in order to drop $\boldsymbol{\delta}, \mathbf{s}$, and $\mathbf{h}$ out of (3.9) and derive the five objective model above, as also discussed in Chapter 6.

3. Stochastic programming, especially in the form of its mean-variance substitute problem, can describe investors' objective more pragmatically than maximizing expected utility approach. Maximizing expected utility approaches can be difficult to apply in practice as pointed out by

> "Maximizing *EU* is typically manyfold more difficult than performing mean-variance analysis. While large institutional investors frequently employ quantitative analysts, few have determined their actual utility function–by specifying their willingness to engage in various simple gambles, as described by von Neumann and Morgenstern (1944). Further, the entire joint distribution of returns must be estimated, which is generally a much larger requirement than estimating the means, variances, and covariances needed for a mean-variance analysis. Finally, the finding of the expected utility-maximizing portfolio typically requires many times more computational resources than does the critical line algorithm for tracing out a mean-variance efficient frontier" Markowitz [64] (1990, p. 100).

However, the contribution of maximizing expected utility approach is recognized and even Markowitz [63] claimed that he is a supporter of this approach.

4. Caballero, Cerdá, Munõz, Rey, and Stancu-Minasian [13] also proposed a safety-first approach and a value-at-risk approach as substitute problems for (3.6). These two approaches can be applied to (3.8) to formulate risk more naturally as argued by Ziemba [98]

> "Stochastic programming models handle extreme event scenarios in a natural way. There is little chance of anyone predicting such events as the 9/11 attacks, but scenarios that represent the effect of such events in terms of their impact on market returns can be included" (2003, p. 49).

Note that this dissertation will focus only on a mean-variance substitute problem at this stage.

3.2 Markowitz's Mean-variance Approach

In the e-constraint approach (3.5), a minimizing portfolio can be obtained for a given e. As e varies, a series of minimizing portfolios can be obtained. A frontier can be obtained by plotting and connecting these portfolios' images in (standard deviation, expected return) space. The frontier is called minimum-variance frontier. Its upper part is called nondominated frontier. Although standard deviation is just the square root of variance, standard deviation is more widely utilized because it is in the same units as expected return and allows many elegant results in financial mathematics. However, in computation it is more helpful to use variance upon which quadratic programming can be employed.

Also Markowitz [63, 66] proposed the following model

$$\begin{aligned} \min\,&\{V[R(\mathbf{x},\mathbf{r})] = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}\} \\ \max\,&\{E[R(\mathbf{x},\mathbf{r})] = \boldsymbol{\mu}^T\mathbf{x}\} \\ \text{s.t.}\quad & \mathbf{1}^T\mathbf{x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{3.10}$$

He called this model the standard portfolio selection model and devised what is known as the critical line algorithm to compute the nondominated frontier. His critical line algorithm is more efficient than e-constraint approach (3.5). His algorithm is outlined in Chapter 4.

In several finance textbooks such as Huang and Litzenberger [43] and Ingersoll [46], a variant of (3.10) is deployed as follows

$$\begin{aligned} \min\{\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} = z\} \\ \text{s.t. } \boldsymbol{\mu}^T\mathbf{x} = e \\ \mathbf{1}^T\mathbf{x} = 1 \end{aligned} \tag{3.11}$$

where e is some target expected return. This model can be re-expressed in two criterion form as follows

$$\begin{aligned} \min\,&\{V[R(\mathbf{x},\mathbf{r})] = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}\} \\ \max\,&\{E[R(\mathbf{x},\mathbf{r})] = \boldsymbol{\mu}^T\mathbf{x}\} \\ \text{s.t.}\quad & \mathbf{1}^T\mathbf{x} = 1 \end{aligned} \tag{3.12}$$

This model is called unlimited short-sales allowed model, because, without any lower bound on the x_i, unlimited short sales are allowed. The advantages of (3.11) are

1. All the results are available closed-form formulae as shown in Merton [70] and Huang and Litzenberger [43]. This avoids the necessity of mathematical programming, which brings convenience in research and teaching.

2. It can serve as the basis for Capital Asset Pricing Model (CAPM) by Sharpe [83].

An example is introduced to serve for three purposes.

1. The example illustrates the criterion space, feasible region Z in criterion space, nondominated set N, decision space, feasible region S in decision space, and efficient set E of a nonlinear case.

2. The example illustrates a portfolio selection problem through the view of multiple criteria optimization. For example, the whole feasible region Z is portrayed. Whereas, in some finance texts (for example Huang and Litzenberger [43], and Bodie, Kane, and Marcus [10]), only a minimum-variance frontier or a nondominated frontier is illustrated, and the right border is ignored.

3. The example shows the difference between the two models, (3.10) and (3.12).

Example 1. An investor is considering the information of three securities as follows

$$\boldsymbol{\mu} = \begin{bmatrix} 0.03 \\ 0.10 \\ 0.07 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 0.036 & 0.005 & 0.007 \\ 0.005 & 0.030 & 0.005 \\ 0.007 & 0.005 & 0.042 \end{bmatrix}$$

The investor applies model (3.10) and sees the feasible region Z in (standard deviation, expected return) space in Figure 3.1 *top*. Note that (standard deviation, expected return) space is deployed as usual in finance, although the criterion space of (3.10) is (variance-expected return) space. The feasible region Z is represented by the shaded region in Figure 3.1 *top*. The three securities are represented by dots $\mathbf{z}^1 = (0.19, 0.03)$, $\mathbf{z}^2 = (0.17, 0.10)$ and $\mathbf{z}^3 = (0.21, 0.07)$. The nondominated frontier N is marked by the thick curve. The feasible region $Z \subset \mathbb{R}^2$ is continuous and bounded. The decision space is $\mathbb{R}^3$. The three points $\mathbf{x}^1 = (1, 0, 0)$, $\mathbf{x}^2 = (0, 1, 0)$, and $\mathbf{x}^3 = (0, 0, 1)$ stand for the three securities in Figure 3.1 *bottom*. The feasible region S is illustrated as the shaded triangular area. The efficient set E is composed of two linear segments as the thick part in graph.

Switching to model (3.12), the investor observes N as the upper part of an unbounded

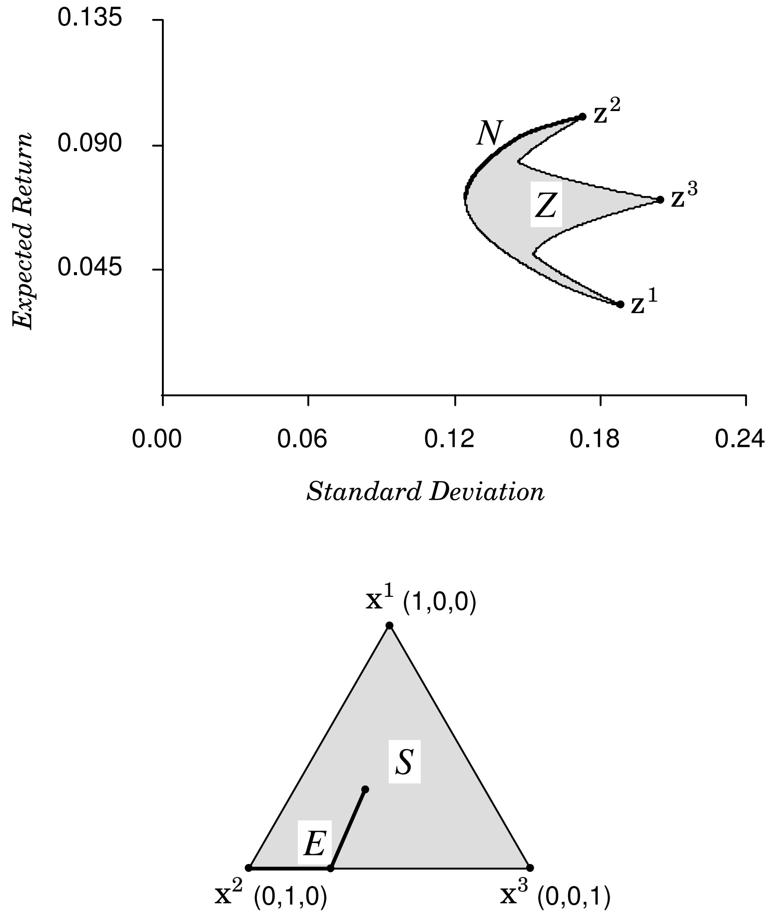

Figure 3.1: Feasible regions Z and S, nondominated frontier, and efficient set of (3.10)

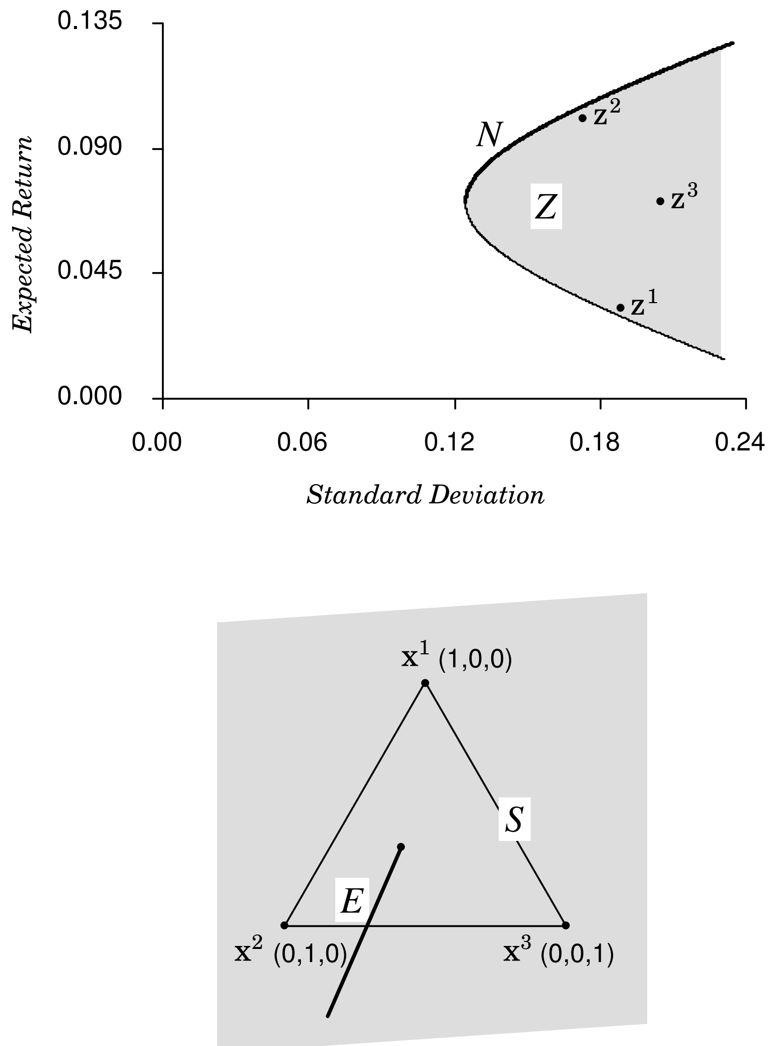

Figure 3.2: Feasible regions Z and S, nondominated frontier, and efficient set of (3.12)

hyperbola in Figure 3.2 *top*. The Z is represented by the shaded area in Figure 3.2 *top*. The Z is still continuous but unbounded, because the S is unbounded. The S is the shaded area in Figure 3.2 *bottom*. Representing the plane $x_1 + x_2 + x_3 = 1$ in $\mathbb{R}^3$, S is unbounded. The E is an unbounded ray on the plane. ◀

Chapter 4

Quadratic Parametric Programming for Portfolio Selection

This chapter outlines Markowitz's [66] critical line algorithm and Hirschberger's procedure developed in Hirschberger, Qi and Steuer [40] for computing nondominated frontiers of mean-variance portfolio selection problems. The outline can form the basis of a future project to systematically compare the two methods. Whereas initial comparisons show that Markowitz's critical line algorithm becomes faster than Hirschberger's procedure after about 200 securities, one advantage of Hirschberger's procedure is that it can be extended to additional linear criteria.

Both Markowitz's critical line algorithm and Hirschberger's procedure utilize a weighted-sums approach similar to

$$\max\{-\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} + \lambda\,\boldsymbol{\mu}^T\mathbf{x} = z_w\} \quad \lambda \geq 0 \qquad (4.1)$$
$$s.t. \qquad \mathbf{x} \in S$$

where λ is a parameter. As λ changes continuously from 0 to ∞, all efficient portfolios are located. That is, Markowitz's critical line algorithm and Hirschberger's procedure belong to quadratic parametric programming.

The difference between Markowitz's critical line algorithm and e-constraint approach is now graphically demonstrated, based on the three-security portfolio selection problem where the only constraints are $\mathbf{1}^T\mathbf{x} = 1$ and $\mathbf{x} \geq \mathbf{0}$ of Example 1 of Chapter 3.

First, Markowitz's critical line algorithm starts with $\lambda = \infty$ and locates $\mathbf{x}^2$ as the corresponding maximizing solution of (4.1). Second, Markowitz's critical line algorithm determines $\lambda = 0.515$ and locates $\mathbf{x}^4$ as the corresponding maximizing solution. At last, Markowitz's critical line algorithm determines $\lambda = 0$ and locates $\mathbf{x}^5$ as the corresponding maximizing solution. The algorithm stops at $\lambda = 0$. The nondominated frontier is in Figure 4.1 *top*. The images of $\mathbf{x}^2, \mathbf{x}^4$, and $\mathbf{x}^5$ are dots $\mathbf{z}^2, \mathbf{z}^4$, and $\mathbf{z}^5$, respectively. The nondominated frontier consists of two hyperbolic segments, with one segment from $\mathbf{z}^2$ to $\mathbf{z}^4$ and the other segment from $\mathbf{z}^4$ to $\mathbf{z}^5$. The expressions of the hyperbolic segments can be computed by Markowitz's critical line algorithm.

The situation in decision space is in Figure 4.1 *bottom*. The three securities' portfolio weights are $\mathbf{x}^1 = (1, 0, 0), \mathbf{x}^2 = (0, 1, 0)$, and $\mathbf{x}^3 = (0, 0, 1)$. The efficient set consists of two linear segments, $\mathbf{x}^2$ to $\mathbf{x}^4$ and $\mathbf{x}^4$ to $\mathbf{x}^5$. As $\lambda \in [0, 0.515]$, the maximizing solutions of (4.1) are on the linear segment from $\mathbf{x}^5$ to $\mathbf{x}^4$. As $\lambda \in [0.515, \infty]$, the maximizing solutions of (4.1) are on the linear segment from $\mathbf{x}^4$ to $\mathbf{x}^2$.

In finance, e-constraint approach (3.5) is often used. It works as follows, (a) change the expected return objective into an equality e-constraint, (b) specify a set of right-hand-side (parameters) of the e-constraint. Take, for example, four right-hand-side as 0.030, 0.050, 0.080, and 0.100, (c) solve the e-constraint problem four times based on the four right-hand-side and get four portfolios, (d) connect the images of these portfolios to form a minimum-variance frontier in (standard deviation, expected return) space, (e) obtain the nondominated frontier from the minimum-variance frontier. The images of the four portfolios are $\mathbf{z}^6, \mathbf{z}^7, \mathbf{z}^8$, and $\mathbf{z}^9$ in Figure 4.2. The minimum-variance frontier consists of the linear segments from $\mathbf{z}^6$ to $\mathbf{z}^7$, $\mathbf{z}^7$ to $\mathbf{z}^8$, and $\mathbf{z}^8$ to $\mathbf{z}^9$. The nondominated frontier consists of the linear segment from $\mathbf{z}^8$ to $\mathbf{z}^9$. It must be pointed out that the e-constraint approach does not locate the true (complete) nondominated frontier and minimum-variance frontier (which is the gray line in Figure 4.2). What the e-constraint approach provides is a discretized representation of the true (complete) nondominated frontier and minimum-variance frontier. Based on these figures, it can be seen that Markowitz's critical line algorithm has at least the following properties.

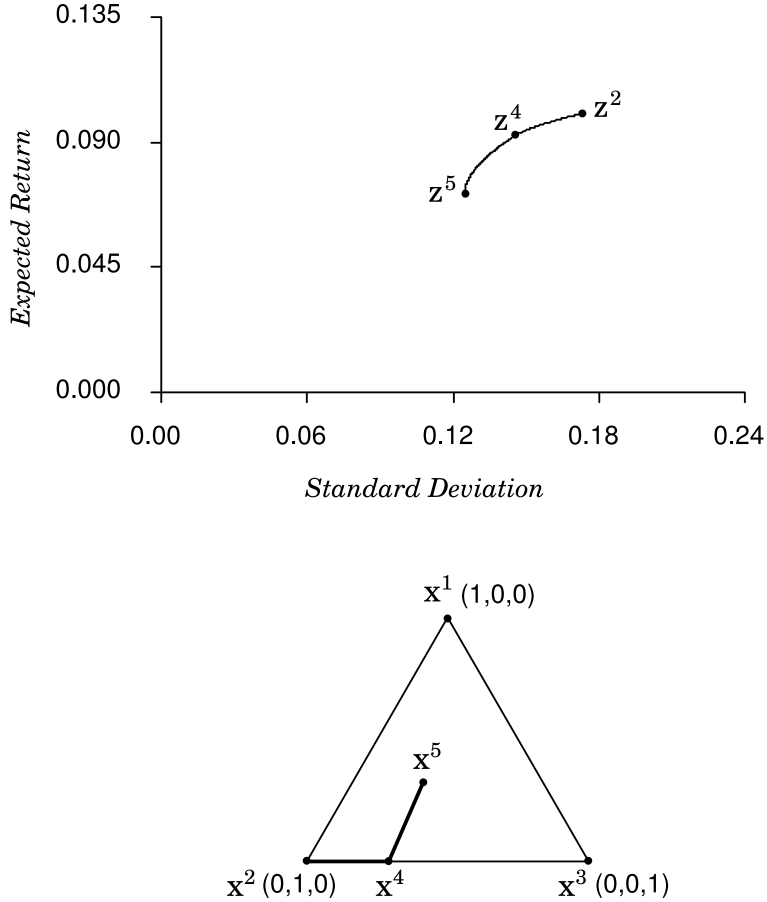

Figure 4.1: The nondominated frontier and efficient set of a three-security portfolio selection computed by Markowitz's critical line algorithm

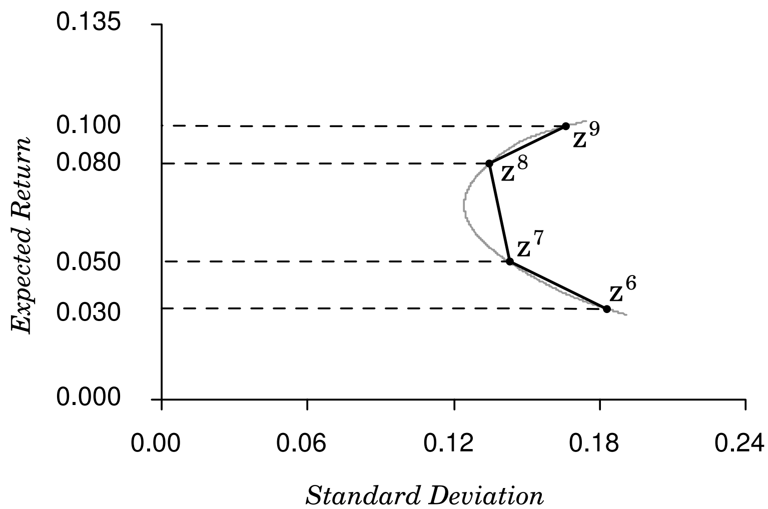

Figure 4.2: The nondominated frontier of a three-security portfolio selection computed by e-constraint approach

1. Markowitz's critical line algorithm computes the exact expression of a nondominated frontier, instead of repetitive solving.

2. Markowitz's critical line algorithm demonstrates the piece-wise hyperbolic structure of nondominated frontiers which method just described does not.

4.1 Outline of Critical Line Algorithm

The solution process for Markowitz's [62, 66] critical line is as follows.

1. Form

$$\begin{aligned} \min\{\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} - \lambda_E\boldsymbol{\mu}^T\mathbf{x} = z_w\} \\ \text{s.t.} \qquad \mathbf{A}\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{4.2}$$

The idea is that this problem will locate all the efficient points for $\lambda_E \in [0, \infty)$.

2. Set up Lagrangian function as follows.

$$\begin{aligned} \min\{\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} - \lambda_E\boldsymbol{\mu}^T\mathbf{x} + \boldsymbol{\lambda}^T\mathbf{A}\mathbf{x} = L(\mathbf{x}, \lambda_E, \boldsymbol{\lambda})\} \\ \text{s.t.} \quad \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{4.3}$$

Apply Kuhn-Tucker Conditions [55] to (4.3) with constraints $\mathbf{x} \geq \mathbf{0}$ handled implicitly by sets IN and OUT.

3. Based on sets IN and OUT, $\mathbf{x}$ with $\lambda_E > 0$ is efficient if

$$\begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \boldsymbol{\alpha} + \lambda_E\boldsymbol{\beta} \tag{4.4}$$

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{A}^T & \boldsymbol{\mu} \end{bmatrix}_{n\times(n+m+1)} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \\ -\lambda_E \end{bmatrix}_{(n+m+1)\times 1} \geq \mathbf{0} \tag{4.5}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be computed without depending on $\mathbf{x}, \lambda_E$ and $\boldsymbol{\lambda}$. (4.4) defines a straight line which is named as *critical line* by Markowitz.

4. Compute λ_{low} and λ_{high}. If

$$\lambda_E \in [\lambda_{low}, \lambda_{high}] \tag{4.6}$$

and (4.4) and other conditions hold, then (4.4) and (4.6) define an efficient linear segment whose end points—$\mathbf{x}$ with respect to $\lambda_E = \lambda_{low}$ and $\lambda_E = \lambda_{high}$ are named *corner portfolios* by Markowitz. Nevertheless, the word "portfolio" suggests criterion space and an investor will not necessarily see a corner on a nondominated frontier. Therefore, the author calls these $\mathbf{x}$ *turning points*.

5. The starting point for the algorithm is computed by solving linear programming problem

$$\begin{aligned} \max\ \boldsymbol{\mu}^T\mathbf{x} &= z_h \\ \text{s.t.} \qquad \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \tag{4.7}$$

for the $\mathbf{x}$ that generates the highest expected return. At this point, sets IN and OUT can be identified, $\lambda_{high} = \infty$, and then λ_{low} is computed to delineate the first segment.

6. The λ_{high} of the next iteration based on next IN and OUT equals λ_{low} of the current iteration based on current IN and OUT and the corresponding turning points coincide i.e. adjacent efficient linear segments are connected.

7. Based on **6**, perform an iteration (changing IN and OUT), and repeat **3** and **4** until the domain of $\lambda_E \in [0, \infty)$ is covered. Then the efficient set is continuous piecewise linear and the nondominated frontier is continuous piecewise hyperbolic.

Example 1. Continue on the three-security portfolio selection problem of Example 1 in Chapter 3. The investor utilizes standard portfolio selection model (3.10) ($\mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}$). The nondominated frontier in (standard deviation, expected return) space is in Figure 4.1 *top*. There are three turning points, $\mathbf{x}^2 = (0, 1, 0)$, $\mathbf{x}^4 = (0, 0.72, 0.28)$, and $\mathbf{x}^5 = (0.32, 0.42, 0.26)$. The images of the turning points are $\mathbf{z}^2$, $\mathbf{z}^4$, and $\mathbf{z}^5$. The nondominated frontier consists of

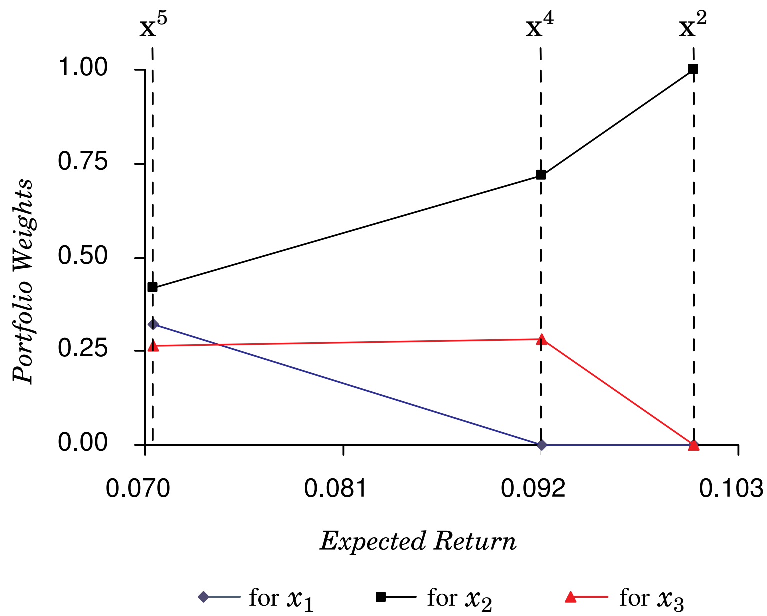

Figure 4.3: Portfolio weights (x_1, x_2, x_3) of the efficient set of Example 1

two hyperbolic pieces from $\mathbf{z}^5$ to $\mathbf{z}^4$ and from $\mathbf{z}^4$ to $\mathbf{z}^2$. The efficient set consists of two linear segments from $\mathbf{x}^5$ to $\mathbf{x}^4$ corresponding to $\lambda_E \in [0, 0.515]$ and from $\mathbf{x}^4$ to $\mathbf{x}^2$ corresponding to $\lambda_E \in [0.515, \infty)$. The efficient set is depicted by the thick line in Figure 4.1 *bottom*.

The portfolio weights $\mathbf{x} = (x_1, x_2, x_3)$ for the efficient set is in Figure 4.3. The three lines represent the three securities' weights over the efficient set, that is, x_1, x_2, and x_3 with respect to expected returns of the efficient portfolios. For example, the top line is for x_2. This line contains three ■ corresponding to x_2 of turning points $\mathbf{x}^5, \mathbf{x}^4$, and $\mathbf{x}^2$ from left to right. That is, $x_2^5 = 0.42, x_2^4 = 0.72$, and $x_2^2 = 1$. ◀

The piece-wise linear segment structure of an efficient set indicates that on one efficient linear segment excluding the turning points, some weights keep the status of positive, while the others are always 0 i.e.

$$x_i > 0 \text{ for } i \in \text{ set } IN, \ \lambda_E \in (\lambda_{low}, \lambda_{high}) \tag{4.8}$$

$$x_i = 0 \text{ for } i \in \text{ set } OUT, \lambda_E \in (\lambda_{low}, \lambda_{high})$$

The financial interpretation is that the portfolios on the corresponding piece of frontier are composed of by the same group of securities and the difference is just the varying weights. This is valuable in rebalancing portfolios. An investor can locate the efficient linear segment closest to her or his current portfolio and determine her or his target portfolio weight on this segment.

Example 2. Continue on Example 1. The portfolio weights $\mathbf{x} = (x_1, x_2, x_3)$ for the efficient set is in Figure 4.4. Suppose the investor's current portfolio weight is $(.15, .5, .35)$, which is not efficient. Then from the efficient set, the linear segment from $\mathbf{x}^5$ to $\mathbf{x}^4$ with $x_1 > 0, x_2 > 0$, and $x_3 > 0$ excluding $\mathbf{x}^4$ is a good match. If the investor wants to keep $x_1 = .15$, then he draws a horizontal broken line from .15 to hit the line of x_1 at point $x_1^6 = .15$. Then he draws a vertical broken line passing through $x_1^6 = .15$ to hit the line of x_2 at point $x_2^6 = .58$ and the line of x_3 at $x_3^6 = .27$. Therefore, the target portfolio weight $\mathbf{x}^6 = (.15, .58, .27)$ is obtained. ◀

4.2 Hirschberger's Procedure

Hirschberger, Qi and Steuer [40] proposed a procedure to compute nondominated frontiers of general portfolio selection problems as well. The procedure was coded in Java for public domain use on modern desktops and laptops by Hirschberger. Hirschberger's procedure computes the same results as Markowitz's critical line algorithm. Compared to Markowitz's critical line algorithm, Hirschberger's procedure has the following characteristics.

1. Hirschberger's procedure adopts well-known methods from operations research including Simplex method, Kuhn-Tucker conditions, Wolfe's method and Guddat's method and therefore can have greater understandability.
2. Hirschberger's procedure is purposely designed to permit its current *uni*-parametric capability to be generalized to embrace *multi*-parametric quadratic applications with additional linear objectives in portfolio.

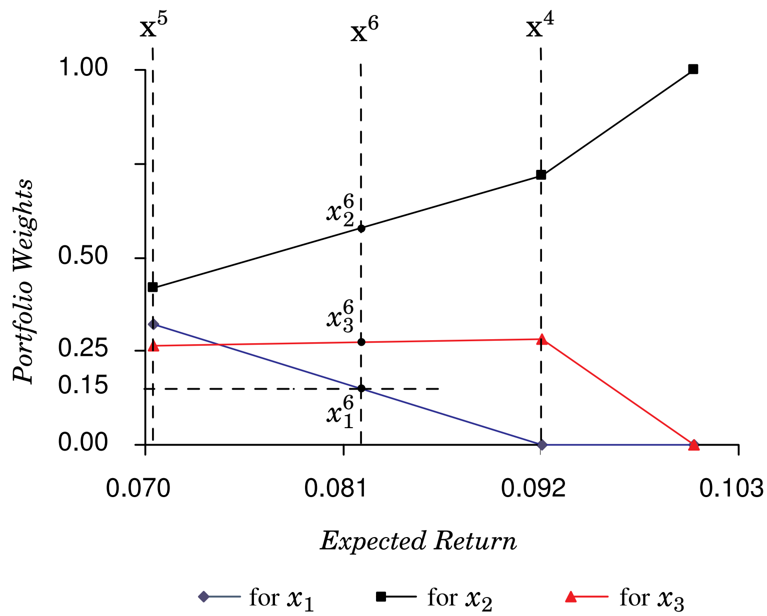

Figure 4.4: Usage of the piece-wise linear segment structure of an efficient set to obtain $\mathbf{x}^6$ in rebalancing portfolios

3. Hirschberger's procedure can compute a whole feasible region Z, in addition to a nondominated frontier.

Hirschberger's procedure is outlined is as follows.

1. Form a weighted-sums approach as follows

$$\max\{-\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} + \lambda\,\boldsymbol{\mu}^T\mathbf{x} = z_w\} \quad \lambda \geq 0 \tag{4.9}$$

$$s.t. \quad \check{\mathbf{H}}\mathbf{x} = \check{\mathbf{d}}$$

$$\mathbf{G}\mathbf{x} \leq \check{\mathbf{b}}$$

$$\mathbf{x} \leq \boldsymbol{\omega}$$

$$\mathbf{x} \geq \boldsymbol{\ell}$$

As λ changes from 0 to ∞, all the efficient points are computed.

2. Simplify (4.9) by translating the axis system to $\boldsymbol{\ell}$ and obtain.

$$\max\{-\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} + \lambda\,\boldsymbol{\mu}^T\mathbf{x} - 2\,\boldsymbol{\ell}^T\boldsymbol{\Sigma}\mathbf{x} = z\} \quad \lambda \geq 0 \tag{4.10}$$

$$\begin{aligned} s.t. \quad \mathbf{H}\mathbf{x} &= \mathbf{d} \\ \mathbf{G}\mathbf{x} &\leq \mathbf{b} \\ \mathbf{x} &\leq \boldsymbol{\beta} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

3. Apply Kuhn-Tucker conditions to (4.10), reduce the problem size by combining unimportant vectors, and obtain.

$$2\boldsymbol{\Sigma}\bar{\mathbf{x}} + \mathbf{H}^T\mathbf{v} + \mathbf{G}^T\mathbf{u}^s + \mathbf{D}\bar{\mathbf{u}} = -2\boldsymbol{\Sigma}\boldsymbol{\ell} + \lambda\boldsymbol{\mu} \tag{4.11}$$

$$\mathbf{H}\bar{\mathbf{x}} = \mathbf{d} \tag{4.12}$$

$$\mathbf{G}\bar{\mathbf{x}} + \mathbf{I}_m\mathbf{s} = \mathbf{b} \tag{4.13}$$

$$\bar{\mathbf{u}}^T\bar{\mathbf{x}} = 0\,, \qquad \mathbf{u}^{s\,T}\mathbf{s} = 0 \tag{4.14}$$

$$\bar{\mathbf{x}} \geq \mathbf{0}\,,\; \bar{\mathbf{u}} \geq \mathbf{0}\,,\; \mathbf{u}^s \geq \mathbf{0}\,,\; \mathbf{s} \geq \mathbf{0} \tag{4.15}$$

$$\mathbf{v} \text{ unrestricted} \tag{4.16}$$

4. Three phases are employed. In Phase I, a feasible solution to (4.12–4.13) is found. In Phase II the solution $(\bar{\mathbf{x}}, \mathbf{s})$ from Phase I is utilized to obtain a solution that satisfies (4.11-4.16) when $\lambda = 0$. The solution obtained in Phase II is the minimum-variance portfolio of the nondominated frontier. Based on the result of Phase II, Phase III parameterizes λ from 0 to ∞ to compute the other solutions of (4.11-4.16) corresponding to the rest of the nondominated frontier.

5. Phase I is done by solving the linear program

$$\min \{ \mathbf{1}^T \mathbf{a}^1 + \mathbf{1}^T \mathbf{a}^2 = z \} \tag{4.17}$$

$$\begin{aligned} s.t. \quad \mathbf{H}\bar{\mathbf{x}} \qquad\qquad\qquad + \mathbf{I}_l \mathbf{a}^2 &= \mathbf{d} \\ \mathbf{G}\bar{\mathbf{x}} + \mathbf{I}_m \mathbf{s} - \mathbf{I}_m \mathbf{a}^1 \qquad &= \mathbf{b} \\ \bar{\mathbf{x}} \geq \mathbf{0}\,,\ \mathbf{s} \geq \mathbf{0}\,,\ \mathbf{a}^1 \geq \mathbf{0}\,,\ \mathbf{a}^2 &\geq \mathbf{0} \end{aligned}$$

6. Phase II is achieved by solving the linear program

$$\min \{ \mathbf{1}^T \mathbf{a}^3 = z \} \tag{4.18}$$

$$2\boldsymbol{\Sigma}\bar{\mathbf{x}} + \mathbf{H}^T\mathbf{v} + \mathbf{G}^T\mathbf{u}^s + \mathbf{D}\bar{\mathbf{u}} \qquad + \mathbf{E}\mathbf{a}^3 = -2\boldsymbol{\Sigma}\boldsymbol{\ell} \tag{4.19}$$

$$\mathbf{H}\bar{\mathbf{x}} \qquad\qquad\qquad = \mathbf{d} \tag{4.20}$$

$$\mathbf{G}\bar{\mathbf{x}} \qquad\qquad + \mathbf{I}_m\mathbf{s} \qquad = \mathbf{b} \tag{4.21}$$

$$\bar{\mathbf{x}} \geq \mathbf{0},\ \bar{\mathbf{u}} \geq \mathbf{0}\,,\ \mathbf{u}^s \geq \mathbf{0},\ \mathbf{s} \geq \mathbf{0}, \mathbf{a}^3 \geq \mathbf{0} \tag{4.22}$$

$$\mathbf{v} \ \text{unrestricted} \tag{4.23}$$

Wolfe's method [95] with upper bound is utilized.

7. Phase III relies on Guddat's method [32]. The key steps are computing $\boldsymbol{\xi}^h$, $\boldsymbol{\Delta}^h$, and λ^{h+1} and setting up $\mathbf{x}^{h+1} = \boldsymbol{\xi}^h + \lambda^{h+1}\boldsymbol{\Delta}^h$ which is a turning point of an efficient linear segment and proceed to the next iteration with $h = h + 1$.

8. It stops if $\lambda^{h+1} = \infty$.

Markowitz's critical line algorithm was coded in Visual Basic for Applications (VBA). Unfortunately, due to the 256 column limit of Excel, the program of Markowitz's critical line algorithm can only solve portfolio selection problems for less than or equal to 256 securities, while the program of Hirschberger's procedure can solve portfolio selection problems up to 3,000 securities. An initial computational comparison between Markowitz's critical line algorithm and Hirschberger's procedure is performed. The comparison is based on randomly generated standard portfolio selection problems ($\mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}$). Randomly generating portfolio selection problems with specified distributional characteristics is discussed in Chapter

Table 4.1: Results for computing the nondominated frontiers of problems with 60, 120, 180, 200, 210, 245 securities by Markowitz's critical line algorithm and Hirschberger's procedure

Securities	Sample size (problems)	CPU time (seconds) of critical line algorithm		CPU time (seconds) of Hirschberger's procedure	
n		total time	solving time per problem	total time	solving time per problem
60	750	888	0.35	166	0.07
120	150	306	1.23	146	0.48
180	50	212	2.49	150	1.93
200	45	219	2.92	183	2.74
210	40	212	3.20	200	3.54
245	20	142	4.46	153	5.59

5. The results are in Table 4.1. Table 4.1 demonstrates the solution time. For example, the first row means that this is a 60 security portfolio selection model, 750 problems of this model is tested, the total running time including inputting the problems, solving the problems, and outputting the solutions by Markowitz's critical line algorithm is 888 seconds, the average time for solving a problem by Markowitz's critical line algorithm is 0.35 second, the total running time by Hirschberger's procedure is 888 seconds, the average time for solving a problem by Hirschberger's procedure is 0.07 second. It can be seen that Hirschberger's procedure is faster for problems less than 200 securities, while Markowitz's critical line algorithm is faster beyond 200 securities. More comprehensive comparisons between Markowitz's critical line algorithm and Hirschberger's procedure will be taken in the future, for example, the sensitivity of time, number of turning points, and number of securities to $n \times n$ size of covariance matrix, upper bounds, randomly generated data versus historical data.

In another experiment only for Hirschberger's procedure, with a sample size of 30 in each case, computational results for standard portfolio selection problems with 500, 1,000, 1,500

n	% Point up Frontier	Cumulative Turning Points		Numb of Secs at % Point		Incremental CPU secs.	
		Ave	Stdev	Ave	Stdev	Ave	Stdev
500	0	1.0	0.0	35.4	3.3	70.71	19.51
	10	58.6	12.3	52.8	4.3	–	–
	25	101.0	18.4	51.5	5.2	–	–
	50	153.4	18.7	32.7	11.1	–	–
	75	178.6	14.9	10.9	5.1	–	–
	100	188.3	14.0	1.0	0.0	1.46	0.13
Total	–	–	–	–	–	72.17	19.55
1000	0	1.0	0.0	40.2	3.9	592.43	193.81
	10	84.7	21.5	58.6	4.2	–	–
	25	143.3	32.3	56.2	6.5	–	–
	50	200.3	30.3	31.2	13.6	–	–
	75	225.2	20.9	10.0	6.1	–	–
	100	234.1	17.4	1.0	0.0	9.52	1.31
Total	–	–	–	–	–	601.95	193.73
1500	0	1.0	0.0	41.4	3.9	2560.0	618.1
	10	88.7	20.3	60.8	4.1	–	–
	25	148.2	21.7	62.7	4.1	–	–
	50	212.2	25.0	41.3	13.5	–	–
	75	248.4	20.2	14.4	7.7	–	–
	100	263.9	20.4	1.0	0.0	23.6	2.3
Total	–	–	–	–	–	2583.6	617.5

and 2,000 securities are shown in Table 4.2. In the Cumulative Turning Point columns one can see how the nondominated frontier consists of many hyperbolic segments (where the number of hyperbolic segments is one less than the number of turning points) and how the bulk of the hyperbolic segment activity takes place, at least in the larger problems, in the first 25% up the nondominated frontier. In the Number of Securities columns, one can see how the number of securities in portfolios along the nondominated frontier at first increases, hitting its peak at about the 10% mark, and then from there commences its long decline.

n	% Point up Frontier	Cumulative Turning Points		Numb of Secs at % Point		Incremental CPU secs.	
		Ave	Stdev	Ave	Stdev	Ave	Stdev
2000	0	1.0	0.0	43.0	4.7	6246.2	1400.0
	10	101.7	25.1	64.0	3.6	–	–
	25	166.9	33.9	61.3	10.6	–	–
	50	232.5	33.8	40.1	17.6	–	–
	75	267.6	23.2	14.2	11.1	–	–
	100	280.8	20.2	1.0	0.0	49.1	4.3
Total	–	–	–	–	–	6295.3	1400.7

Table 4.2: Results of runs of 30 problems each for computing the nondominated frontiers of problems with 500, 1,000, 1,500 and 2,000 securities

One can also note how the number of securities only grows modestly with n. In the Incremental CPU Seconds columns, what is remarkable is how quickly the nondominated frontier can be calculated once the minimum standard deviation point has been obtained, and how the distinction becomes more pronounced with problem size. For instance, in 200-security problems (not shown), the split is roughly 96% vs. 4%, but in the 2,000-security problems, the split widens to more than 99% vs. less than 1%. Whereas large scale was roughly 500 securities in 1984 at the time of Perold's paper [78], with Hirschberger's procedure and the corresponding code, problems (to which no covariance matrix simplification techniques have been applied) with several times this many securities are now within reach of anyone even on a laptop. All CPU times are from a Dell 3.06GHz desktop at the University of Georgia.

4.3 Qi's Graphing Program Connected to Hirschberger's Procedure

Markowitz [66] proposed to obtain the lower part of a minimum-variance frontier (lower left border) of a portfolio selection problem as follows

$$\begin{aligned} &\min \{V[R(\mathbf{x},\mathbf{r})] = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}\} \\ &\min \{E[R(\mathbf{x},\mathbf{r})] = \boldsymbol{\mu}^T\mathbf{x}\} \\ &\text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{4.24}$$

Markowitz also proposed an algorithm to obtain the right border of a portfolio selection problem. Hirschberger's procedure can compute the feasible region Z of a problem by computing the nondominated frontier, the lower part of a minimum-variance frontier (lower left border), and the right border. Qi programmed in Visual Basic for Applications (VBA) to plot the n securities, turning points, and the feasible region Z for Hirschberger's procedure.

Example 3. Randomly select 250 constituents from Standard & Poor's Super Composite 1500 Index from January 1997 to December 2002 (monthly returns) as the asset universe. [1] Take standard portfolio selection model ($\mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}$). Apply Hirschberger's procedure to solve. The results are in Figure 4.5 *top*. The 250 constituents are plotted as the hollow black dots. The 70 turning points are plotted as the solid black dots on the nondominated frontier. The feasible region Z is the shaded area. Then randomly add another 250 constituents to increase the universe to 500 constituents and repeat the process. The 500 constituents, 104 turning points, and feasible region Z are in Figure 4.5 *bottom*. ◀

[1] Data source: the Center for Research in Security Prices (CRSP) through Wharton Research Data Service (WRDS) at the University of Pennsylvania and the website of Standard & Poor's 500 Index http://www.spglobal.com on July 20, 2003. Because of corporations merges, acquisitions, dropping out of the index, or possess missing values, 1078 constituents with full records are utilized instead of 1500 constituents.

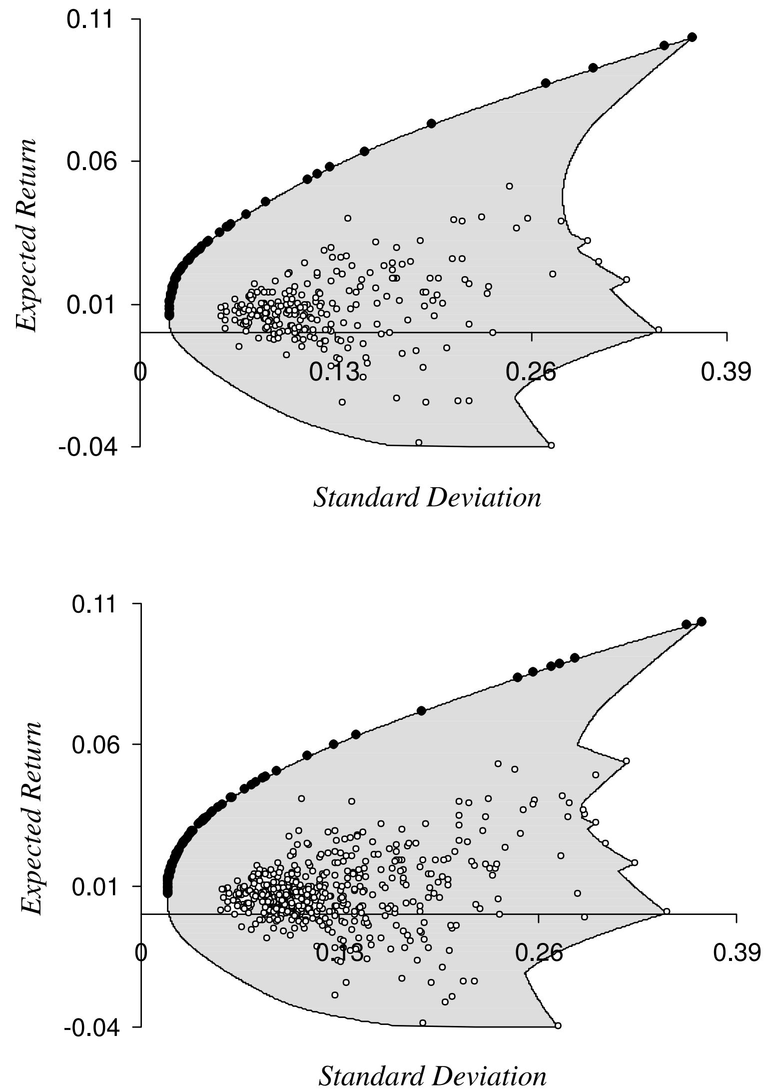

Figure 4.5: The asset universes, turning points, and feasible regions of a 250-constituents portfolio selection problem and a 500-constituents portfolio selection problem selected from Standard & Poor's Super Composite 1500 Index

Chapter 5

Generating Covariance Matrices

This chapter outlines a procedure to generate a series of covariance matrices proposed by Hirschberger, Qi and Steuer [41]. This procedure allows the elements in the covariance matrices to have distributional characteristics closely matching those of empirical data. Furthermore, it also allows (a) the main diagonal (variance) elements of the matrices to possess a pre-specified expected value and standard deviation and (b) the off-diagonal (covariance) elements to possess a likely different pre-specified expected value and standard deviation. With the ability to generate a series of such matrices that match pre-specified distributional characteristics, the procedure can easily provide an unlimited supply of covariance matrices for benchmarking and comparing portfolio selection optimization algorithms, whereas this would be difficult, if not impossible with historical data.

A square matrix is a covariance matrix if and only if it is *positive semidefinite* described by Brockwell and Davis [12]. This is an important property of covariance matrices.

Definition 5.1. *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$*. Then* $\mathbf{A}$ *is* positive semidefinite *if and only if* $\mathbf{A}$ *is symmetric and*

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$$

for all $\mathbf{x} \in \mathbb{R}^n$*.*

One of practical ways to check for positive semidefiniteness is as follows. A square matrix is positive semidefinite if and only if all of its eigenvalues are nonnegative. Furthermore, a covariance matrix is *positive definite* if and only if it is invertible as pointed out by Markowitz and Todd [66].

Definition 5.2. *Let* $\mathbf{A} \in \mathbb{R}^{n \times n}$*. Then* $\mathbf{A}$ *is* positive definite *if and only if* $\mathbf{A}$ *is symmetric and*

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$$

for all $\mathbf{x} \in \mathbb{R}^n$*.*

5.1 Motivation

Since the birth of mean-variance approach by Markowitz [61], the research of proposing models and solution algorithms has developed considerably. For example, see Markowitz [62, 66], Markowitz, Todd, Xu and Yamane [68, 67], Elton, Gruber and Padberg [26, 27, 28], Pang [76], and Perold [78], and in the multiple criteria arena, see Bana e Costa and Soares [6], Bouri, Martel and Chabchoub [11], Ehrgott, Klamroth and Schwehm [24], Guerard and Mark [33], Konno, Shirakawa and Yamazaki [51], Konno and Suzuki [52], Ogryczak [73], Steuer, Qi and Hirschberger [90], and Hirschberger, Qi and Steuer [39]. However, there is a dearth of computational result papers which report the time to solve portfolio selection problems, the number of turning points, and the number of securities in portfolios of different sizes (especially large-scale), and other characteristics.

The reason may be that large numbers of different covariance matrices of all sizes are required in the computations, but this universe of covariance matrices is unavailable. Deploying historical data or simply creating the matrices without guidance will be proven hardly achievable. If this is the case, then how has conventional portfolio selection optimization [1] managed to survive over the years? Apart from obtaining simplistic covariance matrices from averaging techniques as described in Chapter 8 of Elton, Gruber, Brown and Goetzman [25], the field has been resigned, since there has been no other way, to generate the necessary covariance matrices from historical data. While this may come with the territory on applications, this hardly meets the needs of research when all manners of covariance matrices may

[1] The dissertation is focused on models based on standard covariance matrices ushered by Markowitz [61]. Computational simplification models such as factor models are not included. However, the author acknowledges their contributions.

be needed for experimenting with, testing, benchmarking and comparing various strategies, approaches and algorithms.

Equipped with the procedure developed by Hirschberger, Qi and Steuer [41], researchers can not only generate all kinds of covariances matrices of any size, but can also require the diagonal elements of the generated covariance matrices to possess a pre-specified expected value and standard deviation and the off-diagonal elements to possess a likely different pre-specified expected value and standard deviation.

Utilizing empirical data has been the mainstay of portfolio selection problems, but five difficulties can be expected when attempting to obtain covariance matrices of meaningful size from historical data.

1. Access to a sufficient database must be obtained, which is often expensive.

2. Users must often endure the headache of debugging missing or defective data and the task of transforming the format, typically time consuming processes.

3. The random sampling from historical data can produce similar distributional characteristics of the covariance matrices. Distributional characteristics include mean, standard deviation, and skewness of the variances (diagonal elements) and covariances (off-diagonal elements) of a covariance matrix $\mathbf{\Sigma}$.

Example 1. Randomly select four groups of stocks from the securities from the Standard & Poor's Super Composite 1500 Index from January 1997 to December 2002 (monthly returns) with 250 stocks in each group and no overlapping of stocks among the groups. [2] For each group, compute the 250 variances and $\frac{250\times(250-1)}{2} = 31,125$ covariances of the returns of the 250 stocks. The distributions in the form of kernel density functions of the variances and covariances of the four groups are represented by the four graphs in Figure 5.1. A kernel density function is an estimate of the density function and can be considered as a refinement of a histogram or frequency plot, see

[2] Data source: the Center for Research in Security Prices (CRSP) through Wharton Research Data Service (WRDS) at the University of Pennsylvania and the website of the Index at http://www.spglobal.com on July 20, 2003.

Table 5.1: Means and standard deviations of the four random non-overlapping 250-stock groups from Standard & Poor's Super Composite 1500 Index of Example 1

		Group 1	Group 2	Group 3	Group 4
variances	mean, $\bar{e}$	.017527	.017729	.016242	.029842
	standard deviation, $\sqrt{\bar{v}}$	.017606	.015539	.015242	.028037
covariances	mean, e	.002505	.002852	.002521	.004497
	standard deviation, $\sqrt{v}$	.003488	.003616	.002857	.006349

SAS [3]. The thick lines and thin lines symbolize the distributions of variances and covariances, respectively. The variances express much greater variation (dispersion) than the covariances. The mean of the variances is bigger than that of the covariances. Both the variances and covariances are right-skewed. In fact, this becomes the general pattern of later experiments. Surprisingly and interestingly, all the groups present similar distributions and characteristics. The means and standard deviations of the variances, denoted by $\bar{e}$ and $\sqrt{\bar{v}}$ and those of the covariances, denoted by e and $\sqrt{v}$ are listed in Table 5.1. These four groups of samples can mislead researcher in the camouflage of four different samples. But the mean and standard deviation of the variances and those of the covariance are fairly alike, therefore these four groups of samples can give rise to similar results in testing and comparing algorithms. ◀

4. Users have no control over the distributional characteristics of the variances and covariances. Having to utilize what they get, they may not be able to determine whether a feature in testing and comparing algorithms is caused by the algorithms or by the input from historical data.

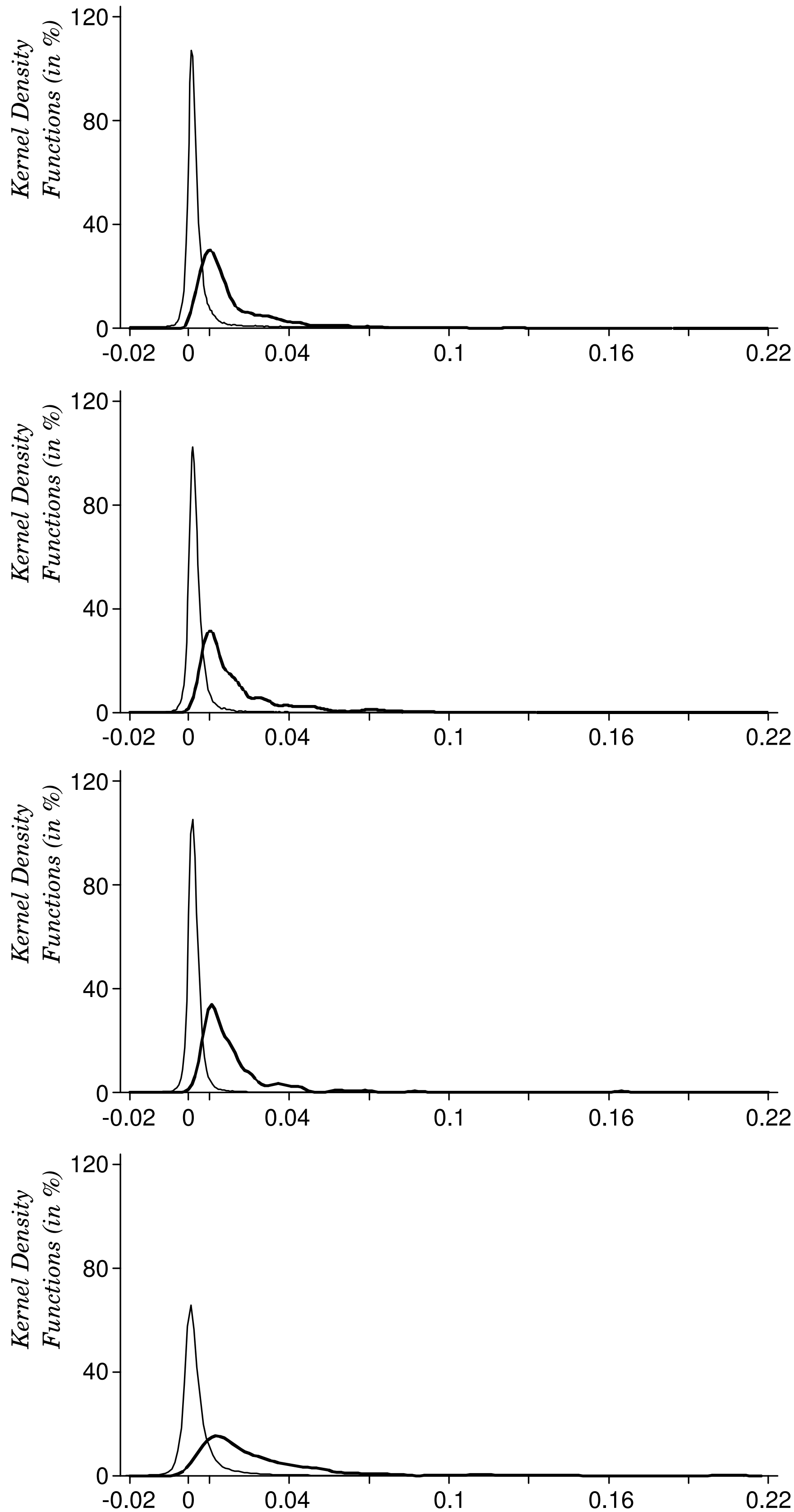

Figure 5.1: Distributions of the 250 variances (thick lines) and 31,125 covariances (thin lines) of the four random non-overlapping 250-stock groups from Standard & Poor's Super Composite 1500 Index of Example 1

5. Databases will typically have difficulty in providing data in large-scale computation. For example the 30 tests of 2000-security portfolio selection problems in Hirschberger, Qi and Steuer [40] will deplete a lot of databases.

It is difficult to do a literature review on this research since there is approximately no other research with the same purpose. Nevertheless, there has been a sprinkling of papers on the construction, generation, and use of random correlation matrices, mostly in other disciplines. These papers aim at the random generation of correlation matrices to match given sets of eigenvalues to meet application needs predominately in numerical statistical analysis (factor analysis, stepwise regression, principal components analysis, and so forth) and signal processing for example by Davies and Higham [18], Holmes [42] and Lin and Bendel [56].

Moving closer is the work of Pafka and Kondor [74], and Pafka, Potters and Kondor [75] in which they use results from the theory of random matrices to reduce noise in empirical covariance matrices. Chopra and Ziemba [15] randomly perturb the elements of a covariance matrix to test the effects of changes in the elements on optimal portfolio choice. However, they did not realize that this perturbation may knock the matrices away from being covariance matrices. Also relevant is the literature of matrix nearness problems (surveyed by Higham [36]). Motivated by a mutual fund situation, Higham [37] also proposes to compute from the cone of positive semidefinite matrices the matrix that is nearest to a matrix not in the cone. This is useful in portfolio selection optimization to correct the situation where analysts divide a large covariance matrix into submatrices, each analyst estimates one such submatrix, every submatrix is positive semidefinite, but the whole matrix is not. Perhaps it may be possible to combine Higham's paper and this chapter in some kind of complementary fashion in future research.

While all but Chopra and Ziemba [15] work on correlation matrices, this is equivalent to working on covariance matrices, because, when all standard deviations $\sigma_i = \sqrt{\sigma_{ii}}$ are nonzero, the associated correlation matrices and covariance matrices, C and $\mathbf{\Sigma}$, respectively, are simply re-scaled versions of one another in the following sense

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_n)\ C\ \text{diag}(\sigma_1, \ldots, \sigma_n)$$
$$C = \text{diag}(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_n})\ \mathbf{\Sigma}\ \text{diag}(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_n})$$

5.2 Characteristics of Historical Covariance Matrices

In this section more samples from the Standard & Poor's 500 Index, Nikkei 225 Index, and FTSE 100 Index are used to try to have a glimpse of distributional characteristics of historical covariance matrices.

Example 2. Select $1,078$ securities from the Standard & Poor's Super Composite 1500 Index from January 1997 to December 2002 (monthly returns). [3] Compute the $1,078$ variances and $\frac{1078\times(1078-1)}{2} = 580,503$ covariances of the returns of the selected securities and plot the distributions of the variances and covariances in Figure 5.2. The thick line and thin line denote the distributions of variances and covariances, respectively. Similar to Figure 5.1, the variances express much greater variation (dispersion) than the covariances and the mean of the variance is bigger than that of the covariances. ◀

Example 3. Select 471 securities from the Standard & Poor's 500 Index from January 1997 to December 2002 (monthly returns) [4] and apply the same method as Example 2. Analogous shapes of distributions show up in Figure 5.3. ◀

---

[3]Data source: the Center for Research in Security Prices (CRSP) through Wharton Research Data Service (WRDS) at the University of Pennsylvania and the website of the Index at http://www.spglobal.com on July 20, 2003. Because of corporation merges, acquisitions, dropouts from the index, or missing values, $1,078$ securities with full records are utilized instead of the $1,500$ constituents.

[4]Data source: the Center for Research in Security Prices (CRSP) through Wharton Research Data Service (WRDS) at the University of Pennsylvania and the website of the Index at http://www.spglobal.com on July 20, 2003. Because of corporation merges, acquisitions, dropouts from the index, or missing values, 471 securities with full records are utilized instead of the 500 constituents.

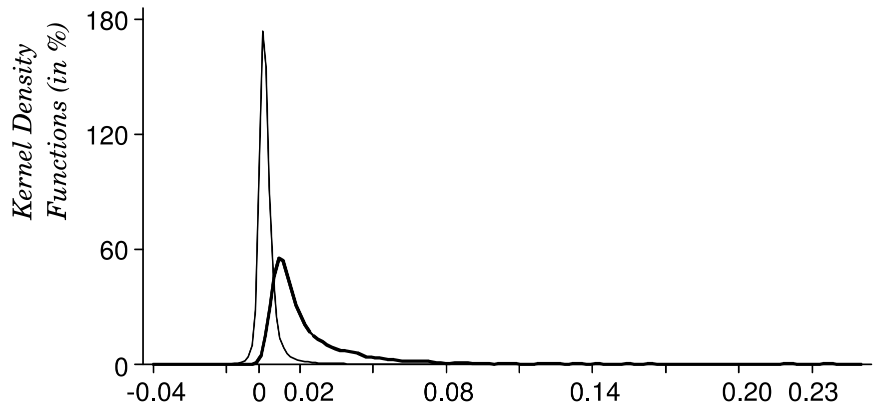

Figure 5.2: Distributions of the 1, 078 variances (thick line) and 580, 503 covariances (thin line) of the 1, 078 securities from the Standard & Poor's Super Composite 1500 Index of Example 2

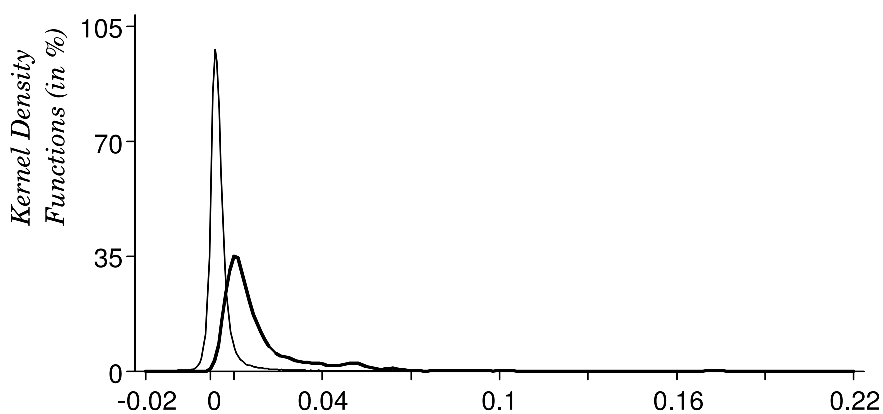

Figure 5.3: Distributions of the 471 variances (thick line) and 110, 685 covariances (thin line) of the 471 securities from the Standard & Poor's 500 Index of Example 3

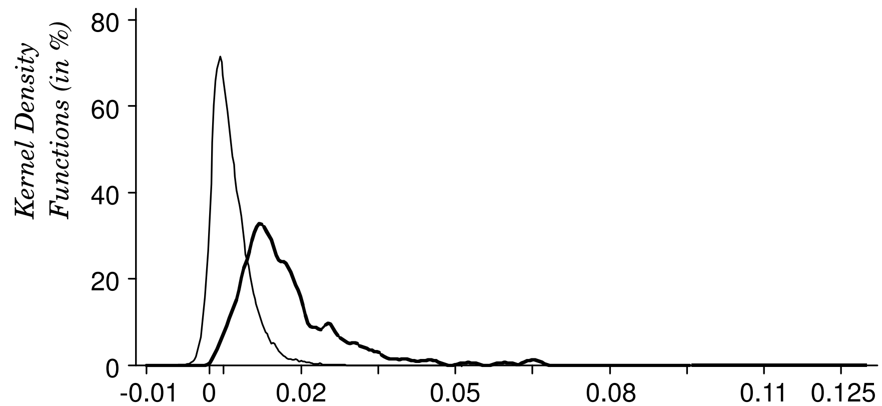

Figure 5.4: Distributions of the 214 variances (thick line) and 22, 791 covariances (thin line) of the 214 securities from the Nikkei 225 Index of Example 4

Example 4. Select 214 securities from the Nikkei 225 Index from January 1997 to December 2002 (monthly returns) [5] and apply the same method as Example 2. Analogous shapes of distributions show up in Figure 5.4. ◀

Example 5. Select 85 securities from the FTSE 100 Index from January 1997 to December 2002 (monthly returns) [6] and apply the same method as Example 2. Analogous shapes of distributions show up in Figure 5.5. ◀

The statistics of the variances and the covariances of Example 2-Example 5 are listed in Table 5.2. All the samples produce relatively similar shapes of distributions. However, it is realized that the covered empirical data is very limited. Beneath this tip of iceberg, further sampling by incorporating more securities, adding more time periods, and using weekly or daily observations may unearth a much more comprehensive and interesting picture.

[5]Data source: DataStream of Thomson Financial, September 6, 2003. Because of corporation merges, acquisitions, dropouts from the index, or missing values, 214 constituents with full records are utilized instead of the 225 constituents.

[6]Data source: DataStream of Thomson Financial, September 6, 2003. Because of corporation merges, acquisitions, dropouts of the index, or missing values, 85 constituents with full records are utilized instead of the 100 constituents.

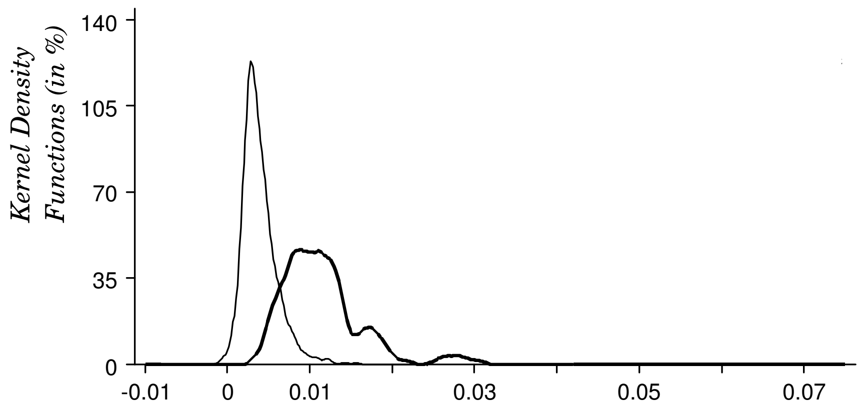

Figure 5.5: Distributions of the 85 variances (thick line) and 3, 570 covariances (thin line) of the 85 securities from the FTSE 100 Index of Example 5

Table 5.2: The means and standard deviations of the variances and covariances of Examples 2 to 5

		Example 2	Example 3	Example 4	Example 5
variances	mean, $\bar{e}$	.020317	.017049	.015888	.010151
	standard deviation, $\sqrt{\bar{v}}$	.021210	.016482	.010350	.005083
covariances	mean, e	.002914	.002962	.004373	.002225
	standard deviation, $\sqrt{v}$	.003976	.003764	.003752	.002219

5.3 Experiments in Trying to Invent Covariance Matrices

This section demonstrates the difficulty investors might encounter in trying to invent (simulate) a covariance matrix without the aid of a procedure such as by Hirschberger, Qi and Steuer [41]. To illustrate the effectiveness of this invention, the results of eighty experiments are reported. In the experiments, the goal is to invent a 50×50 covariance matrix and a 100×100 covariance matrix by simply randomly selecting elements from reasonable distributions to populate the matrix.

5.3.1 By Uniform Distribution and Normal Distribution

In each trial of the experiments corresponding to Examples 2 to 5, a symmetric 50×50 matrix is populated with diagonal elements randomly drawn from one distribution and off-diagonal elements randomly drawn from another with the means of these two distributions closely matching the $\bar{e}$ and e of Table 5.2, respectively. Since such matrices are positive semidefinite if and only if all the eigenvalues are nonnegative, one can then compute all the eigenvalues of all principal minors starting at 1×1 and going to 50×50 [7] . If $(p+1) \times (p+1)$ is the first principal minor to have negative eigenvalues, then one knows that the $p \times p$ principal minor is the largest covariance matrix that can be salvaged from that trial. Utilizing values in a neighborhood about the values for e, $\sqrt{v}$, $\bar{e}$, $\sqrt{\bar{v}}$ shown in Table 5.2, with a sample size of 40 for each experiment, the results of twenty experiments are shown in each of columns (a), (b), (c) and (d) of Table 5.3. Columns (a), (b), (c) and (d) correspond to Example 2, 3, 4 and 5, respectively, with five experiments in each column.

To explain, consider column (a) corresponding to Example 2. For the five experiments of this column, all diagonal elements were randomly drawn from the same uniform distribution $U[0, .040634]$ with lower bound 0 and upper bound .040634 as indicated, while the off-diagonal elements were randomly drawn from the normal distribution $N(.002914, \sqrt{v})$

[7] A principal minor starting at 1×1 and going to $p \times p$ of a $n \times n$ matrix $\mathbf{\Sigma}$ is formed by keeping the 1st, 2nd, 3rd, ... pth rows and columns of $\mathbf{\Sigma}$ and deleting the other rows and columns of $\mathbf{\Sigma}$.

Table 5.3: Results of twenty experiments (40 trials in each experiment) to invent a 50×50 covariance matrix

	(a)		(b)		(c)		(d)	
Diag	$U[0, .040634]$		$U[0, .034098]$		$U[0, .031776]$		$U[0, .020302]$	
Off-diag	$N(.002914, \sqrt{v})$		$N(.002962, \sqrt{v})$		$N(.004373, \sqrt{v})$		$N(.002225, \sqrt{v})$	
$\sqrt{v}$	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}
.0042	4.93	12	4.78	8	3.30	6	2.52	5
.0036	4.80	11	4.68	10	3.70	9	3.40	7
.0030	5.45	15	4.78	12	4.40	9	3.78	8
.0024	6.45	16	6.03	17	5.08	11	3.83	7
.0018	9.23	25	6.48	22	5.05	16	4.98	11

with mean of .002914 and standard deviations $\sqrt{v}$ of .0042, .0036, .0030, .0024 and .0018 for five experiments as indicated, so that the means of $U[0, .040634]$ and $N(.002914, \sqrt{v})$ equal $\bar{e} = .020317$ and $e = .002914$ in Table 5.2, respectively. The "4.93" and "12" of the first experiment mean the following. In the 40 attempts to generate a 50×50 covariance matrix of this experiment, the largest covariance matrix one was able to generate was a 12×12, with an average largest covariance matrix size of 4.93 over the 40 attempts.

Note that over the ranges of values refined from historical data, one was unable to generate even one 50×50 covariance matrix over the 800 trials of the 20 experiments. The largest covariance matrix that one was able to generate was a 25×25 in one of the trials of the last experiment in column (a). The next largest was a 22×22 in one of the trials of the last experiment in column (b), and so forth. However, it is observable that as $\sqrt{v}$ decreases, the ability to invent a covariance matrix appears to improve. But the side effect of decreasing $\sqrt{v}$ much below the lowest value of .0018 shown is to leave the realm of covariances matrices that are realistic in portfolio selection.

Table 5.4: Results of twenty experiments (40 trials in each experiment) to invent a 100×100 covariance matrix

	(a)		(b)		(c)		(d)	
Diag	$U[0, .040634]$		$U[0, .034098]$		$U[0, .031776]$		$U[0, .020302]$	
Off-diag	$N(.002914, \sqrt{v})$		$N(.002962, \sqrt{v})$		$N(.004373, \sqrt{v})$		$N(.002225, \sqrt{v})$	
$\sqrt{v}$	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}
.0042	5.83	12	4.30	9	3.60	6	3.18	7
.0036	6.32	13	4.05	8	3.58	9	2.63	5
.0030	6.48	15	4.08	9	4.45	11	3.78	8
.0024	7.80	18	5.08	12	4.80	13	4.18	8
.0018	8.93	25	6.08	20	4.90	14	5.18	13

This result is confirmed by another 20 experiments to invent a 100×100 covariance matrix with similar designs in Table 5.4. Quite analogous are these two results, which can be explained by that the failure (non-positive-semidefiniteness) is caused by misalignments of the elements and the first misalignments happen approximately in the same interval of dimensions.

5.3.2 By Triangular Distribution and Normal Distribution

An attempt to fine-tune the experiments is to employ for the variances triangular distributions which may fit better than the uniform distributions. The basic preparations are as follows.

1. Construct a triangular distribution so that its mode matches that of the distribution of the variances as in Figure 5.6. The thin line and thick line represent the distribution of variances and the triangular distribution. Let m, h and l denote the mode, height and upper bound of the triangular distribution whose lower bound is 0. Since the area under the thick line is 1 i.e. $\frac{h \times l}{2} = 1$, $l = 2/h$.

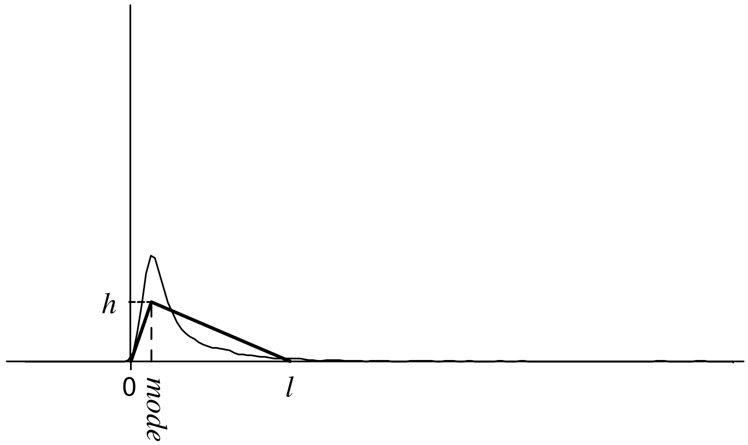

Figure 5.6: Triangular distribution (thick line) to approximate the distribution of the variances (thin line)

2. The density function is

$$p(x) = \begin{cases} \frac{h}{m}x & x \in [0, m] \\ \frac{h}{m-l}(x-l) & x \in (m, l] \\ 0 & \text{otherwise} \end{cases}$$

3. The distribution function is

$$F(x) = prob\{\text{the triangular distribution} \leq x\} = \int_{-\infty}^{x} p(t)\, dt$$

a)

$$F(x) = 0, \text{ if } x < 0$$

b)

$$\begin{aligned} F(x) &= \int_0^x \frac{h}{m} t\, dt, \text{ if } x \in [0, m] \\ &= \frac{h}{2m} t^2 \mid_0^x \\ &= \frac{h}{2m} x^2 \end{aligned}$$

c)

$$
\begin{aligned}
F(x) &= \int_0^x p(t)\,dt, \text{ if } x \in (m, l] \\
&= \int_0^m p(t)\,dt + \int_m^x p(t)\,dt \\
&= \frac{h}{2m} m^2 + \int_m^x \frac{h}{m-l}(t-l)\,dt \\
&= \frac{hm}{2} + \frac{h}{2(m-l)}(t-l)^2 \,|_m^x \\
&= \frac{hm}{2} + \frac{h}{2(m-l)}(x^2 - 2lx + (2lm - m^2)) \\
&= \frac{h}{2(m-l)}x^2 - \frac{hl}{m-l}x + \frac{h(2lm - m^2)}{2(m-l)} + \frac{hm}{2}, \quad \text{since } hl = 2 \\
&= \frac{h}{2(m-l)}x^2 - \frac{2}{m-l}x + \frac{2hlm - hm^2}{2(m-l)} + \frac{hm}{2}, \quad \text{since } hl = 2 \\
&= \frac{h}{2(m-l)}x^2 - \frac{2}{m-l}x + \frac{4m - hm^2}{2(m-l)} + \frac{hm^2 - hlm}{2(m-l)}, \quad \text{since } hl = 2 \\
&= \frac{h}{2(m-l)}x^2 - \frac{2}{m-l}x + \frac{m}{(m-l)} \\
&= \frac{1}{2(m-l)}(hx^2 - 4x + 2m)
\end{aligned}
$$

d)

$$F(x) = 1, \text{ if } x > l$$

4. The distribution function $F(x)$ is strictly increasing and piecewise quadratic on $x \in [0, l]$. The inverse function F^{-1} exists and can be constructed as follows

a)

$$
\begin{aligned}
y = F(x) &= \frac{h}{2m}x^2, \quad \text{if } x \in [0, m] \\
F^{-1}(y) &= \sqrt{\frac{2my}{h}}, \quad \text{if } y \in [0, \frac{hm}{2}]
\end{aligned}
$$

b)

$$y = F(x) = \frac{1}{2(m-l)}(hx^2 - 4x + 2m), \quad \text{if } x \in (m, 1] \text{ , re-arrange to obtain}$$

$$hx^2 - 4x + 2(m + (l - m)y) = 0$$

then [8]

$$x = \frac{4 \pm \sqrt{16 - 8h(m + (l - m)y)}}{2h}$$
$$= \frac{2 \pm \sqrt{4 - 2h(m + (l - m)y)}}{h}$$

then [9]

$$F^{-1}(y) = \frac{2 - \sqrt{4 - 2h(m + (l - m)y)}}{h}, \quad \text{if } y \in (\frac{hm}{2}, 1]$$

5. The triangular distribution can be obtained from $F^{-1}(U[0,1])$ where $U[0,1]$ is a uniform distribution with lower bound 0 and upper bound 1.

Then the experiments to invent 50×50 and 100×100 covariance matrices are proceeded. The results are in Table 5.5 and Table 5.6 where for example $\triangle(.009600, 30)$ means a triangular distribution with model .009600 ($m = .009600$) and height 30 ($h = 30$). Although the results are better than those in Table 5.3 and Table 5.4, all experiments except one in Table 5.5 fail to generate a 50×50 covariance matrix, and all experiments in Table 5.6 fail to generate a 100×100 covariance matrix.

These experiments illustrate that this inventing method can produce covariance matrices up to size 20×20, but the success rate quickly declines toward zero as n gets larger. For instance, a single 50×50 covariance matrix was not able to be generated in the 800 tries, and a single 100×100 covariance matrix was not able to be generated in the 1600 tries. This is serious because covariance matrices in portfolio selection are often not small. Sizes of $1,000 \times 1,000$ are not uncommon.

---

[8]

$$\begin{aligned}\text{The discriminant}\Delta &= 16 - 8h(m + (l - m)y)\\ &\geq 16 - 8h(m + (l - m)), \quad \text{since } 0 \leq y \leq 1\\ &= 16 - 8hl, \quad \text{since } hl = 2\\ &= 0\end{aligned}$$

[9] One root $x = \frac{2+\sqrt{4-2h(m+(l-m)y)}}{h} \geq \frac{2}{h} = l$ is discarded because $x \in (m, l]$.

Table 5.5: Results of twenty experiments (40 trials in each experiment) to invent a 50×50 covariance matrix

	(a)		(b)		(c)		(d)	
Diag	$\triangle(.009600, 30)$		$\triangle(.010500, 21)$		$\triangle(.011900, 20)$		$\triangle(.010000, 35)$	
Off-diag	$N(.002914, \sqrt{v})$		$N(.002962, \sqrt{v})$		$N(.004373, \sqrt{v})$		$N(.002225, \sqrt{v})$	
$\sqrt{v}$	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}
.0042	10.13	16	11.60	20	13.88	22	8.80	16
.0036	11.78	18	15.03	25	16.73	26	10.55	19
.0030	14.20	23	18.68	36	21.78	38	14.13	23
.0024	18.65	27	25.13	39	28.33	48	18.35	30
.0018	30.53	47	37.65	49	43.40	50	29.30	43

Table 5.6: Results of twenty experiments (40 trials in each experiment) to invent a 100×100 covariance matrix

	(a)		(b)		(c)		(d)	
Diag	$\triangle(.009600, 30)$		$\triangle(.010500, 21)$		$\triangle(.011900, 20)$		$\triangle(.010000, 35)$	
Off-diag	$N(.002914, \sqrt{v})$		$N(.002962, \sqrt{v})$		$N(.004373, \sqrt{v})$		$N(.002225, \sqrt{v})$	
$\sqrt{v}$	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}	p^{ave}	p^{max}
.0042	9.53	18	13.75	22	15.20	25	9.95	14
.0036	11.03	21	15.98	27	18.38	31	10.93	19
.0030	13.93	24	18.28	31	23.43	37	13.25	21
.0024	19.50	31	25.78	40	30.08	46	18.00	29
.0018	27.23	41	38.00	51	44.53	65	26.95	41

5.4 Random Covariance Matrix Generation

Cholesky decomposition (see for example Lyuu [59]) is employed in financial research to simulate return vectors $\mathbf{r} \in \mathbb{R}^n$ from a given covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ as follows.

1. Matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ is positive semidefinite if and only if there exists a *root matrix* $\mathbf{L} \in \mathbb{R}^{n \times q}$, such that

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T \tag{5.1}$$

2. Matrix $\mathbf{L}$ can be interpreted as a matrix of factor loadings for some vector of (uncorrelated) risk factors $\mathbf{y} \in \mathbb{R}^q$ as follows

$$\mathbf{r} = \boldsymbol{\mu} + \mathbf{L}\mathbf{y} \tag{5.2}$$

By successively drawing $\mathbf{y}$-vectors from a standard multivariate normal distribution, (5.2) can be used for the random generation of return vectors $\mathbf{r}$ with expected value $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$.

Although the $\mathbf{r}$-vectors can be re-used to calculate a covariance matrix $\hat{\mathbf{\Sigma}}$. $\hat{\mathbf{\Sigma}}$ is merely an estimate of the original $\mathbf{\Sigma}$, more close for more $\mathbf{r}$-vectors. Quite differently, the purpose of the procedure involved is to randomly generate a series of covariance matrices. Although such matrices can be different from one another, they have approximately the same distributional characteristics. Nevertheless, the procedure also starts from (5.1) in the sense that $n \times n$ matrix $\mathbf{F}\mathbf{F}^T$ is positive semidefinite for any matrix $\mathbf{F} \in \mathbb{R}^{n \times m}$ where $m \geq 1$. This means that a random covariance matrix may be obtained by creating a random matrix $\mathbf{F}$ first. The strategy is to develop a method for randomly generating the f_{ij} elements of an $\mathbf{F}$ matrix from independent and identical distributions such that the resulting positive semidefinite matrix $\mathbf{F}\mathbf{F}^T$ has the diagonal and off-diagonal element distributional characteristics desired.

For a random variable X, let $E[X]$ denote expected value, and let variance, standard deviation, skewness, and kurtosis be defined and denoted as follows

$$\begin{aligned} V[X] &= E[(X - E[X])^2] \\ \text{Std}[X] &= V[X]^{1/2} \\ \text{Sk}[X] &= E[(X - E[X])^3] \\ K[X] &= E[(X - E[X])^4] \end{aligned}$$

Four (sometimes five) quantities as inputs— $\bar{e}$ and $\sqrt{\bar{v}}$ denoting the expected value and standard deviation of the diagonal elements (variances) and e and $\sqrt{v}$ denoting those of the off-diagonal elements (covariances) are used as inputs.

$$\begin{aligned} &\bar{e} = E[\sigma_{ii}] \qquad \sqrt{\bar{v}} = \text{Std}[\sigma_{ii}] \qquad \text{for all } i \\ &e = E[\sigma_{ij}] \qquad \sqrt{v} = \text{Std}[\sigma_{ij}] \qquad \text{over all pairs } i, j, \quad i \neq j \end{aligned}$$

A fifth input s, reflecting the skewness of the off-diagonal covariances, where

$$s = \text{Sk}[\sigma_{ij}] \qquad \text{over all pairs } i, j, \quad i \neq j$$

is sometimes required.

Based on $\mathbf{\Sigma} = \mathbf{F}\mathbf{F}^T$, the goal is to utilize the above inputs to construct the f_{ij} elements of an $\mathbf{F}$ in such a way that matrix multiplication $\mathbf{F}\mathbf{F}^T$ results in a $\mathbf{\Sigma}$ possessing the desired diagonal and off-diagonal distributional characteristics by computing $\hat{e}, \hat{v}, \hat{s}$, $\hat{k}$ and m values where

$$\begin{gathered} \hat{e} = E[f_{ij}] \quad \hat{v} = V[f_{ij}] \quad \hat{s} = \text{Sk}[f_{ij}] \quad \hat{k} = K[f_{ij}] \\ \text{over all pairs } i, j, \quad i = 1, \ldots, n \quad j = 1, \ldots, m \end{gathered}$$

To develop the $\hat{e}, \hat{v}$ and $\hat{s}$ moments of the distribution from which the f_{ij} are drawn so as to generate the off-diagonal elements of a covariance matrix that have expected value e and standard deviation $\sqrt{v}$, there is Theorem 5.1.

Theorem 5.1. *Let the* f_{ij}*,* $i = 1, \ldots, n$*,* $j = 1, \ldots, m$*, be independent random variables, identically distributed by some probability distribution* $\mathfrak{F}$*. Then the off-diagonal elements (covariance)* σ_{ij}*,* $i, j = 1, \ldots, n$*,* $i \neq j$*, in* $\mathbf{FF}^T$ *have expected value* $e \geq 0$*, variance* $v \geq 0$ *and skewness* $s > 0$ *if and only if*

$$\hat{e} = \sqrt{\frac{e}{m}} \tag{5.3}$$

$$\hat{v} = -\hat{e}^2 + \sqrt{\hat{e}^4 + \frac{v}{m}} \tag{5.4}$$

$$\hat{s} = -\hat{e}^3 - 3\hat{e}\hat{v} + \sqrt{3\hat{e}^2\hat{v}^2 + 6\hat{e}^4\hat{v} + \hat{e}^6 + \frac{s}{m}} \tag{5.5}$$

where formula (5.5) *is valid if and only if*

$$s \geq -\frac{1}{m}\left[\frac{e^3}{m} + 3ve\right] \tag{5.6}$$

The proof and other proofs in this chapter are in Hirschberger, Qi and Steuer [41]. To determine m, the number of columns in $\mathbf{F}$, and the fourth central moment $\hat{k}$ of the distribution from which the f_{ij} are to be drawn so that the diagonal elements of a covariance matrix have expected value $\bar{e}$ and standard deviation $\sqrt{\bar{v}}$, there is Theorem 5.2.

Theorem 5.2. *Under the assumptions of Theorem 5.1, the diagonal elements (variances)* σ_{ii}*,* $i = 1, \ldots, n$*, in the* $\mathbf{FF}^T$ *have expected value* $\bar{e} \geq 0$ *and variance* $\bar{v} \geq 0$ *if and only if*

$$\bar{e} = m(\hat{v} + \hat{e}^2) \tag{5.7}$$

$$\bar{v} = m(\hat{k} + 4\hat{s}\hat{e} - \hat{v}^2 + 4\hat{v}\hat{e}^2) \tag{5.8}$$

In terms of original inputs, (5.7) can be expressed as

$$m = \frac{\bar{e}^2 - e^2}{v} \tag{5.9}$$

This means that m can be approximately obtained by rounding $\frac{\bar{e}^2 - e^2}{v}$ to the nearest positive integer. Conditions $\bar{e} \geq e$ and thus $m \geq 0$ are supported by the fact that the mean of the variances is greater than or equal to that of the covariances. Reexpressing (5.8)

$$\hat{k} = \frac{\bar{v}}{m} + \hat{v}^2 - 4\hat{s}\hat{e} - 4\hat{v}\hat{e}^2 \tag{5.10}$$

Distributions from the family of Johnson [48] distributions are a good candidate for the distribution $\mathfrak{F}$ (f_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, are independent random variables, identically distributed by $\mathfrak{F}$). Because there are algorithms to determine the family of Johnson distributions by the first four moments, and closed-form formulae or at least good approximations for the corresponding inverse distribution function. However, lognormal distributions, members of the family of Johnson distributions are preferred, because most cases can be covered by lognormal distributions and lognormal distributions are much easier to work with than the family of Johnson distributions .

The procedure of generating covariance matrices is organized as follows.

1. Compute m by (5.9).

2. Compute $\hat{e}$ and $\hat{v}$ by (5.3) and (5.4).

3. Two conditions must be met to solve for $\hat{s}$ and $\hat{k}$. One condition is that $\hat{k}$ and $\hat{s}$ are related through τ as described in Johnson [48] and Hill, Hill and Holder [38] as follows

$$\frac{\hat{k}(\tau)}{\hat{v}^2} = (\tau^4 + 2\tau^3 + 3\tau^2 - 3) \qquad \frac{\hat{s}(\tau)}{\hat{v}^{3/2}} = \pm\sqrt{(\tau-1)(\tau+2)^2}$$

for $\tau \geq 1$. Parameterizing τ from $1 \to +\infty$, points

$$\Big(\frac{\hat{k}(\tau)}{\hat{v}^2}, \frac{\hat{s}(\tau)}{\hat{v}^{3/2}}\Big) \tag{5.11}$$

form a bullet-shaped curve. The other condition is (5.10). Then $\hat{s}$ and $\hat{k}$ can be located at the intersection points between (5.11) and (5.10). However, if they do not intersect, which can be rare and has not happened yet in the experiments, the family of Johnson distributions can be utilized.

4. Compute the parameters of the lognormal distribution as described by Johnson [48] and Hill, Hill and Holder [38] as follows

$$\delta = (\ln\tau)^{-\frac{1}{2}} \qquad \gamma = \tfrac{1}{2}\delta\ln[\tau(\tau-1)/\hat{v}]$$

$$\lambda = \operatorname{sgn}(\hat{s}) \qquad \xi = \lambda\hat{e} - \exp\big[\,(\tfrac{1}{2\delta} - \gamma)/\delta\,\big]$$

5. The lognormally distributed variable X may be produced from a standard normally distributed variable Z using the transformation

$$X = \xi + \lambda \exp\left[\frac{Z-\gamma}{\delta}\right]$$

6. Obtain the $\mathbf{F}$ matrix by

a) Generate random numbers u_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, from $U[0, 1]$.

b) Let $q_{ij} = Q(u_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, where Q is an approximation of the inverse distribution function of $N(0, 1)$.

c) Let $f_{ij} = \xi + \lambda \exp[(q_{ij} - \gamma)/\delta]$, $i = 1, \ldots, n$, $j = 1, \ldots, m$.

7. Compute $\mathbf{\Sigma} = \mathbf{F}\mathbf{F}^T$.

5.5 Illustrations and Generating Expected Returns

To illustrate, use $\bar{e}, \bar{v}, e$ and v in Table 5.2 as inputs and start the procedure. Because $\bar{e}, \bar{v}, e$ and v are computed from empirical data, the elements of the randomly generated covariance matrices should have distributional characteristics closely matching $\bar{e}, \bar{v}, e$ and v. Take Example 2 with the $n = 1,078$ securities from the Standard & Poor's Super Composite 1500 Index, get the column of $\bar{e}, \bar{v}, e$ and v of Example 2 in Table 5.2, and start the generation. The distributions of the randomly generated variances (thin line) and the historical variances (thick line) are in Figure 5.7 *top*. One can detect a slight difference. The distributions of randomly generated covariances (thin line) and the historical covariances (thick line) are in Figure 5.7 *bottom*. It is hard to observe any difference because the two distributions fall approximately on top of each other. Graphs in Figure 5.7 are typical of the results one can expect to obtain.

Likewise, get the columns of $\bar{e}, \bar{v}, e$ and v of Example 3 in Table 5.2 and start the generation. The distributions of randomly generated variances and covariances (thin line), and the historical variances and covariances (thick line) are in Figure 5.8. Get the columns of $\bar{e}, \bar{v}, e$ and v of Example 4 in Table 5.2 and start the generation. The distributions of randomly

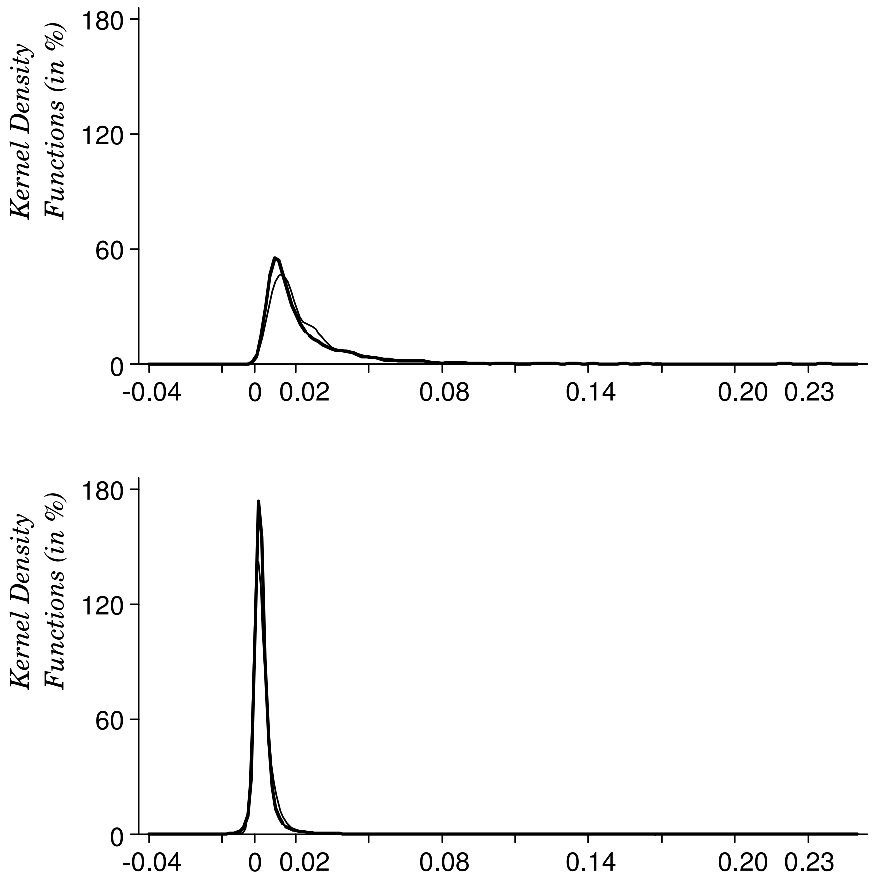

Figure 5.7: Distributions of randomly generated variances (thin line) and historical variances (thick line) in *top* graph, and distributions of randomly generated covariances (thin line) and historical covariances (thick line) in *bottom* graph, based on the $1,078$ securities from the Standard & Poor's Super Composite 1500 Index of Example 2

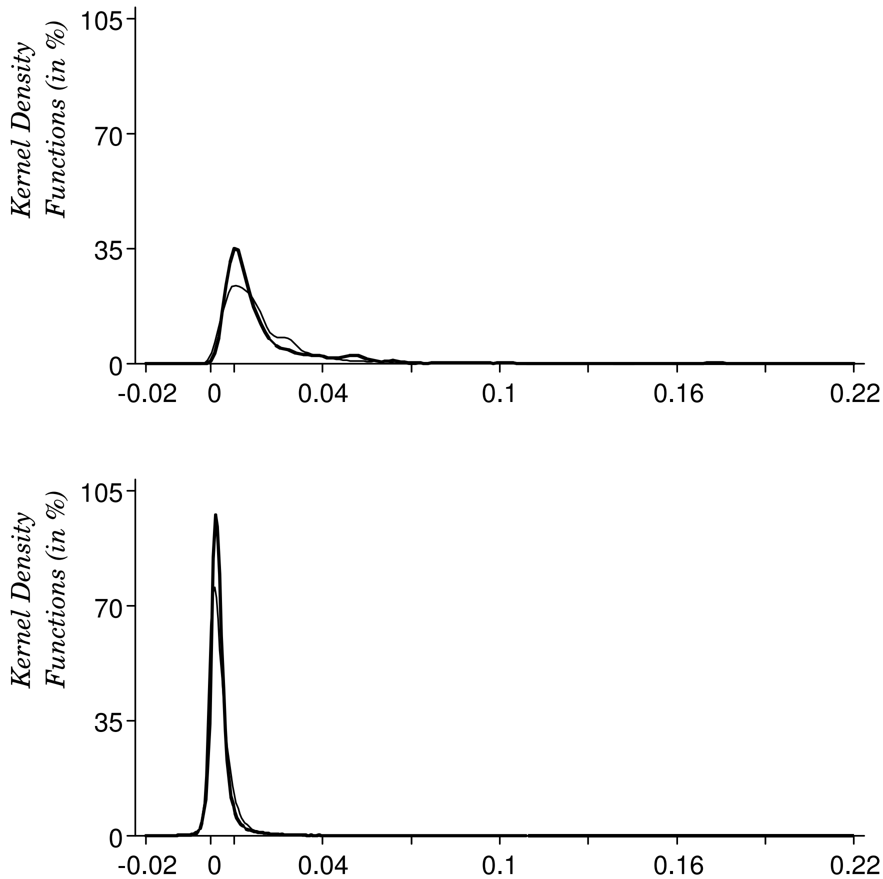

Figure 5.8: Distributions of randomly generated variances (thin line) and historical variances (thick line) in *top* graph, and distributions of randomly generated covariances (thin line) and historical covariances (thick line) in *bottom* graph, based on the 471 securities from the Standard & Poor's 500 Index of Example 3

generated variances and covariances (thin line), and the historical variances and covariances (thick line) are in Figure 5.9. Get the columns of $\bar{e}, \bar{v}, e$ and v of Example 5 in Table 5.2 and start the generation. The distributions of randomly generated variances and covariances (thin line), and the historical variances and covariances (thick line) are in Figure 5.10. The shapes of distributions of these three figures are similar to those of Figure 5.7.

The means and standard deviations ($\bar{e}^r$, $\sqrt{\bar{v}}^r$, e^r and $\sqrt{v}^r$) of randomly generated variances and covariances of Examples 2 to 5 are listed in Table 5.7. The means and standard deviations are similar to their counterparts in Table 5.2.

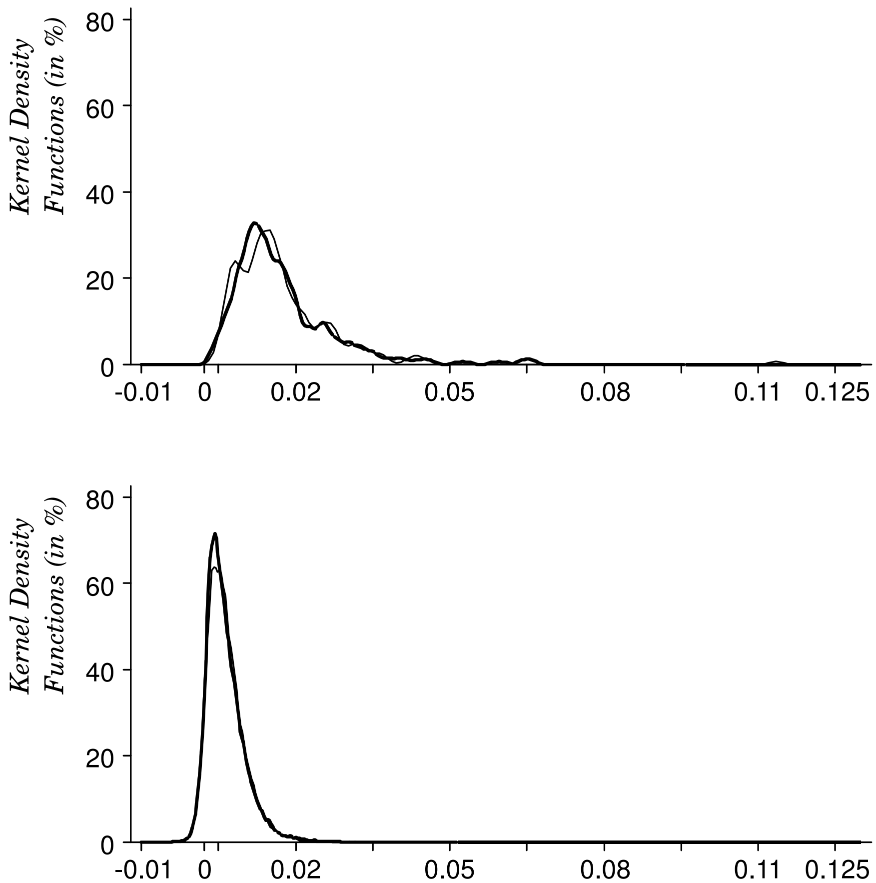

Figure 5.9: Distributions of randomly generated variances (thin line) and historical variances (thick line) in *top* graph, and distributions of randomly generated covariances (thin line) and historical covariances (thick line) in *bottom* graph, based on the 214 securities from the Nikkei 225 Index of Example 4

Table 5.7: Means and standard deviations of randomly generated variances and covariances of Examples 2 to 5

		Example 2	Example 3	Example 4	Example 5
variances	mean, $\bar{e}^r$	.021614	.018453	.015619	.010984
	standard deviation, $\sqrt{\bar{v}}^r$	.019067	.018249	.010875	.006317
covariances	mean, e^r	.003154	.003099	.004411	.002468
	standard deviation, $\sqrt{v}^r$	.004190	.003975	.003642	.002508

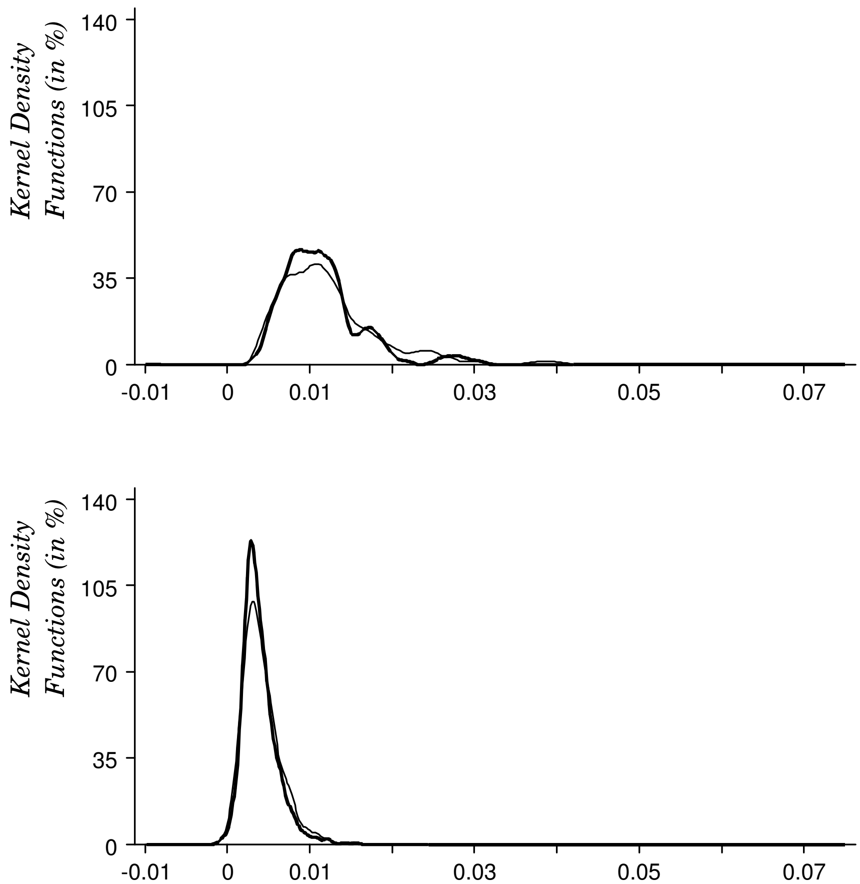

Figure 5.10: Distributions of randomly generated variances (thin line) and historical variances (thick line) in *top* graph, and distributions of randomly generated covariances (thin line) and historical covariances (thick line) in *bottom* graph, based on the 85 securities from the FTSE 100 Index of Example 5

To make the procedure complete, an ability to generate expected returns is offered. This ability is based on the general principle of high return with high risk and low return with low risk. Take the square root of each σ_{jj} generated covariance matrix diagonal element. Let μ_s and σ_s denote the mean and standard deviation of the $\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \ldots, \sqrt{\sigma_{nn}}$. Compute

$$t_{s_j} = \frac{\sqrt{\sigma_{jj}} - \mu_s}{\sigma_s}$$

Then compute the expected return for security j as

$$\mu_{e_j} = \mu_e + t_{s_j} \rho \, \sigma_e \tag{5.12}$$

where μ_e and σ_e are the mean and standard deviation of the securities' of expected returns, respectively and ρ is the correlation coefficient between generated standard deviations and generated expected returns. It is acknowledged that (5.12) is very primitive and generating expected returns is one of the most difficult challenges of portfolio selection.

The author believes that it may be possible to take the equal portfolio weights $(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ as the market portfolio and apply Capital Asset Pricing Model which gives the relationship between the risk of an asset and its expected return, in order to enable this generating expected returns to be more consistent with popular pricing theories.

Chapter 6

Arguments and Formulation for Multiple Criteria Portfolio Selection

This chapter provides the theoretical arguments for multiple criteria portfolio selection by introducing multiple criteria stochastic programming and extended utility function, and practical considerations of the multiple criteria. It also points out the difficulty of standard portfolio selection theory in treating these criteria as constraints. It reminds investors the imprecise definition of nondominated frontiers as well. The modelling of these criteria is proposed.

6.1 Arguments for Multiple Criteria Portfolio Selection

A single criterion stochastic programming model is described in Section 3.1 as follows

$$\begin{aligned} &\max \{ f(\mathbf{x}, \mathbf{r}) = z \} \\ &\text{s.t.} \qquad \mathbf{x} \in S \end{aligned}$$

where $\mathbf{x}$ is the decision vector, S is the feasible region, and $\mathbf{r}$ is a random vector to model uncertainty. The objective function is f. Note that $f(\mathbf{x}, \mathbf{r})$ is still a random variable. Based on this model, Stancu-Minasian [87] and Prékopa [79] proposed multiple criteria stochastic programming model as follows

$$\begin{aligned} &\max \{ f_1(\mathbf{x}, \mathbf{r}) = z_1 \} \\ &\max \{ f_2(\mathbf{x}, \mathbf{r}) = z_2 \} \\ &\qquad \vdots \\ &\max \{ f_k(\mathbf{x}, \mathbf{r}) = z_k \} \\ &\text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{6.1}$$

Substitute problems are needed to operationalize (6.1). Caballero, Cerdá, Munõz, Rey, and Stancu-Minasian [13] proposed and compared several substitute problems of (6.1). One of these substitute problems is

$$\begin{aligned}
\min \{V[f_1(\mathbf{x},\mathbf{r})] &= z_1\} \\
\max \{E[f_1(\mathbf{x},\mathbf{r})] &= z_2\} \\
\min \{V[f_2(\mathbf{x},\mathbf{r})] &= z_3\} \\
\max \{E[f_2(\mathbf{x},\mathbf{r})] &= z_4\} \\
&\vdots \\
\min \{V[f_k(\mathbf{x},\mathbf{r})] &= z_{2k-1}\} \\
\max \{E[f_k(\mathbf{x},\mathbf{r})] &= z_{2k}\} \\
\text{s.t.} \qquad \mathbf{x} &\in S
\end{aligned} \tag{6.2}$$

Model (6.2) is closely related to this dissertation. As argued in the introduction in Section 1.1, investors may have more criteria than portfolio return as put forward by Markowitz [65] as

> "This single period utility function may depend on portfolio return and perhaps other state variables ... it seems to me that the theory of rational behavior under uncertainty can continue to provide insights as to which practicable procedures provide near optimum results. In particular, it can further help evaluate the adequacy of mean and variance, or alternate practical measures, as criteria" (1991, p. 471 & 476).

and by Fisher and Statman [30] as

> "Mean-variance investors care only about risk and expected returns, but behavioral investors care about more than that ... A central feature of behavioral portfolio theory is that investors view their portfolios not as a whole, as prescribed by Markowitz, but as distinct layers in a pyramid of assets, where layers

are associated with particular goals, and where each layer has its own particular risks" (1997, p. 13).

Dividend yield, social responsibility and R&D can be examples of these criteria. Moreover, such three criteria are stochastic. That is, a portfolio's dividend yield, social responsibility and R&D are still unknown, even after the portfolio's weights $\mathbf{x}$ are determined. Therefore at the very beginning, suitable-portfolio investors bear in their mind a multiple criteria stochastic programming portfolio selection model for example as follows

$$\begin{aligned} &\max \{R(\mathbf{x}, \mathbf{r}) = \text{return}\} \\ &\max \{f_2(\mathbf{x}, \boldsymbol{\delta}) = \text{dividend yield}\} \\ &\max \{f_3(\mathbf{x}, \mathbf{s}) = \text{social responsibility}\} \\ &\max \{f_4(\mathbf{x}, \mathbf{h}) = \text{R\&D}\} \\ &\quad \text{s.t.} \quad \mathbf{x} \in S \end{aligned} \tag{6.3}$$

where the random vectors $\boldsymbol{\delta}, \mathbf{s}$, and $\mathbf{h}$ represent the uncertainty of dividend yield, social responsibility, and R&D, respectively. To avoid double-counting dividends, return can be computed from capital gains only. Taking (6.3) only as a guidance, a suitable-portfolio investor can customize his or her model by adding or removing criteria.

However, simply applying the substitute problem (6.2) of (6.3) will result in an eight objective model which can be difficult to solve for the following reasons. (a) A suitable-portfolio investor can not find an algorithm for computing the nondominated surface of four quadratic objectives plus four linear objectives. (b) A suitable-portfolio investor will have to estimate the expected value and covariance matrix of each of $\mathbf{r}, \boldsymbol{\delta}, \mathbf{s}$, and $\mathbf{h}$, which is a statistical burden. Therefore, the following assumption is made.

Assumption 6.1. *Dividend yields, social responsibility, and R&D are deterministic criteria.*

This assumption may be feasible, because people may not worry too much about uncertainty of these criteria. However, the weakness of this assumption is acknowledged, and research (for example, in Chapter 8) is being undertaken to relax this assumption. Helped by this assumption, a suitable-portfolio investor is facing the following model

$$\begin{aligned} \max\{R(\mathbf{x},\mathbf{r}) &= \text{return}\} \\ \max\{f_2(\mathbf{x}) &= \text{dividend yield}\} \\ \max\{f_3(\mathbf{x}) &= \text{social responsibility}\} \\ \max\{f_4(\mathbf{x}) &= \text{R\&D}\} \\ \text{s.t.} \quad \mathbf{x} &\in S \end{aligned} \tag{6.4}$$

where $\mathbf{r}$ is the n securities' returns. Then the variances of dividend yield, social responsibility, and R&D are all zero and can be ignored. The expected values of dividend yield, social responsibility, and R&D are just themselves. Then the substitute problem of (6.3) becomes

$$\begin{aligned} \min\{V[R(\mathbf{x},\mathbf{r})] &= \text{variance}\} \\ \max\{E[R(\mathbf{x},\mathbf{r})] &= \text{expected return}\} \\ \max\{f_2(\mathbf{x}) &= \text{dividend yield}\} \\ \max\{f_3(\mathbf{x}) &= \text{social responsibility}\} \\ \max\{f_4(\mathbf{x}) &= \text{R\&D}\} \\ \text{s.t.} \quad \mathbf{x} &\in S \end{aligned} \tag{6.5}$$

Investors can add

$$\begin{aligned} \min\{f_5(\mathbf{x}) &= \text{number of securities in portfolio}\} \\ \min\{f_6(\mathbf{x}) &= \text{maximum portfolio weight}\} \\ \min\{f_7(\mathbf{x}) &= \text{short selling}\} \\ \min\{f_8(\mathbf{x}) &= \text{turnover}\} \\ \min\{f_9(\mathbf{x}) &= \text{tracking error}\} \end{aligned}$$

where number of securities in portfolio, maximum portfolio weight, short selling, and turnover are deterministic objectives on their own and do not need the help of Assumption 6.1. That is, number of securities, maximum portfolio weight, short selling, and turnover are completely determined after the portfolio weight $\mathbf{x}$ is determined. The modelling of the criteria is in Section 6.5.

Considering (6.4), suitable-portfolio investors can introduce an *extended* utility function as follows

$$U(z_1, z_2, z_3, z_4) : \mathbb{R}^4 \to \mathbb{R} \tag{6.6}$$
$$\text{where} z_1 = R(\mathbf{x}, \mathbf{r})$$
$$z_2 = f_2(\mathbf{x})$$
$$z_3 = f_3(\mathbf{x})$$
$$z_4 = f_4(\mathbf{x})$$

The word "extended" suggests that the utility of suitable-portfolio investors depends on other variable in addition to return. This approach is endorsed by Markowitz [65] as

> "utility functions can contain state-variables in addition to return (or end of period wealth). Expected utility, in this case, can be estimated from return and state-variable means, variance and covariances provided that utility is approximately quadratic in the relevant region . . . To my knowledge, no one has investigated such quadratic approximation for cases in which state variables other than portfolio value are needed in practice" (1991, p. 476).

Suitable-portfolio investors can require $U(z_1, z_2, z_3, z_4)$ as coordinatewise increasing with respect to z_1, z_2, z_3, z_4 and concave with respect to z_1. Suitable-portfolio investors can set up the extended maximizing expected utility approach as follows

$$\max\ E[U(R(\mathbf{x}, \mathbf{r}), f_2(\mathbf{x}), \ldots, f_4(\mathbf{x}))] \tag{6.7}$$
$$\text{s.t. } \mathbf{x} \in S$$

The following theorem proves that (6.5) is consistent with (6.7) under the conditions of quadratic utility function or a multivariate normal distribution.

Theorem 6.1. *If*

(i) $U(z_1, z_2, z_3, z_4)$ *is quadratic with respect to the first variable* z_1 *and the security returns* $(r_1, \ldots, r_n)$ *are continuously distributed, or*

(ii) *the security returns* $(r_1, \ldots, r_n)$ *are multivariate normally distributed,*

then all optimal solutions of (6.7) *can be obtained by solving* (6.5) *i.e. tracing out its efficient set.*

The proof is in Steuer, Qi and Hirschberger [90].

6.2 Practical Considerations for Multiple Criteria Portfolio Selection

Practical considerations for the multiple criteria are proposed as follows. Minimizing the number of securities in a portfolio is designed to alleviate the burden of portfolio management because for individuals, every extra security is a paperwork headache, and for every mutual fund, every extra security is a managerial distraction. A look is taken at the heavy-weight player of capital markets—mutual funds. The number of holdings (securities) of the 100 largest (in net assets) mutual funds in the category "large value" by Morningstar is scatter-plotted in Figure 6.1. [1] For instance, the largest fund—American Funds Inv Co Amer A with about 50 billion in assets only holds 184 securities. The sample mean of the number of securities of these 100 mutual funds is 135. Other than for the ones that are index funds, no one appears to be volunteering to hold more securities than necessary, even though they would have the money to really spread out. Once again, as long as the performance on the other measures does not suffer greatly, fewer securities is preferred to more.

Minimizing maximum portfolio weight can originate from an investor's intention to have a relatively even portfolio weight as prescribed by Bodie and Kane and Marcus [9]

[1]Data source: Morningstar Principia Pro for Mutual Funds, November 2001. For a more revealing view, two outliers—Fund Consulting Group Large Capital Value and Fund SEI Institutional Large Capital Value A with number of holdings 743 and net asset of \$1.45 billion dollars and 693 and \$3.62 billion dollars, respectively are not included in the graph.

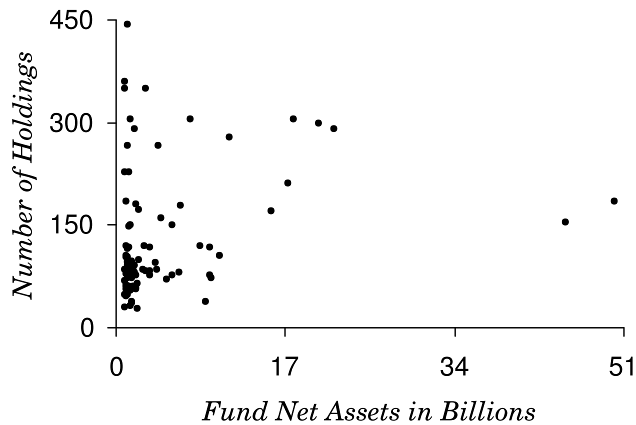

Figure 6.1: Numbers of holdings (securities) of the 100 largest “large value” mutual funds

> “If you invest half of your risky portfolio in ExxonMobil, leaving the other half in Dell ... But why stop at only two stocks? Diversifying into many more securities continues to reduce exposure to firm-specific factors” (2004, p. 170).

The same equally weighted was adopted by Elton, Gruber, Brown and Goetzmann [25]. This can be a natural style of investors without advanced knowledge of portfolio optimization. The maximum portfolio weight of the nondominated frontier of the three-security portfolio selection in Figure 4.3 is in Figure 6.2. For example, in the nondominated portfolio marked by $\mathbf{z}^4$ that has 0.092 expected return, roughly 0.72 is the largest weight in the portfolio with 0.28 in another security and nothing in the third. It can be imagined that many investors would feel uncomfortable being overweighted to such an extent in any single security. Furthermore they could be startled to try $\mathbf{z}^2$ with the maximum weight 1—investing all the money into the second security which is totally against the doctrine of diversification. A seemingly easy way to avoid to $\mathbf{z}^2$ is to install upper bounds $\mathbf{x} \leq \boldsymbol{\omega}$ to control the maximum weight. But this comes at the high price of losing a lot of promising portfolios as discussed in Section 6.3, especially in the case of large number of securities.

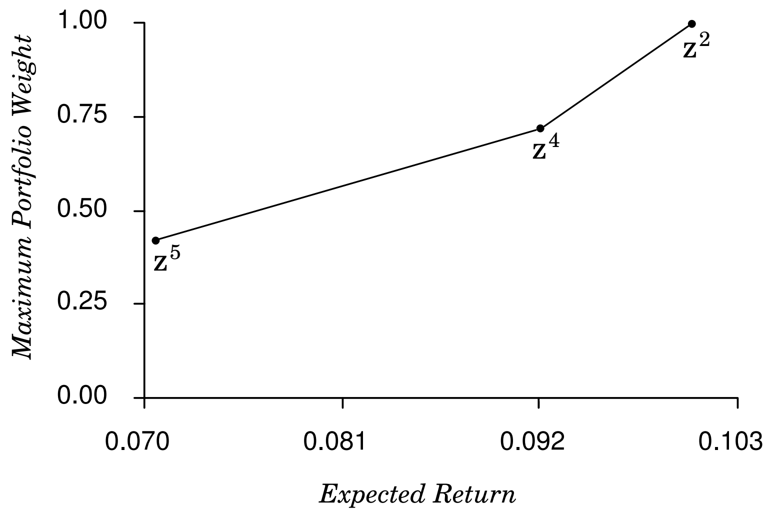

Figure 6.2: Maximum portfolio weight of the nondominated frontier of the three-security portfolio selection in Figure 4.3

Minimizing short selling is motivated by the facts that (a) Securities and Exchange Commission requires the proceeds from a short sale be kept on the account with the broker and not be available to purchase other securities, and (b) short selling increases risk.

Minimizing tracking error is designed for index fund managers and passive-management investors whose mission is to try to match the performance of a broad market index. Tracking error can be interpreted as the performance difference between the index and the managers' portfolio. The managers try to reduce the gap by minimizing tracking error.

Minimizing turnover is based on investors' consideration to control transaction costs when rebalancing portfolios. A T-test was undertaken on the 5-year annualized total return between the first 100 lowest turnover ratio "large value" mutual funds and the first 100 highest turnover ratio "large value" mutual funds by Morningstar. [2] Investors rejected the hypothesis $\mu_{\text{high turnover}} \geq \mu_{\text{low turnover}}$ at $\alpha = 0.01$ level.

Maximizing social responsibility is set up for investors to tilt their portfolios toward companies engaged in socially positive activities such as health care and environment protection.

[2]Data source: Morningstar Principia Pro for Mutual Funds, November 2001.

Maximizing R&D represents an investor's preference for the companies emphasizing research and development.

6.3 Sensitivity of Nondominated Frontiers by Standard Investors' Approach

Standard investors view suitable-portfolio investors' model (6.5) by e-constraint approach (2.3) —treating the additional criteria as constraints as follows

$$\begin{aligned} \min \{V[R(\mathbf{x},\mathbf{r})] &= \text{variance}\} \\ \text{s.t. } E[R(\mathbf{x},\mathbf{r})] &= e \\ f_2(\mathbf{x}) &\geq e_2 \\ f_3(\mathbf{x}) &\geq e_3 \\ f_4(\mathbf{x}) &\geq e_4 \\ \mathbf{x} &\in S \end{aligned} \tag{6.8}$$

as prescribed by Markowitz and Todd [66]

> "When mean-variance is used for actual money management, more complex constraints are imposed on the choice of portfolio reflecting either legal requirements or management policy. Upper bounds on individual holdings of securities as discussed in chapter 1, are a frequent requirement" (2000, p. 44).

After picking a setting of e_2, e_3 and e_4 standard investors try to obtain a nondominated frontier by changing e. But this approach leaves open the sensitivity of the nondominated frontiers to the setting of e_2, e_3 and e_4.

Example 1. Continue on (the three-security portfolio selection of) Example 1 in Chapter 3. Following the upper bound suggestion from Markowitz and Todd [66], the investor considers the model as follows

$$\min\{V[R(\mathbf{x},\mathbf{r})] = \text{variance}\} \tag{6.9}$$
$$\text{s.t.} \quad \boldsymbol{\mu}^T\mathbf{x} = e$$
$$\mathbf{1}^T\mathbf{x} = 1$$
$$\mathbf{x} \leq \boldsymbol{\omega}$$
$$\mathbf{x} \geq \mathbf{0}$$

where $\boldsymbol{\omega} = (\omega,\omega,\omega)^T$ is the upper bound vector. Then the experiment is conducted as follows.

1. Put ω to an acceptable setting.

2. Successively solve (6.9) for a sequence of 30 different right-hand-side e settings. From the solution of each e-constraint run, extract standard deviation value to form a point.

3. Connect the 30 resulting points on a graph to form a frontier.

4. Locate the nondominated frontier from the frontier obtained in Step **3** which inherently contains dominated portfolios by e-constraint approach.

5. Go to Step 1.

The investor installs three settings of ω by $\omega = 1, \omega = 0.6$ and $\omega = 0.4$, cycles through the 5-step procedure once for each ω setting, and obtains three nondominated frontiers. The three nondominated frontiers marked by $\omega = 1, \omega = 0.6$ and $\omega = 0.4$ are in Figure 6.3 *top*. The nondominated frontier of $\omega = 1$ is the longest and highest. The nondominated frontier of $\omega = 0.6$ is in the middle, with most part coinciding with the nondominated frontier of $\omega = 1$. The nondominated frontier of $\omega = 0.4$ is the shortest and lowest, with most part dominated by the nondominated frontier of $\omega = 1$. Shorter and lower nondominated frontiers mean less freedom to choose and worse portfolio performance. The nondominated frontiers undergo major changes as moving through different settings of ω. This represents considerable sensitivity. The reason is that the constraint $\mathbf{x} \leq \boldsymbol{\omega}$ can significantly reduce the original feasible region $S_{\omega=1} = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\}$ of (6.9). This $S_{\omega=1}$ is marked by the shaded region and labelled as $S_{\omega=1}$ in Figure 6.3 *middle left*. The resultant feasible regions of $\omega = 0.6$ and $\omega = 0.4$ of (6.9) are $S_{\omega=0.6} = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \leq \mathbf{0.6}, \mathbf{x} \geq \mathbf{0}\}$ and $S_{\omega=0.4} = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \leq \mathbf{0.4}, \mathbf{x} \geq \mathbf{0}\}$. These two feasible regions are shaded and marked by $S_{\omega=0.6}$ and $S_{\omega=0.4}$ in Figure 6.3 *middle right* and *bottom*, respectively. It can be seen that about half of $S_{\omega=1}$ is lost by setting $\omega = 0.6$, and most of $S_{\omega=1}$ is lost by setting $\omega = 0.4$. The financial interpretation is that sheer amount of rewarding portfolios are chopped away by some settings of this e-constraint approach so that the portfolio performance can be destructively downgraded. ◀

Example 2. Continue on (the three-security portfolio selection of) Example 1 in Chapter 3. The investor adds dividend yield into consideration as follows

$$\begin{aligned} \min \{V[R(\mathbf{x}, \mathbf{r})] &= \text{variance}\} \\ \text{s.t.} \quad \boldsymbol{\mu}^T\mathbf{x} &= e \\ \boldsymbol{\delta}^T\mathbf{x} &\geq e_2 \\ \mathbf{1}^T\mathbf{x} &= 1 \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \tag{6.10}$$

where $\boldsymbol{\delta} = (0.5\%, 0.1\%, 0.3\%)^T$ is the dividend yield vector for the three securities and the portfolio's dividend yield is formulated as $\boldsymbol{\delta}^T\mathbf{x}$. Then the experiment is conducted as follows.

1. Put e_2 to an acceptable setting.

2. Successively solve (6.10) for a sequence of 30 different right-hand-side e settings. From the solution of each e-constraint run, extract standard deviation value to form a point.

3. Connect the 30 resulting points on a graph to form a frontier.

4. Locate the nondominated frontier from the frontier obtained in Step **3** which inherently contains dominated portfolios by e-constraint approach.

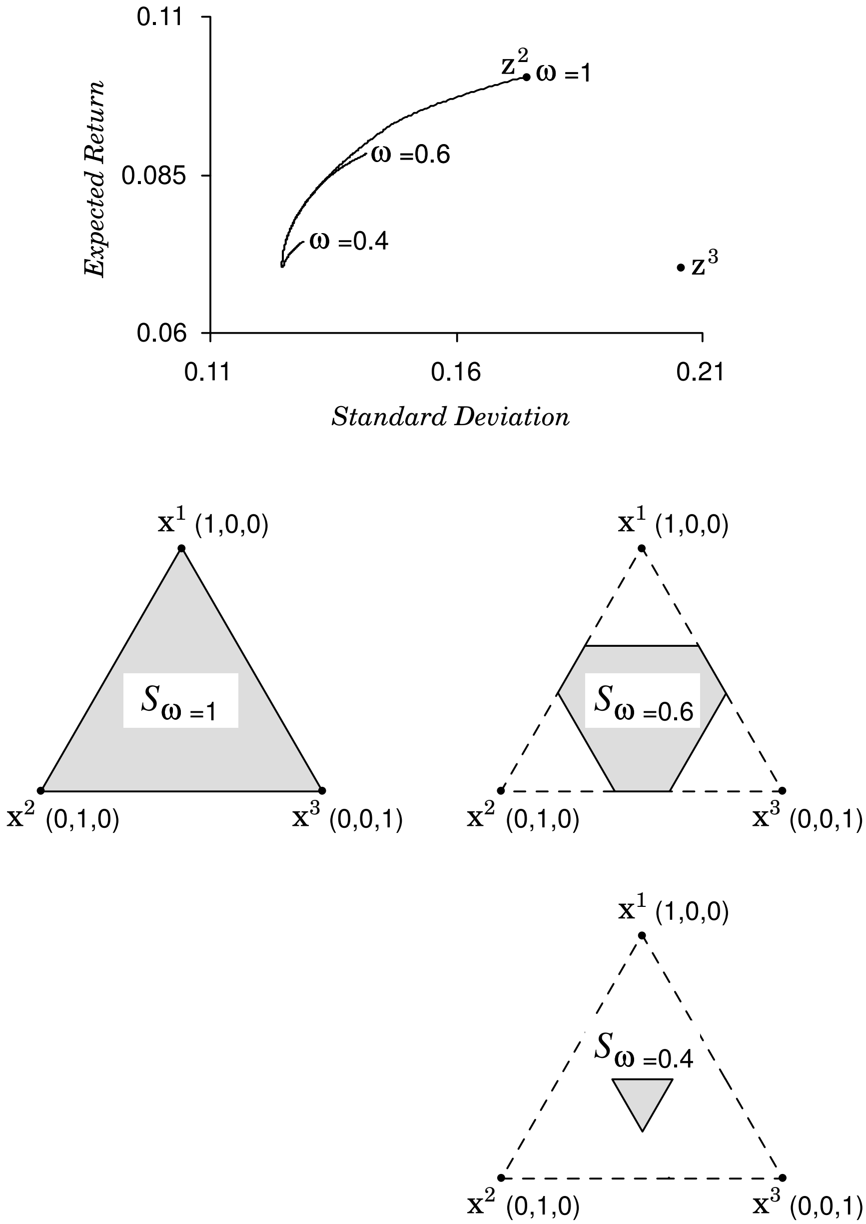

Figure 6.3: Sensitivity of nondominated frontiers to upper bound settings of $\omega = 1, \omega = 0.6$ and $\omega = 0.4$, and the corresponding feasible regions

5. Go to Step 1.

The investor installs three settings of e_2 by $e_2 = 0, e_2 = 0.2\%$ and $e_2 = 0.4\%$, cycles through the 5-step procedure once for each e_2 setting, and obtains three nondominated frontiers. The three nondominated frontiers marked by $e_2 = 0, e_2 = 0.2\%$ and $e_2 = 0.4\%$ are in Figure 6.4 *top*. The nondominated frontier of $e_2 = 0$ is the longest and highest. The nondominated frontier of $e_2 = 0.2\%$ is in the middle, with two thirds coinciding with the nondominated frontier of $e_2 = 0$. The nondominated frontier of $e_2 = 0.4\%$ is the shortest and lowest, completely dominated by the nondominated frontier of $e_2 = 0$. Shorter and lower nondominated frontiers mean less freedom to choose and worse portfolio performance. The nondominated frontiers undergo major changes as moving through different settings of e_2. This represents considerable sensitivity. The reason is that the constraint $\boldsymbol{\delta}^T\mathbf{x} \geq e_2$ can significantly reduce the original feasible region $S_{e_2=0} = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\}$ of (6.9). This $S_{e_2=0}$ is marked by the shaded region and labelled as $S_{e_2=0}$ in Figure 6.4 *middle left*. The resultant feasible regions of $e_2 = 0.2\%$ and $e_2 = 0.4\%$ of (6.9) are $S_{e_2=0.2\%} = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{1}^T\mathbf{x} = 1, \boldsymbol{\delta}^T\mathbf{x} \geq 0.2\%, \mathbf{x} \geq \mathbf{0}\}$ and $S_{e_2=0.4\%} = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{1}^T\mathbf{x} = 1, \boldsymbol{\delta}^T\mathbf{x} \geq 0.4\%, \mathbf{x} \geq \mathbf{0}\}$. These two feasible regions are shaded and marked by $S_{e_2=0.2\%}$ and $S_{e_2=0.4\%}$ in Figure 6.4 *middle right* and *bottom*, respectively. It can be seen that about one seventh of $S_{e_2=0}$ is lost by setting $e_2 = 0.2\%$, and most of $S_{e_2=0}$ is lost by setting $e_2 = 0.4\%$. The financial interpretation is that sheer amount of rewarding portfolios are chopped away by some settings of this e-constraint approach so that the portfolio performance can be destructively downgraded. ◀

Similar to dividend yield criterion, social responsibility and R&D can be modelled linearly and display analogous sensitivities. If more experiments are continued to test the sensitivity of the nondominated frontiers to short selling limitations, the number of securities in a portfolio, and so forth, investors would typically see more of the same. The experiments above are based on only three securities, much more complex and time-consuming it will be for individual investors to handle tens of securities and institutional investors to cope with

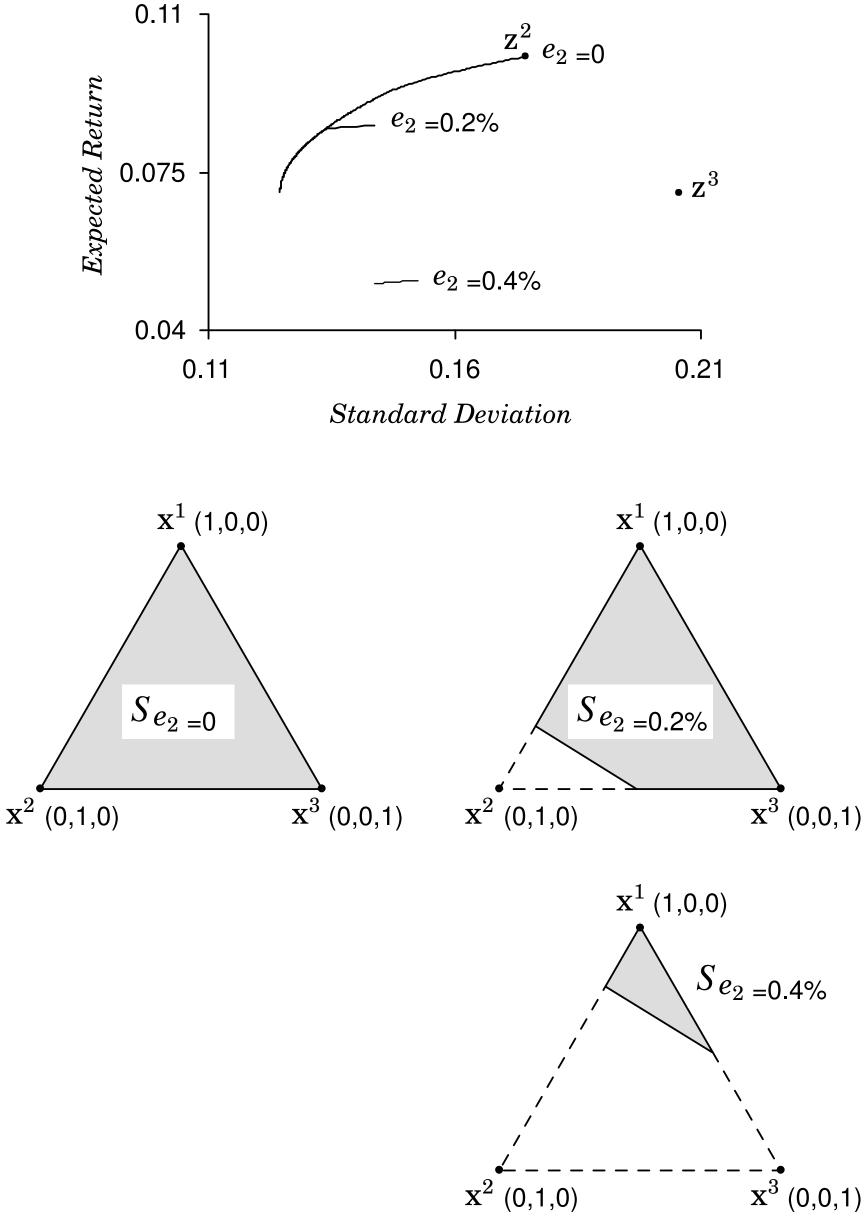

Figure 6.4: Graphs of Example 2, Sensitivity of nondominated frontiers to the dividend yield constraint

hundreds of assets in practice. The experiments above are based on fine-tuning e_2, e_3 and e_4 separately. Much more difficult it will be for investors to confront the e_2, e_3 and e_4 jointly.

This would not be so serious if an investor were in a position to know *a priori* his or her most appropriate e_2, e_3 and e_4. But with all of the tradeoffs among standard deviation, expected return, and perhaps other criteria, it is approximately not possible to know optimal e_2, e_3 and e_4 in advance. This is the rationale as to why, if dividend yield, social responsibility and others are of concern to an investor, then they should be modelled as objectives rather than constraints.

In multiple criteria decision making, there is always a line that must be drawn between what is most appropriately modelled as an objective and what is most appropriately modelled as a constraint, but in portfolio selection in finance, the line has most likely been drawn too soon. Generally, in multiple criteria, decision makers recognize a constraint from an objective as follows. If, when modelling a constraint, the decision makers realize that they can not easily fix a right-hand-side setting without knowing the sensitivity to the other criteria, then they are probably looking at an objective. In fact, this is the case with upper bound e_2, e_3 and e_4 and, depending on the investor, can easily be the case with dividends, social responsibility, number of securities in the portfolio, and short selling. This is the reason why these concerns above are modelled as criteria rather than constraints in (6.5).

6.4 Standard Investors' Approach vs. Suitable-Portfolio Investors'

Standard investors take e-constraint approach for example (6.8) as their major method, and e-constraint approach is also widely used in textbooks of finance such as Huang and Litzenberger [43], Bodie, Kane and Marcus [10], and Merton [70]. The advantage of e-constraint approach is its directness and simplicity by changing a multiple criteria optimization into a single criterion optimization. It can successfully solve the standard mean-variance problems and offer the insight of nondominated frontiers of short-sales allowed models. However, its disadvantage is as follows.

1. The ranges of $f_2(\mathbf{x}), f_3(\mathbf{x})$ and $f_4(\mathbf{x})$ should be approximately known, which is a heavy computational burden. Since e_2, e_3 and e_4 should fall into the ranges, otherwise investors may encounter infeasibility.

2. Even with e_2, e_3 and e_4 in the ranges, the e-constraints can be at odds, for example for some given e_2 and e_3, $\{\mathbf{x} \in \mathbb{R}^n \mid f_2(\mathbf{x}) \geq e_2, f_2(\mathbf{x}) \geq e_3\} = \emptyset$.

3. This approach quickly runs out of steam in model with more than three objectives, because there are at least 2^3 directions (either increasing or decreasing e_j) to modulate the e_2, e_3 and e_4, which makes systematic improvements difficult.

4. The e-constraints can substantially reduce the original feasible region. Therefore, lots of promising portfolios can be lost and the performance of portfolios can be destructively downgraded.

5. It inherently locates inefficient solutions. This can be serious if a model has more than three criteria. For example, it can be an educated guess that about $\frac{1}{2}$ of the optimal solutions of the e-constraint approach of a two objective model are inefficient, about $\frac{1}{2^2}$ of the optimal solutions of the e-constraint approach of a three objective model are inefficient, about $\frac{1}{2^3}$ of the optimal solutions of the e-constraint approach of a four objective model are inefficient, and so forth. Therefore, as the number of objectives increase, one can typically observe that more optimal solutions of the e-constraint approach are inefficient.

Suitable-portfolio investors' approach, for example (6.5) can overcome the disadvantage in the sense that the feasible region in decision space S stays the same and the feasible region in decision space Z adds dimensions, with one dimension for one additional criterion. Suitable-portfolio investors can see a panorama of tradeoffs among the criteria. However, the disadvantage of suitable-portfolio investors' approach is the difficulty to locate the nondominated surface and pinpoint the optimal solution as discussed in Section 1.1.

Relying on e-constraint approach, standard investors define nondominated frontiers as in Kroll, Levy and Markowitz [54]

> "A portfolio is mean-variance efficient if it maximizes expected rate of return (E) for a given variance (V), and minimizes the variance for a given expected return" (1984, p. 49).

and Jones [49]

> "An efficient portfolio has the highest expected return for a given level of risk, or the lowest level of risk for a given level of expected return" (2000, p. 526).

and Mayo [69] (2000, p. 163) and Perold [78] (1984, p.1144) and numerous other works. But this definition is imprecise. The feasible region Z of the three-security portfolio selection (3.10) of Example 1 in Chapter 3 is represented by the shaded region in Figure 6.5. The nondominated frontier N is the thick line. Investors will obtain portfolios $\mathbf{z}^8$ by utilizing the definition above for the given standard deviation 0.20 in Figure 6.5 *top* and $\mathbf{z}^9$ for the given expected return 0.060 in Figure 6.5 *bottom*. But neither $\mathbf{z}^8$ nor $\mathbf{z}^9$ are nondominated. They are dominated by $\mathbf{z}^2$ and $\mathbf{z}^5$, respectively.

The correct way is to define a nondominated frontier by Definition 2.2 as the nondominated set of a two objective (variance-expected return) portfolio selection model, for example as follows

$$\begin{aligned} \min \{V[R(\mathbf{x},\mathbf{r})] &= \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}\} \\ \max \{E[R(\mathbf{x},\mathbf{r})] &= \boldsymbol{\mu}^T\mathbf{x}\} \\ \text{s.t.} \qquad & \mathbf{x} \in S \end{aligned}$$

6.5 Modelling the Criteria

In this section the additional criteria are modelled. The author tries to formulate the criteria linearly, so that the procedure proposed by Hirschberger and Qi and Steuer [39] can be applied to solve some multiple objective quadratic linear programming models. Similar to standard portfolio selection in Chapter 3, assume

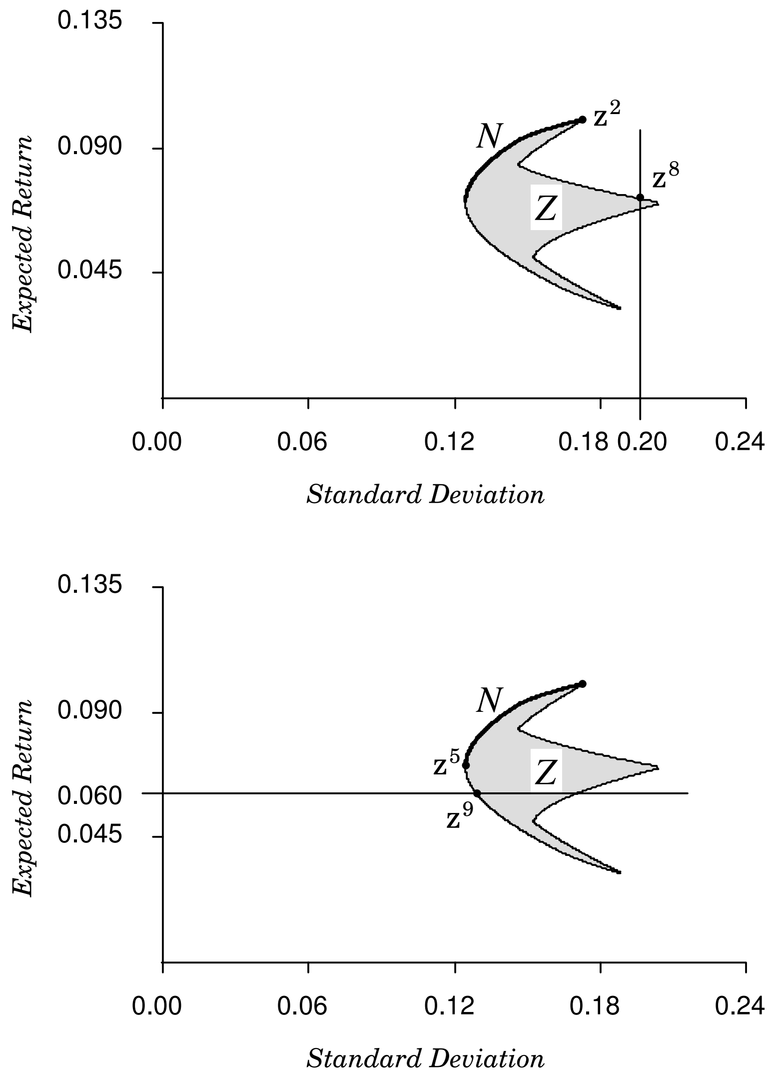

Figure 6.5: Graphs illustrating imprecise definitions of nondominated frontiers

(a) n securities

(b) an initial sum of money to be invested

(c) the beginning of a holding period

(d) the end of the holding period.

Let $\mathbf{r}$ be the n securities' returns. Let $\mathbf{x}$ be an investor's portfolio weight.

6.5.1 Dividend Yield

A security's dividend yield is defined as dividend over price. The historical data can be found from The Center for Research in Security Prices (CRSP) via wrds.wharton.upenn.edu. By Assumption 6.1, dividend yields, social responsibility, and $R\&D$ are deterministic criteria. Denote the n securities' dividend yield as $\boldsymbol{\delta}$. Then a portfolio's dividend yield can be modelled linearly as

$$\text{portfolio's dividend yield} = \boldsymbol{\delta}^T \mathbf{x} \tag{6.11}$$

6.5.2 Social Responsibility

Similarly denote the n securities' social responsibility as $\mathbf{s}$. Then a portfolio's social responsibility can be modelled linearly as

$$\text{portfolio's social responsibility} = \mathbf{s}^T \mathbf{x} \tag{6.12}$$

One difficulty is lack of historical data. What available is only the Domini 400 Social Index^{SM}. [3] It is suggested that one can randomly generate the n securities' social responsibility by uniform distribution or normal distribution.

[3] The Domini 400 Social Index^{SM} is a market capitalization-weighted common stock index. It monitors the performance of 400 U.S. corporations that pass multiple, broad-based social screens at http://www.domini.com. The index was created by the social research firm of KLD Research & Analytics, Inc. in 1989.

6.5.3 R&D

A security's R&D can be measured by dividing the annual Compustat R&D expenditures by the corresponding monthly market capitalization by Guerard and Mark [33]. A portfolio's R&D can be modelled linearly as the inner product of the n securities' R&D and the portfolio weight.

6.5.4 Number of Securities in Portfolio

Still denote the lower bound vector and the upper bound vector of portfolio weights as $(\ell_1, \ell_2, \ldots, \ell_n)$ and $(\omega_1, \omega_2, \ldots, \omega_n)$ respectively. Assume $\ell_1 < 0, \ell_2 < 0, \ldots, \ell_n < 0$. That is, limited short sales are allowed. Then investors can control the numbers of securities long and short as follows

$$\begin{aligned} \min\, & \{y_1^+ + y_2^+ + \ldots + y_n^+ = z_1\} \\ \min\, & \{y_1^- + y_2^- + \ldots + y_n^- = z_2\} \\ \text{s.t.}\quad & \omega_i y_i^+ \geq x_i,\ i = 1, 2, \ldots, n \\ & \ell_i y_i^- \leq x_i,\ i = 1, 2, \ldots, n \\ & y_i^+ \in \{0, 1\},\ i = 1, 2, \ldots, n \\ & y_i^- \in \{0, 1\},\ i = 1, 2, \ldots, n \\ & (x_1, x_2, \ldots, x_n) \in S \end{aligned} \tag{6.13}$$

where y_i^+ and y_i^- are variable names, not meaning $y_i^+ = max\{0, y_i\}$ and $y_i^- = -min\{0, y_i\}$. y_i^+ and y_i^- are binary variables to mark the status of x_i by the following three situations. The relationships between x_i and y_i^+ and y_i^- is described by the following theorem.

Theorem 6.2. *Suppose* $(x_1, x_2, \ldots, x_n,\ y_1^+, y_2^+, \ldots, y_n^+,\ y_1^-, y_2^-, \ldots, y_n^-)^T$ *is an efficient solution of* (6.13). *Then the status of* x_i*, the status of* y_i^+ *and* y_i^-*, and the relationships between* x_i *and* y_i^+ *and* y_i^- *are exhaustively and exclusively classified as follows*

(i) $x_i > 0$ *if and only if* $y_i^+ = 1$ *and* $y_i^- = 0$,

(ii) *$x_i < 0$ if and only if $y_i^+ = 0$ and $y_i^- = 1$,*

(iii) *$x_i = 0$ if and only if $y_i^+ = 0$ and $y_i^- = 0$,*

(iv) *the status $y_i^+ = 1$ and $y_i^- = 1$ is impossible.*

Proof. First take the $x_i \Rightarrow (y_i^+, y_i^-)$ direction. If $x_i > 0$, then $y_i^+ = 1$ by the constraint $\omega_i y_i^+ \geq x_i$. However, y_i^- can be both 0 and 1, because the constraint $\ell_i y_i^- \leq x_i$ does not provide useful information about y_i^-. But $y_i^- = 0$ instead of $y_i^- = 1$ is chosen, because $\min\{y_1^- + y_2^- + \ldots + y_n^- = z_2\}$ and $(x_1, x_2, \ldots, x_n,\, y_1^+, y_2^+, \ldots, y_n^+,\, y_1^-, y_2^-, \ldots, y_n^-)^T$ is an efficient solution. That is

$$\text{If } x_i > 0, \text{ then } y_i^+ = 1 \text{ and } y_i^- = 0 \tag{6.14}$$

If $x_i < 0$, then $y_i^- = 1$ by the constraint $\ell_i y_i^- \leq x_i$. However, y_i^+ can be both 0 and 1, because the constraint $\omega_i y_i^+ \geq x_i$ does not provide useful information about y_i^+. But $y_i^+ = 0$ instead of $y_i^+ = 1$ is chosen, because $\min\{y_1^+ + y_2^+ + \ldots + y_n^+ = z_1\}$ and $(x_1, x_2, \ldots, x_n,\, y_1^+, y_2^+, \ldots, y_n^+,\, y_1^-, y_2^-, \ldots, y_n^-)^T$ is an efficient solution. That is

$$\text{If } x_i < 0, \text{ then } y_i^+ = 0 \text{ and } y_i^- = 1 \tag{6.15}$$

If $x_i = 0$, then y_i^- can be both 0 and 1, and so does y_i^+ because the constraints $\ell_i y_i^- \leq x_i$ and $\omega_i y_i^+ \geq x_i$ do not provide useful information about y_i^- and y_i^+. But $y_i^- = 0$ instead of $y_i^- = 1$ is chosen, because $\min\{y_1^- + y_2^- + \ldots + y_n^- = z_2\}$ and $(x_1, x_2, \ldots, x_n,\, y_1^+, y_2^+, \ldots, y_n^+,\, y_1^-, y_2^-, \ldots, y_n^-)^T$ is an efficient solution. Likewise, $y_i^+ = 0$ instead of $y_i^+ = 1$ is chosen, because $\min\{y_1^+ + y_2^+ + \ldots + y_n^+ = z_1\}$ and

$(x_1, x_2, \ldots, x_n,\, y_1^+, y_2^+, \ldots, y_n^+,\, y_1^-, y_2^-, \ldots, y_n^-)^T$ is an efficient solution. That is

$$\text{If } x_i = 0, \text{ then } y_i^+ = 0 \text{ and } y_i^- = 0 \tag{6.16}$$

Next take the $(y_i^+, y_i^-) \Rightarrow x_i$ direction. If $y_i^+ = 1$ and $y_i^- = 0$, then $x_i \geq 0$ by the constraint $\ell_i y_i^- \leq x_i$. The status $x_i = 0$ will result in $y_i^+ = 0$ and $y_i^- = 0$ by (6.16), which contradicts $y_i^+ = 1$ and $y_i^- = 0$. Therefore $x_i > 0$. That is

$$\text{If } y_i^+ = 1 \text{ and } y_i^- = 0, \text{ then } x_i > 0 \tag{6.17}$$

If $y_i^+ = 0$ and $y_i^- = 1$, then $x_i \leq 0$ by the constraint $\omega_i y_i^+ \geq x_i$. The status $x_i = 0$ will result in $y_i^+ = 0$ and $y_i^- = 0$ by (6.16), which contradicts $y_i^+ = 0$ and $y_i^- = 1$. Therefore $x_i < 0$. That is

$$\text{If } y_i^+ = 0 \text{ and } y_i^- = 1, \text{ then } x_i < 0 \tag{6.18}$$

If $y_i^+ = 0$ and $y_i^- = 0$, then $x_i = 0$ by the constraints $\ell_i y_i^- \leq x_i$ and $\omega_i y_i^+ \geq x_i$. That is

$$\text{If } y_i^+ = 0 \text{ and } y_i^- = 0, \text{ then } x_i = 0 \tag{6.19}$$

If $y_i^+ = 1$ and $y_i^- = 1$, then x_i can be $x_i > 0$, or $x_i < 0$, or $x_i = 0$, because the constraints $\ell_i y_i^- \leq x_i$ and $\omega_i y_i^+ \geq x_i$ can not lock the status of x_i. However, the status $x_i > 0$ will result in $y_i^+ = 1$ and $y_i^- = 0$ by (6.14), which contradicts $y_i^+ = 1$ and $y_i^- = 1$. The status $x_i < 0$ will result in $y_i^+ = 0$ and $y_i^- = 1$ by (6.15), which contradicts $y_i^+ = 1$ and $y_i^- = 1$.The status $x_i = 0$ will result in $y_i^+ = 0$ and $y_i^- = 0$ by (6.16), which contradicts $y_i^+ = 1$ and $y_i^- = 1$. That is, all the possibilities $x_i > 0$, or $x_i < 0$, or $x_i = 0$ are impossible. Therefore $y_i^+ = 1$ and $y_i^- = 1$ is impossible. That is

$$\text{The status } y_i^+ = 1 \text{ and } y_i^- = 1 \text{ is impossible} \tag{6.20}$$

Conclusions (6.14) and (6.17) result in **(i)**. Conclusions (6.15) and (6.18) result in **(ii)**. Conclusions (6.16) and (6.19) result in **(iii)**. Conclusion (6.20) results in **(iv)**. Conclusions **(i)**, **(ii)**, **(iii)** and **(iv)** describe the status of x_i, the status of y_i^+ and y_i^-, and the relationships between x_i and y_i^+ and y_i^- exhaustively and exclusively. □

This formulation is based on the work of Chang, Meade, Beasley, and Sharaiha [14]. Chang, Meade, Beasley, and Sharaiha used the following portfolio selection model (in their notations)

$$\min \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} w_i \mu_i = R^*$$

$$\sum_{i=1}^{N} w_i = 1$$

$$\sum_{i=1}^{N} z_i = K$$

$$\epsilon_i z_i \leq w_i \leq \delta_i z_i,\ i = 1, 2, \ldots, N$$

$$z_i \in [0, 1],\ i = 1, 2, \ldots, N$$

where $\sum_{i=1}^{N} w_i \mu_i = R^*$ "ensures that the portfolio has an expected return of R^*" and $\sum_{i=1}^{N} z_i = K$ "ensures that exactly K assets are held" (2000, p. 1276). The $w_1, w_2, \ldots, w_N$ are portfolio weight. The $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ are the lower bound vector of portfolio weight. The $\delta_1, \delta_2, \ldots, \delta_N$ are the upper bound vector of portfolio weight.

The potential difficulty of the work of Chang, Meade, Beasley, and Sharaiha [14] is that the relationships between $w_1, w_2, \ldots, w_N$ and $z_1, z_2, \ldots, z_N$ is for only one direction, not "if and only if". For example, suppose $(w_1, w_2, \ldots, w_N, z_1, z_2, \ldots, z_N)$ is an optimal solution. Then the following relationships hold

$$w_i > 0 \Rightarrow z_1 = 1$$

$$w_i < 0 \Rightarrow z_1 = 1$$

$$w_i = 0 \Rightarrow z_1 = 1 \text{ or } z_1 = 0$$

$$z_i = 0 \Rightarrow w_i = 0$$

$$z_i = 1 \Rightarrow w_i = 0 \text{ or } w_i > 0 \text{ or } w_i < 0$$

where $\Rightarrow$ means "if, then". For example, $w_i > 0 \Rightarrow z_1 = 1$ means that if $w_i > 0$, then $z_1 = 1$. Because the relationships above are not "if and only if", their proposal that "$\sum_{i=1}^{N} z_i = K$ ensures that exactly K assets are held" may be difficult to realize.

Model (6.13) can avoid this difficulty, based on the "if and only if" relationships between x_i and y_i^+ and y_i^- of Theorem 6.2. Moreover, model (6.13) allows investors to control the number of securities long and the number of securities short separately.

6.5.5 Maximum Portfolio Weight

For a portfolio weight $\mathbf{x}$, the portfolio's maximum portfolio weight is defined as $\max\{x_1, x_2, \ldots, x_n\}$. Minimizing maximum portfolio weight is formulated as follows

$$\begin{aligned} \min \{\max\{x_1, x_2, \ldots, x_n\} &= \text{maximum portfolio weight}\} \\ \text{s.t.} \quad (x_1, x_2, \ldots, x_n) &\in S \end{aligned} \tag{6.21}$$

However, model (6.21) is nonlinear. The definition of being *equivalent* is introduced. Two models are equivalent if they have the same nondominated sets. (see Markowitz and Todd [66], p. 25). Model (6.21) is equivalent to the model as follows

$$\begin{aligned} \min \{y &= \text{maximum portfolio weight}\} \\ \text{s.t.} \quad y &\geq x_1 \\ y &\geq x_2 \\ &\vdots \\ y &\geq x_n \\ (x_1, x_2, \ldots, x_n) &\in S \end{aligned} \tag{6.22}$$

Model (6.22) is linear. Konno and Yamazaki [53] utilized similar equivalent-model technique.

6.5.6 Short Selling

For a portfolio weight $\mathbf{x}$, the portfolio's short selling is defined as $x_1^- + x_2^- + \ldots + x_n^-$ where $x_i^- = -min\{0, x_i\} = max\{0, -x_i\}, i = 1, 2, \ldots, n$. Therefore minimizing short selling is formulated as follows

$$\min \{x_1^- + x_2^- + \ldots + x_n^- = \text{short selling}\} \tag{6.23}$$
$$\text{s.t.} \qquad (x_1, x_2, \ldots, x_n) \in S$$

Although (6.23) is nonlinear, it is equivalent to the model as follows

$$\min \{y_1 + y_2 + \ldots + y_n = \text{short selling}\} \tag{6.24}$$
$$\text{s.t.} \qquad y_i \geq 0,\ i = 1, 2, \ldots, n$$
$$y_i \geq -x_i,\ i = 1, 2, \ldots, n$$
$$(x_1, x_2, \ldots, x_n) \in S$$

Model (6.24) is linear.

6.5.7 Turnover

Turnover is deploy to control transaction cost in rebalancing portfolios. Suppose an investor's current portfolio weight of is $\mathbf{x}^c$ and $\mathbf{x}^c$ is already known. Then for a portfolio weight $\mathbf{x}$, the investor will buy $(x_i - x_i^c)^+ = \max\{x_i - x_i^c, 0\}$ and sell $(x_i - x_i^c)^- = -\min\{x_i - x_i^c, 0\}$ of security i, $i = 1, 2, \ldots, n$ in order to change from $\mathbf{x}^c$ to $\mathbf{x}$. That is,

$$x_i^c + (x_i - x_i^c)^+ - (x_i - x_i^c)^- = x_i,\ i = 1, 2, \ldots, n$$

Therefore make the summation as

$$\sum_{i=1}^n x_i^c + \sum_{i=1}^n (x_i - x_i^c)^+ - \sum_{i=1}^n (x_i - x_i^c)^- = \sum_{i=1}^n x_i$$

Because $\sum_{i=1}^n x_i^c = 1$ and $\sum_{i=1}^n x_i = 1$,

$$\sum_{i=1}^n (x_i - x_i^c)^+ = \sum_{i=1}^n (x_i - x_i^c)^-$$

That is, the total percentage to buy equals the total percentage to sell in rebalancing portfolio. Therefore, the portfolio turnover can be measured by the total percentage to buy, $\sum_{i=1}^{n}(x_i - x_i^c)^+$. Then the model is

$$\begin{aligned} &\min\{(x_1 - x_1^c)^+ + (x_2 - x_2^c)^+ + \ldots + (x_n - x_n^c)^+ = \text{turnover}\} \\ &\text{s.t.} \qquad (x_1, x_2, \ldots, x_n) \in S \end{aligned} \tag{6.25}$$

Although (6.25) is nonlinear, it is equivalent to the model as follows

$$\begin{aligned} &\min\{y_1 + \ldots + y_n = \text{turnover}\} \\ &\text{s.t.} \qquad y_i \geq x_i - x_i^c,\ i = 1, 2, \ldots, n \\ &\qquad\qquad y_i \geq 0,\ i = 1, 2, \ldots, n \\ &\qquad (x_1, x_2, \ldots, x_n) \in S \end{aligned} \tag{6.26}$$

6.5.8 Tracking Error

Investors and mutual fund managers can follow the principle of passive management by trying to match and beat the performance of a broad market index. The portfolio weight of this index is denoted as $\mathbf{x}^t$ and is already known. For a portfolio weight $\mathbf{x}$, the portfolio's tracking error can be defined as $\mathbf{r}^T(\mathbf{x} - \mathbf{x}^t)$. Note that $\mathbf{r}^T(\mathbf{x} - \mathbf{x}^t)$ is a random variable. This passive management can be modelled through stochastic programming as follows

$$\begin{aligned} &\max\{\mathbf{r}^T(\mathbf{x} - \mathbf{x}^t) = \text{tracking error}\} \\ &\text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{6.27}$$

where "max" describes investors' intention to beat the index. Apply the substitute problem (6.2) to (6.27) and obtain the following model.

$$\begin{aligned} &\min\{V[\mathbf{r}^T(\mathbf{x} - \mathbf{x}^t)] = (\mathbf{x} - \mathbf{x}^t)^T \boldsymbol{\Sigma} (\mathbf{x} - \mathbf{x}^t)\} \\ &\max\{E[\mathbf{r}^T(\mathbf{x} - \mathbf{x}^t)] = \boldsymbol{\mu}^T(\mathbf{x} - \mathbf{x}^t)\} \\ &\text{s.t.} \qquad \mathbf{x} \in S \end{aligned} \tag{6.28}$$

where $V[\mathbf{r}^T(\mathbf{x}-\mathbf{x}^t)]$ and $E[\mathbf{r}^T(\mathbf{x}-\mathbf{x}^t)]$ are the variance of and expected value of $\mathbf{r}^T(\mathbf{x}-\mathbf{x}^t)$, respectively. It is proposed that minimizing number of securities in portfolio can be added to (6.28), because broad market indices usually have hundreds even thousands of securities, for example Standard & Poor's 500 Index and Wilshire 5000 Index. Therefore, with n around thousands, the efficient solutions of (6.28) can have hundreds of securities which can be out of control of investors. Note that the 100 largest "large value" mutual funds have only 135 assets on average as described in Section 6.2. Therefore the objective number of securities in portfolio can be added to (6.28) to control the number of securities as follows

$$\begin{aligned}
&\min\{V[\mathbf{r}^T(\mathbf{x}-\mathbf{x}^t)] = \text{tracking error risk}\} \\
&\max\{E[\mathbf{r}^T(\mathbf{x}-\mathbf{x}^t)] = \text{expected tracking error}\} \\
&\min\{y_1^+ + \ldots + y_n^+ = z_1\} \\
&\min\{y_1^- + \ldots + y_n^- = z_2\} \\
&\text{s.t.} \qquad \omega_i y_i^+ \geq x_i,\ i = 1, 2, \ldots, n \\
&\qquad\qquad \ell_i y_i^- \leq x_i,\ i = 1, 2, \ldots, n \\
&\qquad\qquad y_i^+ \in \{0, 1\} \\
&\qquad\qquad y_i^- \in \{0, 1\} \\
&\qquad\qquad \mathbf{x} \in S
\end{aligned} \tag{6.29}$$

Chapter 7

Tracing Out Nondominated Surfaces of Unlimited Short-sales Allowed Models

In this chapter a three objective (variance, expected return, and another linear objective) portfolio selection model is analyzed, as an extension of Merton [70]'s work to derive the nondominated frontier of a standard unlimited short-sales allowed model. With the only constraint $\mathbf{1}^T\mathbf{x} = 1$, approximately all results are in closed-form formulae to bring convenience in teaching and research and offer insight. The standard minimum-variance frontier, a parabola is extended into a minimum-variance surface, a paraboloid. The portfolio weights (inverse images) of the minimum-variance frontier is a subset of the minimum-variance surface. The efficient set of the standard model is a subset of the efficient set of this three objective model. Properties of the minimum-variance surface and nondominated surface are delineated. The connection between a suitable-portfolio investor and a standard investor can be achieved by projecting the suitable-portfolio investor' feasible region Z and nondominated surfaces onto (variance, expected return) space. It is proved that the projection of the feasible region Z of this three objective model is exactly the feasible region Z of the standard investor. The projection of the nondominated surface forms an area on (variance, expected return) space and the nondominated frontier of the standard investor is part of the boundaries of this area. The future research of adding a risk-free asset and trying to derive extended Capital Asset Pricing Model is proposed. A general k objective (variance, expected return, and other linear objectives) portfolio selection model is analyzed. The minimum-variance surface is a paraboloid in k-dimensional space. Furthermore, the previously nondominated portfolios will still stay nondominated as the suitable-portfolio investor gradually adds objectives.

A standard investor can consider a unlimited short-sales allowed model as follows

$$\begin{aligned} &\min \{\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} = z_1 \text{ variance}\} \\ &\max \{ \boldsymbol{\mu}^T\mathbf{x} = z_2 \text{ expected return}\} \\ &\text{s.t.} \quad \mathbf{1}^T\mathbf{x} = 1 \end{aligned} \tag{7.1}$$

where (7.1) can be solved by e-constraint approach. That is, by transforming objective $\boldsymbol{\mu}^T\mathbf{x} = z_2$ into e-constraint $\boldsymbol{\mu}^T\mathbf{x} = e_2$ where e_2 is a target expected return and applying Lagrangian method as in Merton [70] and Huang and Litzenberger [43]. For later convenience, denote

$$a = \boldsymbol{\mu}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu} \qquad c = \boldsymbol{\mu}^T\mathbf{\Sigma}^{-1}\mathbf{1} \qquad f = \mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1} \tag{7.2}$$

Other symbols b, d and e will be defined later. Some of the important known results of (7.1) are as follows. The portfolio weights (inverse images) of the minimum-variance frontier are as follows

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = [\frac{1}{af-cc}(-c\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + a\mathbf{\Sigma}^{-1}\mathbf{1})] + e_2[\frac{1}{af-cc}(f\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - c\mathbf{\Sigma}^{-1}\mathbf{1})], e_2 \in \mathbb{R}\} \tag{7.3}$$

Expression (7.3) is a straight line or a one dimensional affine subspace [1] in decision space $\mathbb{R}^n$, i.e., translating subspace $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = e_2[\frac{1}{af-cc}(f\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - c\mathbf{\Sigma}^{-1}\mathbf{1})], e_2 \in \mathbb{R}\}$ by vector$[\frac{1}{af-cc}(-c\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + a\mathbf{\Sigma}^{-1}\mathbf{1})]$. The minimum-variance frontier is a parabola in (variance, expected return) space as follows

$$z_1 = \frac{1}{af-cc}(fz_2^2 - 2cz_2 + a) \tag{7.4}$$

The portfolio weight, variance, and expected return of the minimum-variance portfolio is

$$\mathbf{x}^1 = \frac{1}{f}\mathbf{\Sigma}^{-1}\mathbf{1} \qquad \text{variance} = \frac{1}{f} \qquad \text{expected return} = \frac{c}{f} \tag{7.5}$$

The efficient set of (7.1) is as follows

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{x}^1 + \lambda_2(\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\mathbf{\Sigma}^{-1}\mathbf{1}), \lambda_2 \geq 0\} \tag{7.6}$$

[1] An affine subspace of $\mathbb{R}^n$ is a set of vectors $\{\mathbf{x}+\mathbf{x}^0 \mid \mathbf{x}$ belongs to a (linear) subspace S_0 of $\mathbb{R}^n\}$ for some $\mathbf{x}^0 \in \mathbb{R}^n$. That is an affine subspace is the translate of S_0 by the vector $\mathbf{x}^0$. The dimension of the affine subspace is the dimension of the S_0. Shifrin [85], p. 284.

Expression (7.6) is a one dimensional translated cone or a ray, and also a subset of (7.3) in decision space. If a risk-free asset with zero variance and deterministic return r_f exists and $r_f < \frac{c}{f}$, then there is a straight line (capital allocation line) passing through $(0, r_f)$ and tangent to the nondominated frontier of (7.1) in (standard deviation, expected return) space. Part of this straight line serves as the new nondominated frontier.

A suitable-portfolio investor can extend (7.1) by considering

$$\begin{aligned} \min \{ \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} &= z_1 \ \text{ variance}\} \\ \max \{ \ \boldsymbol{\mu}^T \mathbf{x} &= z_2 \ \text{ expected return}\} \\ \max \{ \ \boldsymbol{\delta}^T \mathbf{x} &= z_3 \ \text{ dividend yield}\} \\ \text{s.t.} \quad \mathbf{1}^T \mathbf{x} &= 1 \end{aligned} \tag{7.7}$$

They could use other linear objectives instead of dividend yield. The key difference between (7.1) and (7.7) is that (7.7) has one more dimension in criterion space. These two models have the same decision space, which is the basis for comparison, for example in Theorem 7.7. The following assumptions are made in this chapter.

Assumption 7.1. *The number of securities, n in asset universe is greater than the number of criteria, i.e., $n \geq 4$.*

Assumption 7.2. *The covariance matrix $\boldsymbol{\Sigma}$ is invertible and thus positive definite.*

Assumption 7.3. *The expected return vector $\boldsymbol{\mu}$, dividend yield vector $\boldsymbol{\delta}$, and vector one $\mathbf{1}$ are linearly independent.*

Lemma 7.1. *$\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}\mathbf{1}$ are linearly independent.*

Proof. Premultiply the equation $t_1\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + t_2\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + t_3\boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{0}$ by $\boldsymbol{\Sigma}$. Then $t_1 = 0, t_2 = 0$ and $t_3 = 0$ because $\boldsymbol{\mu}, \boldsymbol{\delta}$ and $\mathbf{1}$ are linearly independent by Assumption 7.3. Then the conclusion holds. □

7.1 Solve by e-Constraint Approach

A standard investor treats (7.7) by e-constraint approach (2.3) and obtains as follows

$$\begin{aligned} &\min \{\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} = z_1\} \\ \text{s.t.} \quad &\boldsymbol{\mu}^T\mathbf{x} = e_2 \\ &\boldsymbol{\delta}^T\mathbf{x} = e_3 \\ &\mathbf{1}^T\mathbf{x} = 1 \end{aligned} \tag{7.8}$$

where e_2 and e_3 are target expected return and target dividend yield.

First, the geometrical interpretation of (7.8) is illustrated. The proof of properties of (7.8) will be given in Theorem 7.1. The e-constraints $\boldsymbol{\mu}^T\mathbf{x} = e_2$ and $\boldsymbol{\delta}^T\mathbf{x} = e_3$ correspond to the two planes passing through $(0, e_2, 0)$ and $(0, 0, e_3)$, respectively in the axis order of variance, expected return, and dividend yield in criterion space in Figure 7.1 *top*. These two planes intersect at straight line l represented by the thick line in Figure 7.1 *top*. The lower boundary of the Z of (7.7) with respect to variance is represented by the mesh surface in Figure 7.1 *bottom*. Z is the part above and also including the surface, and represented by the shaded area. Straight line l is only partly contained in Z, and there exists a point where l intersects the surface. This point is represented by $\mathbf{z}$ in Figure 7.1 *bottom*. The points below $\mathbf{z}$ on line l have less variance than $\mathbf{z}$, but are infeasible, while the points above $\mathbf{z}$ on line l have higher variance. That is, $\mathbf{z}$ is the image of the minimizing solution of (7.8). This is the geometric interpretation of (7.8).

A suitable-portfolio investor can get the surface, i.e. the lower boundary of Z with respect to variance by varying e_2 and e_3 continuously. This surface deserves a name, *minimum-variance surface* as the extension of the minimum-variance frontier (7.4) of model (7.1). Next is the algebraic proof backing up the geometrical interpretation.

Theorem 7.1. *For any given e_2 and e_3, let $S_{e_2e_3}$ be the feasible region of* (7.8),

$$S_{e_2e_3} = \{\mathbf{x} \in \mathbb{R}^n \mid \boldsymbol{\mu}^T\mathbf{x} = e_2, \boldsymbol{\delta}^T\mathbf{x} = e_3, \mathbf{1}^T\mathbf{x} = 1\}$$

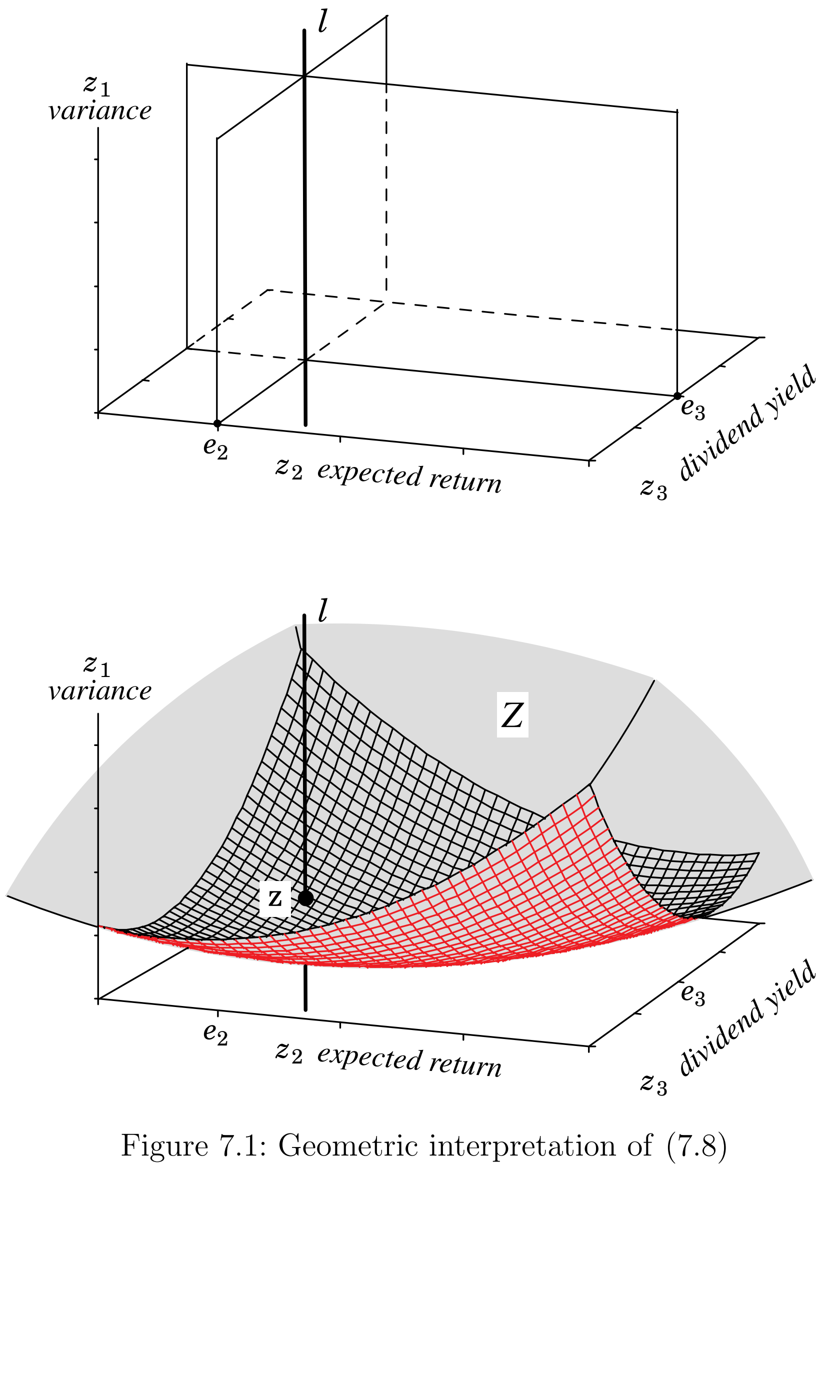

Figure 7.1: Geometric interpretation of (7.8)

Let $Z_{e_2e_3}$ be the range of (7.8)

$$Z_{e_2e_3} = \{z_1 \in \mathbb{R} \mid z_1 = \mathbf{x}^T\mathbf{\Sigma}\mathbf{x}, \mathbf{x} \in S_{e_2e_3}\}$$

Then

(i) *$S_{e_2e_3}$ is unbounded with respect to at least one of its components, i.e., there exists $i \in \{1, 2, \ldots, n\}$ such that for any $y > 0$, there exists $(x_1, x_2, \ldots, x_n) \in S_{e_2e_3}$ with $|x_i| > y$.*

(ii) *$Z_{e_2e_3}$ is lower bounded by 0, i.e., for any $z_1 \in Z_{e_2e_3}, z_1 \geq 0$, but upper unbounded, i.e., for any $y > 0$, there exists $z_1 \in Z_{e_2e_3}$ with $z_1 > y$.*

Proof. $S_{e_2e_3}$ can be re-expressed as the solutions of the linear equations $\begin{bmatrix} \boldsymbol{\mu}^T \\ \boldsymbol{\delta}^T \\ \mathbf{1}^T \end{bmatrix}_{3\times n} \mathbf{x}_{n\times 1} = \begin{bmatrix} e_2 \\ e_3 \\ 1 \end{bmatrix}_{3\times 1}$ where the subscripts $3 \times n, n \times 1$ and 3×1 show the dimensions.

Because $\boldsymbol{\mu}$, $\boldsymbol{\delta}$ and $\mathbf{1}$ are linearly independent by Assumption 7.3, the rank of matrix $\begin{bmatrix} \boldsymbol{\mu}^T \\ \boldsymbol{\delta}^T \\ \mathbf{1}^T \end{bmatrix}$ is 3. Furthermore $n \geq 4$ by Assumption 7.1, then by the fact from linear algebra (for example Kolman [50]) $S_{e_2e_3}$ is not an empty set and

$$S_{e_2e_3} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \bar{\mathbf{x}} + t_1\bar{\mathbf{x}}^1 + t_2\bar{\mathbf{x}}^2 + \ldots + t_{n-3}\bar{\mathbf{x}}^{n-3}, t_1, t_2, \ldots t_{n-3} \in \mathbb{R}\}$$

where $\bar{\mathbf{x}}$ is a feasible solution of (7.8), i.e. $\bar{\mathbf{x}} \in S_{e_2e_3}$ and $\bar{\mathbf{x}}^1 \neq \mathbf{0}, \bar{\mathbf{x}}^2 \neq \mathbf{0}, \ldots \bar{\mathbf{x}}^{n-3} \neq \mathbf{0}$ are the basis of the solutions of the homogeneous linear equations $\begin{bmatrix} \boldsymbol{\mu}^T \\ \boldsymbol{\delta}^T \\ \mathbf{1}^T \end{bmatrix} \mathbf{x} = \mathbf{0}$.

Just take $\mathbf{x} = \bar{\mathbf{x}} + t_1\bar{\mathbf{x}}^1$ by letting $t_2 = 0, t_3 = 0, \ldots, t_{n-3} = 0$. Because $\bar{\mathbf{x}}^1 \neq \mathbf{0}$, there exists $\bar{x}_i^1 \neq 0$. Then $|x_i^1| = |\bar{x}_i + t_1\bar{x}_i^1| \to \infty$ as $t_1 \to \infty$. Therefore, $S_{e_2e_3}$ is unbounded with respect to x_i, which is **(i)**. Then

$$\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} = (\bar{\mathbf{x}} + t_1\bar{\mathbf{x}}^1)^T\mathbf{\Sigma}(\bar{\mathbf{x}} + t_1\bar{\mathbf{x}}^1) = (\bar{\mathbf{x}}^1)^T\mathbf{\Sigma}\bar{\mathbf{x}}^1\, t_1^2 + 2\bar{\mathbf{x}}^T\mathbf{\Sigma}\bar{\mathbf{x}}^1\, t_1 + \bar{\mathbf{x}}^T\mathbf{\Sigma}\bar{\mathbf{x}} \to \infty\,, \text{ as } t_1 \to \infty$$

, because the limit is determined by the coefficient of t_1^2, $(\bar{\mathbf{x}}^1)^T\mathbf{\Sigma}\bar{\mathbf{x}}^1 > 0$ since $\mathbf{\Sigma}$ is positive definite. Then $Z_{e_2e_3}$ is upper unbounded. It is lower bounded by 0, because variances are always nonnegative, which is **(ii)**. □

The straight line l is the image of $\{\mathbf{x} \in \mathbb{R}^n \mid \boldsymbol{\mu}^T\mathbf{x} = e_2, \boldsymbol{\delta}^T\mathbf{x} = e_3\}$. The part of l contained in Z is the image of $S_{e_2e_3}$. Theorem 7.1 reveals the shape of the feasible region Z of (7.7), that is, for given expected return and dividend yield, the variance is lower bounded by 0 but upper unbounded, which implies the existence of the minimum-variance surface. The exact location of this surface will be handled by (7.9)-(7.13). The next theorem proves the continuity of Z of model (7.7) and extends the work of Markowitz and Todd[66] p. 225 who proved the continuity of the feasible region Z of mean-variance models.

Theorem 7.2. *The feasible region Z of* (7.7) *has the following properties.*

(i) *For any given target expected return e_2 and target dividend yield e_3, if (σ_a^2, e_2, e_3) and (σ_c^2, e_2, e_3) are images of two feasible portfolios of* (7.7)*, with $\sigma_a^2 < \sigma_c^2$, then for any $\sigma_b^2 \in [\sigma_a^2, \sigma_c^2]$ there exists a feasible portfolio of* (7.7) *such that its image is (σ_b^2, e_2, e_3).*

(ii) *Z is continuous.*

Proof. Take $\mathbf{x}^a$ and $\mathbf{x}^c$ as the two feasible portfolios whose images are (σ_a^2, e_2, e_3) and (σ_c^2, e_2, e_3), respectively, i.e.,

$$\mathbf{x}^{aT}\mathbf{\Sigma}\mathbf{x}^a = \sigma_a^2 \qquad \boldsymbol{\mu}^T\mathbf{x}^a = e_2 \qquad \boldsymbol{\delta}^T\mathbf{x}^a = e_3$$

$$\mathbf{x}^{cT}\mathbf{\Sigma}\mathbf{x}^c = \sigma_c^2 \qquad \boldsymbol{\mu}^T\mathbf{x}^c = e_2 \qquad \boldsymbol{\delta}^T\mathbf{x}^c = e_3$$

As proved in Theorem 7.1, $S_{e_2e_3} \subseteq S$ is not empty. Moreover, $S_{e_2e_3}$ is a convex set in decision space, because $S_{e_2e_3} = \{\mathbf{x} \in \mathbb{R}^n \mid \boldsymbol{\mu}^T\mathbf{x} = e_2\} \cap \{\mathbf{x} \in \mathbb{R}^n \mid \boldsymbol{\delta}^T\mathbf{x} = e_3\} \cap \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{1}^T\mathbf{x} = 1\}$. Each of the three sets is a hyperplane and therefore convex set in decision space. Then $S_{e_2e_3}$ as the intersection of these three sets is also convex.

It is clear that $\mathbf{x}^a \in S_{e_2e_3}$ and $\mathbf{x}^c \in S_{e_2e_3}$. Then $\mathbf{x}^b = (1-t)\mathbf{x}^a + t\mathbf{x}^c \in S_{e_2e_3}$ and $\mathbf{x}^b \in S$ for any $t \in [0,1]$, because $S_{e_2e_3}$ is convex.

$$\boldsymbol{\mu}^T\mathbf{x}^b = (1-t)\boldsymbol{\mu}^T\mathbf{x}^a + t\boldsymbol{\mu}^T\mathbf{x}^c = (1-t)e_2 + te_2 = e_2$$
$$\boldsymbol{\delta}^T\mathbf{x}^b = (1-t)\boldsymbol{\delta}^T\mathbf{x}^a + t\boldsymbol{\delta}^T\mathbf{x}^c = (1-t)e_3 + te_3 = e_3$$

Let

$$f(t) = \mathbf{x}^{bT}\boldsymbol{\Sigma}\mathbf{x}^b = ((1-t)\mathbf{x}^a + t\mathbf{x}^c)^T\boldsymbol{\Sigma}((1-t)\mathbf{x}^a + t\mathbf{x}^c)$$

$f(t)$ is continuous with respect to t. Furthermore, $f(0) = \sigma_a^2, f(1) = \sigma_c^2$. Therefore, by Mid-Value Theorem for any $\sigma_b^2 \in [\sigma_a^2, \sigma_c^2]$, there exists a $t_b \in [0,1]$ such that $f(t_b) = \sigma_b^2$. Then $\mathbf{x}^b = (1-t_b)\mathbf{x}^a + t_b\mathbf{x}^c$ is the feasible portfolio with the image (σ_b^2, e_2, e_3). This is the proof of **(i)**.

(ii) follows from the fact that the steps above hold for any given e_2 and e_3, and so does Theorem 7.1. □

To solve (7.8) and obtain the minimum-variance surface, apply Lagrangian method by constructing

$$L(\mathbf{x}, \lambda_2, \lambda_3, \lambda_4) = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} + \lambda_2(e_2 - \boldsymbol{\mu}^T\mathbf{x}) + \lambda_3(e_3 - \boldsymbol{\delta}^T\mathbf{x}) + \lambda_4(1 - \mathbf{1}^T\mathbf{x}) \tag{7.9}$$

where λ_2, λ_3 and λ_4 are Lagrangian multipliers. Because $\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}$ is positive semidefinite with respect to $\mathbf{x}$, $\mathbf{x}$ is the minimizing solution of (7.8) if and only if

$$\frac{\partial L}{\partial \mathbf{x}} = 2\boldsymbol{\Sigma}\mathbf{x} - \lambda_2\boldsymbol{\mu} - \lambda_3\boldsymbol{\delta} - \lambda_4\mathbf{1} = \mathbf{0} \tag{7.10}$$
$$\frac{\partial L}{\partial \lambda_2} = e_2 - \boldsymbol{\mu}^T\mathbf{x} = 0 \tag{7.11}$$
$$\frac{\partial L}{\partial \lambda_3} = e_3 - \boldsymbol{\delta}^T\mathbf{x} = 0 \tag{7.12}$$
$$\frac{\partial L}{\partial \lambda_4} = 1 - \mathbf{1}^T\mathbf{x} = 0 \tag{7.13}$$

Because $\boldsymbol{\Sigma}$ is invertible, then (7.10) is

$$\mathbf{x} = \frac{1}{2}(\lambda_2\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \lambda_3\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + \lambda_4\boldsymbol{\Sigma}^{-1}\mathbf{1}) \tag{7.14}$$

Pre-multiply (7.14) by $\boldsymbol{\mu}^T, \boldsymbol{\delta}^T$ and $\mathbf{1}^T$ respectively and substitute the products in (7.11), (7.12), and (7.13), respectively to obtain.

$$\lambda_2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \lambda_3 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} + \lambda_4 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = 2e_2$$

$$\lambda_2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} + \lambda_3 \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} + \lambda_4 \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = 2e_3$$

$$\lambda_2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} + \lambda_3 \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} + \lambda_4 \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = 2$$

Re-express these three equations in matrix form and obtain

$$\begin{bmatrix} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} & \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} & \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \\ \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} & \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} & \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \\ \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} & \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} & \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \end{bmatrix}_{3\times 3} \begin{bmatrix} \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} 2e_2 \\ 2e_3 \\ 2 \end{bmatrix} \tag{7.15}$$

Denote

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}_{3\times 3} \tag{7.16}$$

where

$$a = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \qquad b = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} \qquad c = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \tag{7.17}$$

$$d = \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} \qquad e = \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \qquad f = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$$

Symbols a, c and f are defined the same as in (7.2). By Assumption 7.2, $\boldsymbol{\Sigma}$ is positive definite, and so is $\boldsymbol{\Sigma}^{-1}$ (Huang and Litzenberger [43], p. 64). Therefore

$$a > 0 \qquad d > 0 \qquad f > 0 \tag{7.18}$$

The lemma below reveals the properties of $\mathbf{A}$.

Lemma 7.2. $\mathbf{A}$ *is also a covariance matrix. Furthermore, it is positive definite and thus invertible and the following inequalities hold.*

$$af - cc > 0 \qquad ad - bb > 0 \qquad df - ee > 0 \tag{7.19}$$

Proof. Because a square matrix is positive definite if and only if its inverse is also positive definite (Huang and Litzenberger [43], p. 64), $\mathbf{\Sigma}^{-1}$ is also positive definite by Assumption 7.2. Then it can function as a covariance matrix . That is, there exists a random vector $\mathbf{v}$ such that the covariance matrix of $\mathbf{v}$ is $\mathbf{\Sigma}^{-1}$. Furthermore, $\boldsymbol{\mu}^T\mathbf{v}, \boldsymbol{\delta}^T\mathbf{v}$ and $\mathbf{1}^T\mathbf{v}$ are random variables and the covariance matrix of these three is exactly $\mathbf{A}$ (7.16). A closer look at $\mathbf{A}$ reveals that

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{\mu}^T \\ \boldsymbol{\delta}^T \\ \mathbf{1}^T \end{bmatrix}_{3\times n} \quad \mathbf{\Sigma}^{-1}_{n\times n} \begin{bmatrix} \boldsymbol{\mu} & \boldsymbol{\delta} & \mathbf{1} \end{bmatrix}_{n\times 3}$$

For any $\mathbf{y} \in \mathbb{R}^3$,

$$\mathbf{y}^T_{1\times 3}\mathbf{A}_{3\times 3}\mathbf{y}_{3\times 1} = \mathbf{y}^T \begin{bmatrix} \boldsymbol{\mu}^T \\ \boldsymbol{\delta}^T \\ \mathbf{1}^T \end{bmatrix} \mathbf{\Sigma}^{-1} \begin{bmatrix} \boldsymbol{\mu} & \boldsymbol{\delta} & \mathbf{1} \end{bmatrix} \mathbf{y}$$

Let $\mathbf{x} = \begin{bmatrix} \boldsymbol{\mu} & \boldsymbol{\delta} & \mathbf{1} \end{bmatrix}_{n\times 3} \mathbf{y}_{3\times 1}$. Then

$$\mathbf{y}^T\mathbf{A}\mathbf{y} = \mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} \ > 0, \quad \text{because } \mathbf{\Sigma}^{-1} \text{ is positive definite}$$

Therefore $\mathbf{A}$ is positive definite and thus invertible . Because $\mathbf{A}$ is positive definite and vector $c\boldsymbol{\mu} - a\mathbf{1} \in \mathbb{R}^n$.

$$\begin{aligned}
&(c\boldsymbol{\mu} - a\mathbf{1})^T\mathbf{\Sigma}^{-1}(c\boldsymbol{\mu} - a\mathbf{1}) > 0 \\
&(c\boldsymbol{\mu} - a\mathbf{1})^T\mathbf{\Sigma}^{-1}(c\boldsymbol{\mu} - a\mathbf{1}) = cc\boldsymbol{\mu}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + aa\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1} - 2ac\boldsymbol{\mu}^T\mathbf{\Sigma}^{-1}\mathbf{1}, \text{ by (7.17)} \\
&\qquad\qquad\qquad\qquad\qquad\quad = cca + aaf - 2acc \\
&\qquad\qquad\qquad\qquad\qquad\quad = a(af - cc)
\end{aligned}$$

Because $a > 0$ by (7.18), then $af - cc > 0$. Similarly $ad - bb > 0$ and $df - ee > 0$ by taking the vector as $b\boldsymbol{\mu} - a\boldsymbol{\delta}$ and $e\boldsymbol{\delta} - d\mathbf{1}$ respectively. □

Then re-express (7.15) as

$$\mathbf{A}\begin{bmatrix}\lambda_2\\ \lambda_3\\ \lambda_4\end{bmatrix} = \begin{bmatrix}2e_2\\ 2e_3\\ 2\end{bmatrix} \tag{7.20}$$

By (7.16),

$$|\mathbf{A}| = adf + 2bce - aee - bbf - ccd \tag{7.21}$$

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix}df-ee & ce-bf & be-cd\\ ce-bf & af-cc & bc-ae\\ be-cd & bc-ae & ad-bb\end{bmatrix}_{3\times 3} \tag{7.22}$$

Because $|\mathbf{A}|$ is invertible by Lemma 7.2, then (7.20) is solved by

$$\begin{bmatrix}\lambda_2\\ \lambda_3\\ \lambda_4\end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix}2e_2\\ 2e_3\\ 2\end{bmatrix} \quad \text{then by (7.22)}$$

$$= \frac{2}{|\mathbf{A}|}\begin{bmatrix}e_2(df-ee)+e_3(ce-bf)+(be-cd)\\ e_2(ce-bf)+e_3(af-cc)+(bc-ae)\\ e_2(be-cd)+e_3(bc-ae)+(ad-bb)\end{bmatrix}_{3\times 1} \tag{7.23}$$

Then plug (7.23) into (7.14)

$$\begin{aligned}\mathbf{x} = \frac{1}{|\mathbf{A}|}&[(e_2(df-ee)+e_3(ce-bf)+(be-cd))\mathbf{\Sigma}^{-1}\boldsymbol{\mu}\\ &+(e_2(ce-bf)+e_3(af-cc)+(bc-ae))\mathbf{\Sigma}^{-1}\boldsymbol{\delta}\\ &+(e_2(be-cd)+e_3(bc-ae)+(ad-bb))\mathbf{\Sigma}^{-1}\mathbf{1}]\\ = e_2&\mathbf{d}^2+e_3\mathbf{d}^3+\mathbf{x}^0\end{aligned} \tag{7.24}$$

where

$$\mathbf{d}^2 = \frac{1}{|\mathbf{A}|}[(df-ee)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (ce-bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (be-cd)\boldsymbol{\Sigma}^{-1}\mathbf{1}] \tag{7.25}$$

$$\mathbf{d}^3 = \frac{1}{|\mathbf{A}|}[(ce-bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (af-cc)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (bc-ae)\boldsymbol{\Sigma}^{-1}\mathbf{1}] \tag{7.26}$$

$$\mathbf{x}^0 = \frac{1}{|\mathbf{A}|}[(be-cd)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (bc-ae)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (ad-bb)\boldsymbol{\Sigma}^{-1}\mathbf{1}] \tag{7.27}$$

The following theorem proves that $\mathbf{d}^2$ and $\mathbf{d}^3$ are linearly independent. The interpretation of $\mathbf{x}^0$, $\mathbf{d}^2$ and $\mathbf{d}^3$ will be given later.

Theorem 7.3. *Vectors $\mathbf{d}^2$ and $\mathbf{d}^3$ are linearly independent.*

Proof. By (7.25) and (7.26),

$$\begin{aligned} e_2\mathbf{d}^2 + e_3\mathbf{d}^3 &= \frac{e_2}{|\mathbf{A}|}[(df-ee)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (ce-bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (be-cd)\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\ &\quad + \frac{e_3}{|\mathbf{A}|}[(ce-bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (af-cc)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (bc-ae)\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\ &= \frac{1}{|\mathbf{A}|}\{[e_2(df-ee) + e_3(ce-bf)]\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ &\quad + [e_2(ce-bf) + e_3(af-cc)]\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} \\ &\quad + [e_2(be-cd) + e_3(bc-ae)]\boldsymbol{\Sigma}^{-1}\mathbf{1}\} \end{aligned}$$

By Lemma 7.1, the necessary and sufficient condition of $e_2\mathbf{d}^2 + e_3\mathbf{d}^3 = \mathbf{0}$ is

$$(df-ee)e_2 + (ce-bf)e_3 = 0 \tag{7.28}$$

$$(ce-bf)e_2 + (af-cc)e_3 = 0 \tag{7.29}$$

$$(be-cd)e_2 + (bc-ae)e_3 = 0 \tag{7.30}$$

By (7.29) and $af - cc > 0$ in (7.19),

$$e_3 = \frac{(bf-ce)}{af-cc}\,e_2$$

Plug it into (7.28)

$$\begin{aligned}(df-ee)e_2+(ce-bf)\frac{(bf-ce)}{af-cc}\,e_2&=0\\ \frac{adff-aeef-ccdf+ccee-ccee-bbff+2bcef}{af-cc}\,e_2&=0\\ \frac{f(adf+2bce-aee-bbf-ccd)}{af-cc}\,e_2&=0,\text{ by (7.21)}\\ \frac{f|\mathbf{A}|}{af-cc}\,e_2&=0\end{aligned}$$

Because $f>0$ by (7.18), $|\mathbf{A}|\neq 0$ by Lemma 7.2, and $af-cc>0$ by (7.19), $e_2=0$ and then $e_3=0$.

This $e_2=0$ and $e_3=0$ also satisfy (7.30). Then one knows that the only solution of (7.28)-(7.30) is $e_2=0$ and $e_3=0$, which implies the conclusion. □

The interpretation of $\mathbf{x}^0$ is that it is the minimizing solution of (7.8) when $e_2=0$ and $e_3=0$. $\mathbf{d}^3$ and $\mathbf{d}^2$ are two directions (vectors) so that $\mathbf{d}^2+\mathbf{x}^0$ is the minimizing solution of (7.8) when $e_2=1$ and $e_3=0$ and so is $\mathbf{d}^3+\mathbf{x}^0$ when $e_2=0$ and $e_3=1$. Next introduce

$$\{\mathbf{x}\in\mathbb{R}^n \mid \mathbf{x}=e_2\mathbf{d}^2+e_3\mathbf{d}^3,\; e_2,e_3\in\mathbb{R}\} \tag{7.31}$$

Expression (7.31) is a two dimensional subspace in decision space, because $\mathbf{d}^2$ and $\mathbf{d}^3$ are linearly independent as proved in Theorem 7.3. It is spanned (formed by linear combinations) by $\mathbf{d}^2$ and $\mathbf{d}^3$. Next translate subspace (7.31) by $\mathbf{x}^0$ as follows

$$\{\mathbf{x}\in\mathbb{R}^n \mid \mathbf{x}=e_2\mathbf{d}^2+e_3\mathbf{d}^3+\mathbf{x}^0,\; e_2,e_3\in\mathbb{R}\} \tag{7.32}$$

Expression (7.32) is a two dimensional affine subspace in decision space. For given e_2 and e_3, the minimizing solution of (7.8) is (7.24). As both e_2 and e_3 change from $-\infty$ to ∞, the minimizing solutions of (7.8), are described by (7.32). Therefore, (7.32) is the portfolio weights (inverse images) of minimum-variance surface and extends the portfolio weights of the minimum-variance frontier (7.3) (one dimensional affine subspace). Furthermore, it will be proved in Theorem 7.13 that this one dimensional affine subspace is a subset of the two dimensional affine subspace (7.32).

In the following corollary, Assumption 7.1 is relaxed while Assumptions 7.2 and 7.3 still hold.

Corollary 7.1. *If the asset universe contains only three securities* ($n = 3$)*, then the feasible region* S *of* (7.7) *is the inverse images (portfolio weights) of the minimum-variance surface* (7.32)*, and consequently,* Z *of* (7.7) *is the minimum-variance surface.*

Proof. For $n = 3$, $S = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{1}^T\mathbf{x} = 1\}$ is a two dimensional affine subspace. As a subset of S, the inverse images (portfolio weights) of the minimum-variance surface (7.32) is also a two dimensional affine subspace by Theorem (7.3). Therefore left is the only possibility that these two are the same. Consequently their images match, too. □

Therefore for $n = 3$, Z of (7.7) is unfilled, i.e., composed of only a boundary (shell) which is a paraboloid as proved later in Theorem 7.5. The image of every feasible solution (portfolio) is on this boundary. For suitable-portfolio investors deploying (7.7), this serves the same building-block role as the 2-security portfolio ($n = 2$) of standard investors deploying (7.1) and locating a parabola-shaped Z.

Theorem 7.4. *The inverse images* (*portfolio weights*) *of the minimum-variance surface* (7.32) *has the following properties.*

- **(i)** *Any three points* (*solutions*) $\mathbf{x}^a, \mathbf{x}^b$ *and* $\mathbf{x}^c$ *on it can span* (*generate*) *this affine subspace, as long as* $\mathbf{x}^a, \mathbf{x}^b$ *and* $\mathbf{x}^c$ *are affinely independent.*[2]
- **(ii)** *Any convex combination* [3] *of any* m *points on this affine subspace is still on this affine subspace.*

Proof. **(i)** is based on the fact in algebra that any three affinely independent points determine a two dimensional affine subspace, see Shifrin [85], pp. 290-291. The affine subspace can be

[2] *Points* $\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^m$ *are affinely independent if* $\mathbf{x}^1 - \mathbf{x}^0, \mathbf{x}^2 - \mathbf{x}^0, \ldots, \mathbf{x}^m - \mathbf{x}^0$ *are linearly independent, Shifrin [85], p. 285.*

[3] *A convex combination of* $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m$ *is taken as* $a_1\mathbf{x}^1 + a_2\mathbf{x}^2 + \ldots + a_m\mathbf{x}^m$ *for given* $a_1, a_2, \ldots, a_m \geq 0$ *and* $a_1 + a_2 + \ldots + a_m = 1$.

expressed as $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{x}^a + t_b(\mathbf{x}^b - \mathbf{x}^a) + t_c(\mathbf{x}^c - \mathbf{x}^a),\ t_b, t_c \in \mathbb{R}\}$. The affine subspace is a convex set. **(ii)** is a property of convex sets. □

Theorem 7.4 extends the structure of the portfolio weights of the minimum-variance frontier (7.3) by Huang and Litzenberger [43] that

> "the entire portfolio frontier can be generated by forming portfolios of the two frontier portfolios $\mathbf{g}$ and $\mathbf{g} + \mathbf{h}$ ... the portfolio frontier can be generated by *any* two distinct frontier portfolios" (1988, pp. 65-66).

where $\mathbf{g}$ and $\mathbf{h}$ are $[\frac{1}{af-cc}(-c\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + a\mathbf{\Sigma}^{-1}\mathbf{1})]$ and $[\frac{1}{af-cc}(f\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - c\mathbf{\Sigma}^{-1}\mathbf{1})]$ of (7.3). It also implies that if all investors hold portfolios on the minimum-variance surface, then the market portfolio as a convex combination of all the investors' portfolios by Huang and Litzenberger [43] p. 83 is also on the minimum-variance surface. The subset relationship between the portfolio weights of the minimum-variance frontier (7.3) and the portfolio weights of the minimum-variance surface (7.32) will be proved in Theorem 7.13.

7.2 Paraboloid

Some properties of paraboloid are outlined, because these properties are closely related to portfolio selection. The simple form of paraboloid in y-x_1-x_2-…-x_n space is

$$y = a_1x_1^2 + a_2x_2^2 + \ldots + a_nx_n^2 \tag{7.33}$$

where $a_1, a_2, \ldots, a_n$ are fixed and nonnegative coefficients. If some $a_i = 0$, then it is a degenerate paraboloid or paraboloidal cylinder. For example, a straight line $y = 0$ is a degenerate paraboloid with $a_1 = 0$ in x-y space and so is $y = x_2^2$ with $a_1 = 0$ in x_1-x_2-y space. If all $a_i > 0$, then it is a non-degenerate paraboloid. General forms of paraboloid can be obtained by changes of coordinate systems by rotating and (or) translating the simple form.

It will be proved in Theorem 7.5 that the minimum-variance surface is a non-degenerate paraboloid in (variance, expected return, and dividend yield) space. Another application of paraboloid in portfolio selection is that the variance $V[R(\mathbf{x},\mathbf{r})] = \mathbf{x}^T\mathbf{\Sigma}\mathbf{x}$ defines a paraboloid in variance-x_1-x_2-…-x_n space. To see this, first $\mathbf{\Sigma}$ is a symmetric matrix and all its elements are real numbers. Therefore there exists a normal matrix [4] $\mathbf{N}$ such that

$$\mathbf{\Sigma}_{n\times n} = \mathbf{N}^T_{n\times n} \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & & \\ \vdots & & & \vdots \\ 0 & & \cdots & v_n \end{bmatrix}_{n\times n} \mathbf{N}_{n\times n}$$

where $v_1, v_2, \ldots, v_n$ are the eigenvalues of $\mathbf{\Sigma}$. Because $\mathbf{\Sigma}$ is positive semidefinite, all it eigenvalues $v_1, v_2, \ldots, v_n$ are nonnegative, i.e. $v_1 \geq 0, v_2 \geq 0, \ldots, v_n \geq 0$.

Let $\mathbf{y} = \mathbf{N}_{n\times n}\mathbf{x}_{n\times 1}$. Then

$$\begin{aligned} V[R(\mathbf{x},\mathbf{r})] &= \mathbf{x}^T\mathbf{N}^T \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & & \\ \vdots & & & \vdots \\ 0 & & \cdots & v_n \end{bmatrix} \mathbf{N}\mathbf{x} \\ &= \mathbf{y}^T \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & & \\ \vdots & & & \vdots \\ 0 & & \cdots & v_n \end{bmatrix} \mathbf{y} \\ &= v_1y_1^2 + v_2y_2^2 + \ldots + v_ny_n^2 \end{aligned}$$

which is a paraboloid in variance-y_1-y_2-…-y_n space, and also a paraboloid variance-x_1-x_2-…-x_n space because $\mathbf{y} = \mathbf{N}\mathbf{x}$ is a change of coordinate systems by rotation. Furthermore, if some $v_i = 0$ (or $\mathbf{\Sigma}$ is positive semidefinite as a necessary and sufficient condition), then the paraboloid is degenerate. If all $v_i > 0$ (or $\mathbf{\Sigma}$ is positive definite, as a necessary and sufficient condition), then the paraboloid is non-degenerate.

[4] A matrix $\mathbf{N} \in \mathbb{R}^{n\times n}$ is a normal matrix if $\mathbf{N}^T\mathbf{N} = \mathbf{I}$ where $\mathbf{I} \in \mathbb{R}^{n\times n}$ is an identity matrix.

Lemma 7.3. *A paraboloid has the following properties.*

(i) *A degenerate paraboloid has at least one* cylinder direction [5] *which is a direction such that one can slide along this direction infinitely from any given point belonging to the paraboloid, and the path is still belonging to the paraboloid.*

(ii) *A non-degenerate paraboloid does not have any cylinder direction. Moreover, for any two different points belonging to the paraboloid, there does not exist a linear segment both passing through these two points and belonging to the paraboloid.*

Proof. This proof is based on the simple form of paraboloid (7.33), because general forms can be obtained by changes of coordinate systems by rotating and (or) translating the simple form. For a degenerate paraboloid, assume without loss of generality that $a_1 = 0$. Then take the cylinder direction as $(1, 0, \ldots, 0, 0)^T_{(n+1)\times 1}$. Then from any point belonging to the paraboloid, $(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, \hat{y})$, i.e., $\hat{y} = 0\hat{x}_1^2 + a_2\hat{x}_2^2 + \ldots + a_n\hat{x}_n^2$. The sliding along the cylinder direction can be represented by

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \\ \hat{y} \end{bmatrix}_{(n+1)\times 1} + t \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}_{(n+1)\times 1} = \begin{bmatrix} \hat{x}_1 + t \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \\ \hat{y} \end{bmatrix}_{(n+1)\times 1} \text{, for any } t \in \mathbb{R}$$

The path is still belonging to the paraboloid, because

$$\begin{aligned} \hat{y} &= 0(\hat{x}_1 + t)^2 + a_2\hat{x}_2^2 + \ldots + a_n\hat{x}_n^2 \\ &= 0\hat{x}_1^2 + a_2\hat{x}_2^2 + \ldots + a_n\hat{x}_n^2 \end{aligned}$$

[5] *The author calls it "cylinder direction", because another name of degenerate paraboloid is paraboloidal cylinders formed by sliding the paraboloid along a direction. For example, $y = x_2^2$ with $a_1 = 0$ in x_1-x_2-y space is formed by plotting the parabola $y = x_2^2$ in x_2-y space and then sliding it along the direction of x_1 axis. Markowitz [66] p. 112 calls this direction "zero-variance direction" because moving along it will not change variance in $V[R(\boldsymbol{x}, \boldsymbol{r})] = v_1y_1^2 + v_2y_2^2 + \ldots + v_ny_n^2$ for some $v_i = 0$.*

which is **(i)**. For a non-degenerate paraboloid (7.33) with all $a_i > 0$, it is clear that this paraboloid does not have any cylinder direction, because all $a_i > 0$. Take any two different points belonging to the paraboloid, $(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, \hat{y})$ and $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \bar{y})$. There exists an i such that $\hat{x}_i \neq \bar{x}_i$, otherwise identical $\hat{x}_i = \bar{x}_i, i = 1, 2, \ldots, n$ would lead to the same $\hat{y} = \bar{y}$ which means the two points are the same. Without loss of generality, assume $\hat{x}_1 \neq \bar{x}_1$. Because the paraboloid is smooth, the normal of any point belonging to the paraboloid exists. Take (7.33) into

$$F(x_1, x_2, \ldots, x_n, y) = 0$$

where

$$F(x_1, x_2, \ldots, x_n, y) = a_1 x_1^2 + a_2 x_2^2 + \ldots + a_n x_n^2 - y$$

The normal is obtained by taking partial derivatives

$$\frac{\partial F}{\partial x_1} = 2a_1 x_1 \qquad \frac{\partial F}{\partial x_2} = 2a_2 x_2 \qquad \ldots \qquad \frac{\partial F}{\partial x_n} = 2a_n x_n \qquad \frac{\partial F}{\partial y} = -1$$

The first and last elements of normal of $(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, \hat{y})$ are

$$\frac{\partial F}{\partial x_1}|_{(}\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, \hat{y}) = 2a_1 \hat{x}_1 \qquad \frac{\partial F}{\partial y}|_{(}\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, \hat{y}) = -1$$

while the first and last elements of normal of $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \bar{y})$ are

$$\frac{\partial F}{\partial x_1}|_{(}\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \bar{y}) = 2a_1 \bar{x}_1 \qquad \frac{\partial F}{\partial y}|_{(}\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \bar{y}) = -1$$

The normals of $(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, \hat{y})$ and $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \bar{y})$ are not parallel to each other, because their last elements are equal while their first elements are not. There does not exist a linear segment both passing through these two points and belonging to the paraboloid, because if it were the case, then the normals of these two points would be parallel to each other. $\square$

7.3 Minimum-variance Surface

The minimum-variance surface is obtained by plugging in the expression of the affine subspace (7.32).

$$\begin{aligned} V[R(\mathbf{x},\mathbf{r})] &= (e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \mathbf{x}^0)^T\boldsymbol{\Sigma}(e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \mathbf{x}^0) \\ &= \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 e_2^2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3 e_3^2 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 e_2 e_3 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0 e_2 + 2\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0 e_3 + \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0 \end{aligned}$$

or re-express in the symbols of z_1, z_2 and z_3 of criterion space as

$$z_1 = \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 z_2^2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_3^2 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 z_3 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_2 + 2\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_3 + \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0 \tag{7.34}$$

Re-express $\mathbf{x} = e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \mathbf{x}^0$ (7.32) in matrix format

$$\mathbf{x} = \begin{bmatrix}\mathbf{d}^2 & \mathbf{d}^3 & \mathbf{x}^0\end{bmatrix}_{n\times 3} \begin{bmatrix} e_2 \\ e_3 \\ 1 \end{bmatrix}_{3\times 1}$$

or re-express in the symbols of z_1, z_2 and z_3 of criterion space as

$$\mathbf{x} = \begin{bmatrix}\mathbf{d}^2 & \mathbf{d}^3 & \mathbf{x}^0\end{bmatrix}_{n\times 3} \begin{bmatrix} z_2 \\ z_3 \\ 1 \end{bmatrix}_{3\times 1}$$

Then $V[R(\mathbf{x},\mathbf{r})] = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}$ is

$$z_1 = \begin{bmatrix} z_2 & z_3 & 1 \end{bmatrix}_{1\times 3} \begin{bmatrix} \mathbf{d}^{2^T} \\ \mathbf{d}^{3^T} \\ \mathbf{x}^{0^T} \end{bmatrix}_{3\times n} \boldsymbol{\Sigma}_{n\times n} \begin{bmatrix}\mathbf{d}^2 & \mathbf{d}^3 & \mathbf{x}^0\end{bmatrix}_{n\times 3} \begin{bmatrix} z_2 \\ z_3 \\ 1 \end{bmatrix}_{3\times 1}$$

Denote

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}^{2^T} \\ \mathbf{d}^{3^T} \\ \mathbf{x}^{0^T} \end{bmatrix}_{3\times n} \quad \mathbf{\Sigma}_{n\times n} \begin{bmatrix} \mathbf{d}^2 & \mathbf{d}^3 & \mathbf{x}^0 \end{bmatrix}_{n\times 3}$$

$$= \begin{bmatrix} \mathbf{d}^{2^T}\mathbf{\Sigma}\mathbf{d}^2 & \mathbf{d}^{2^T}\mathbf{\Sigma}\mathbf{d}^3 & \mathbf{d}^{2^T}\mathbf{\Sigma}\mathbf{x}^0 \\ \mathbf{d}^{2^T}\mathbf{\Sigma}\mathbf{d}^3 & \mathbf{d}^{3^T}\mathbf{\Sigma}\mathbf{d}^3 & \mathbf{d}^{3^T}\mathbf{\Sigma}\mathbf{x}^0 \\ \mathbf{d}^{2^T}\mathbf{\Sigma}\mathbf{x}^0 & \mathbf{d}^{3^T}\mathbf{\Sigma}\mathbf{x}^0 & \mathbf{x}^{0^T}\mathbf{\Sigma}\mathbf{x}^0 \end{bmatrix}_{3\times 3} \tag{7.35}$$

Lemma 7.4. *Matrix* $\mathbf{D}$ (7.35) *is a covariance matrix and is positive definite.*

Proof. A closer look at $\mathbf{D}$ reveals that it is the covariance matrix of random variables $\mathbf{d}^{2^T}\mathbf{r}, \mathbf{d}^{3^T}\mathbf{r}$ and $\mathbf{x}^{0^T}\mathbf{r}$. Therefore $\mathbf{D}$ is positive semidefinite. For any $\mathbf{y} \in \mathbb{R}^3$,

$$\mathbf{y}^T\mathbf{D}\mathbf{y} = \mathbf{y}^T \begin{bmatrix} \mathbf{d}^{2^T} \\ \mathbf{d}^{3^T} \\ \mathbf{x}^{0^T} \end{bmatrix} \mathbf{\Sigma} \begin{bmatrix} \mathbf{d}^2 & \mathbf{d}^3 & \mathbf{x}^0 \end{bmatrix} \mathbf{y}$$

Let $\qquad \mathbf{x}_{n\times 1} = \begin{bmatrix} \mathbf{d}^2 & \mathbf{d}^3 & \mathbf{x}^0 \end{bmatrix}_{n\times 3} \mathbf{y}_{3\times 1}$. Then

$$\mathbf{y}^T\mathbf{D}\mathbf{y} = \mathbf{x}^T\mathbf{\Sigma}\mathbf{x} > 0 \quad \text{, because } \mathbf{\Sigma} \text{ is positive definite}$$

Therefore, $\mathbf{D}$ is positive definite. □

Theorem 7.5. *The minimum-variance surface* (7.34) *is a non-degenerate paraboloid in (variance, expected return, and dividend yield) space.*

Proof. Because $\mathbf{D}$ is positive definite by Lemma 7.4, all its eigenvalues v_1, v_2 and v_3 are positive as a necessary and sufficient condition . Because $\mathbf{D}$ is symmetric and all its elements are real numbers, there exists a normal matrix $\mathbf{N}$ such that

$$\mathbf{D} = \mathbf{N}^T_{3\times 3} \begin{bmatrix} v_1 & 0 & 0 \\ 0 & v_2 & 0 \\ 0 & 0 & v_3 \end{bmatrix}_{3\times 3} \quad \mathbf{N}_{3\times 3}$$

Let

$$\mathbf{y} = \mathbf{N} \begin{bmatrix} z_2 \\ z_3 \\ 1 \end{bmatrix} \tag{7.36}$$

Then the minimum-variance surface (7.34) is

$$\begin{aligned} z_1 &= \begin{bmatrix} z_2 & z_3 & 1 \end{bmatrix} \mathbf{D} \begin{bmatrix} z_2 \\ z_3 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} z_2 & z_3 & 1 \end{bmatrix} \mathbf{N}^T \begin{bmatrix} v_1 & 0 & 0 \\ 0 & v_2 & 0 \\ 0 & 0 & v_3 \end{bmatrix} \mathbf{N} \begin{bmatrix} z_2 \\ z_3 \\ 1 \end{bmatrix} \\ &= \mathbf{y}^T \begin{bmatrix} v_1 & 0 & 0 \\ 0 & v_2 & 0 \\ 0 & 0 & v_3 \end{bmatrix} \mathbf{y} \\ &= v_1 y_1^2 + v_2 y_2^2 + v_3 y_3^2 \end{aligned}$$

With $v_1, v_2, v_3 > 0$, this is the expression of a non-degenerate paraboloid in z_1-y_1-y_2-y_3 space, and also in z_1-z_2-z_3 space, because $\mathbf{y} = \mathbf{N} \begin{bmatrix} z_2 \\ z_3 \\ 1 \end{bmatrix}$ can be taken as a change of coordinate system by rotation. □

Corollary 7.2. *The feasible region Z of* (7.7) *is an unbounded continuous set in criterion space, with a non-degenerate paraboloidal boundary.*

Proof. The conclusion follows from the facts that Z is upper unbounded respect to variance by Theorem 7.1 and continuous by Theorem 7.2, the minimum-variance surface is the lower boundary of Z with respect to variance, and the minimum-variance surface is a non-degenerate paraboloid by Theorem 7.5. □

The minimum-variance surface (7.34) can be delineated by calculating the six elements of $\mathbf{D}$ (7.35) next. Because $\mathbf{D}$ is positive definite, the following properties of its elements hold, similar to (7.18) and (7.19) of positive definite matrix $\mathbf{A}$ (7.16).

$$\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 > 0 \qquad \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3 > 0 \qquad \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0 > 0 \tag{7.37}$$

$$\begin{aligned}
(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2)(\mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0) - (\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0)(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0) > 0\\
(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3) - (\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3) > 0\\
(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0) - (\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0) > 0
\end{aligned} \tag{7.38}$$

$$\begin{aligned}
\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 &= \frac{1}{|\mathbf{A}|^2}[(df-ee)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} + (ce-bf)\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1} + (be-cd)\mathbf{1}^T\boldsymbol{\Sigma}^{-1}]\\
&\quad \boldsymbol{\Sigma}\,[(df-ee)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (ce-bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (be-cd)\boldsymbol{\Sigma}^{-1}\mathbf{1}], \text{ by (7.17)}\\
&= \frac{1}{|\mathbf{A}|^2}[(df-ee)^2a + (ce-bf)^2d + (be-cd)^2f\\
&\quad + 2(df-ee)(ce-bf)b + 2(df-ee)(be-cd)c + 2(ce-bf)(be-cd)e]\\
&= \frac{1}{|\mathbf{A}|^2}[(df-ee)^2a + (ce-bf)^2d + (be-cd)^2f\\
&\quad + 2(df-ee)(ce-bf)b + 2(df-ee)(be-cd)c + 2(ce-bf)(be-cd)e]\\
&= \frac{1}{|\mathbf{A}|^2}[(addff - 2adeef + aeeee) + (ccdee - 2bcdef + bbdff)\\
&\quad + (bbeef - 2bcdef + ccddf) + (2bcdef - 2bceee - 2bbdff + 2bbeef)\\
&\quad + (2bcdef - 2bceee - 2ccddf + 2ccdee) + (2bceee - 2bbeef - 2ccdee + 2bcdef)]\\
&= \frac{1}{|\mathbf{A}|^2}[addff - 2adeef + aeeee - bbdff + bbeef + 2bcdef - 2bceee - ccddf + ccdee]\\
&= \frac{1}{|\mathbf{A}|^2}[ad^2f^2 - 2ade^2f + ae^4 - b^2df^2 + b^2e^2f + 2bcdef - 2bce^3 - c^2d^2f + c^2de^2]
\end{aligned} \tag{7.39}$$

Some terms are in gray to denote they are cancelled. For example $-2bcdef$ in the first row of the last equality part is cancelled by $+2bcdef$ in the second row. The gray 2 in $-2bbdff$ in the second row means this term is subtracted by $bbdff$ in the first row.

$$
\begin{aligned}
\mathbf{d}^{3T}\boldsymbol{\Sigma}\,\mathbf{d}^3 &= \frac{1}{|\mathbf{A}|^2}[(ce-bf)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} + (af-cc)\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1} + (bc-ae)\mathbf{1}^T\boldsymbol{\Sigma}^{-1}] \\
&\quad \boldsymbol{\Sigma}\,[(ce-bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (af-cc)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (bc-ae)\boldsymbol{\Sigma}^{-1}\mathbf{1}], \text{ by (7.17)} \\
&= \frac{1}{|\mathbf{A}|^2}[(ce-bf)^2a + (af-cc)^2d + (bc-ae)^2f \\
&\quad + 2(ce-bf)(af-cc)b + 2(ce-bf)(bc-ae)c + 2(af-cc)(bc-ae)e] \\
&= \frac{1}{|\mathbf{A}|^2}[(accee - 2abcef + abbff) + (aadff - 2accdf + ccccd) \\
&\quad + (bbccf - 2abcef + aaeef) + (2abcef - 2abbff - 2bccce + 2bbccf) \\
&\quad + (2bccce - 2bbccf - 2accee + 2abcef) + (2abcef - 2bccce - 2aaeef + 2accee)] \\
&= \frac{1}{|\mathbf{A}|^2}[aadff - aaeef - abbff + 2abcef - 2accdf + accee + bbccf - 2bccce + ccccd] \\
&= \frac{1}{|\mathbf{A}|^2}[a^2df^2 - a^2e^2f - ab^2f^2 + 2abcef - 2ac^2df + ac^2e^2 + b^2c^2f - 2bc^3e + c^4d]
\end{aligned}
\tag{7.40}
$$

$$
\begin{aligned}
\mathbf{x}^{0T}\boldsymbol{\Sigma}\,\mathbf{x}^0 &= \frac{1}{|\mathbf{A}|^2}[(be-cd)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} + (bc-ae)\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1} + (ad-bb)\mathbf{1}^T\boldsymbol{\Sigma}^{-1}] \\
&\quad \boldsymbol{\Sigma}\,[(be-cd)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (bc-ae)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (ad-bb)\boldsymbol{\Sigma}^{-1}\mathbf{1}], \text{ by (7.17)} \\
&= \frac{1}{|\mathbf{A}|^2}[(be-cd)^2a + (bc-ae)^2d + (ad-bb)^2f \\
&\quad + 2(be-cd)(bc-ae)b + 2(be-cd)(ad-bb)c + 2(bc-ae)(ad-bb)e] \\
&= \frac{1}{|\mathbf{A}|^2}[(abbee - 2abcde + accdd) + (bbccd - 2abcde + aadee) \\
&\quad + (aaddf - 2abbdf + bbbbf) + (2bbbce - 2bbccd - 2abbee + 2abcde) \\
&\quad + (2abcde - 2accdd - 2bbbce + 2bbccd) + (2abcde - 2aadee - 2bbbce + 2abbee)] \\
&= \frac{1}{|\mathbf{A}|^2}[aaddf - aadee - 2abbdf + abbee + 2abcde - accdd - 2bbbce + bbbbf + bbccd] \\
&= \frac{1}{|\mathbf{A}|^2}[a^2d^2f - a^2de^2 - 2ab^2df + ab^2e^2 + 2abcde - ac^2d^2 - 2b^3ce + b^4f + b^2c^2d]
\end{aligned}
\tag{7.41}
$$

$$
\begin{aligned}
\mathbf{d}^{2T}\boldsymbol{\Sigma}\,\mathbf{d}^3 &= \frac{1}{|\mathbf{A}|^2}[(df-ee)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} + (ce-bf)\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1} + (be-cd)\mathbf{1}^T\boldsymbol{\Sigma}^{-1}] \\
&\quad \boldsymbol{\Sigma}\,[(ce-bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (af-cc)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (bc-ae)\boldsymbol{\Sigma}^{-1}\mathbf{1}], \text{ by (7.17)} \\
&= \frac{1}{|\mathbf{A}|^2}[(df-ee)(ce-bf)a + (ce-bf)(af-cc)d + (be-cd)(bc-ae)f \\
&\quad + ((df-ee)(af-cc) + (ce-bf)(ce-bf))b \\
&\quad + ((df-ee)(bc-ae) + (be-cd)(ce-bf))c \\
&\quad + ((ce-bf)(bc-ae) + (be-cd)(af-cc))e] \\
&= \frac{1}{|\mathbf{A}|^2}[(acdef - aceee - abdff + abeef) + (acdef - abdff - cccde + bccdf) \\
&\quad + (bbcef - bccdf - abeef + acdef) \\
&\quad + (abdff - abeef - bccdf + bccee) + (bccee - 2bbcef + bbbff) \\
&\quad + (bccdf - bccee - acdef + aceee) + (bccee - cccde - bbcef + bccdf) \\
&\quad + (bccee - bbcef - aceee + abeef) + (abeef - acdef - bccee + cccde)] \\
&= \frac{1}{|\mathbf{A}|^2}[-abdff + abeef + acdef - aceee + bbbff + bccdf - 3bbcef + 2bccee - cccde] \\
&= \frac{1}{|\mathbf{A}|^2}[-abdf^2 + abe^2f + acdef - ace^3 + b^3f^2 + bc^2df - 3b^2cef + 2bc^2e^2 - c^3de]
\end{aligned}
\tag{7.42}
$$

$$\begin{aligned}
\mathbf{d}^{2^T}\boldsymbol{\Sigma}\,\mathbf{x}^0 &= \frac{1}{|\mathbf{A}|^2}[(df-ee)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}+(ce-bf)\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1}+(be-cd)\mathbf{1}^T\boldsymbol{\Sigma}^{-1}] \\
&\quad \boldsymbol{\Sigma}\,[(be-cd)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}+(bc-ae)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}+(ad-bb)\boldsymbol{\Sigma}^{-1}\mathbf{1}],\ \text{by (7.17)} \\
&= \frac{1}{|\mathbf{A}|^2}[(df-ee)(be-cd)a+(ce-bf)(bc-ae)d+(be-cd)(ad-bb)f \\
&\quad +((df-ee)(bc-ae)+(ce-bf)(be-cd))b \\
&\quad +((df-ee)(ad-bb)+(be-cd)(be-cd))c \\
&\quad +((ce-bf)(ad-bb)+(be-cd)(bc-ae))e] \\
&= \frac{1}{|\mathbf{A}|^2}[(abdef-abeee-acddf+acdee)+(bccde-bbcdf-acdee+abdef) \\
&\quad +(abdef-acddf-bbbef+bbcdf) \\
&\quad +(bbcdf-bbcee-abdef+abeee)+(bbcee-bbbef-bccde+bbcdf) \\
&\quad +(acddf-acdee-bbcdf+bbcee)+(bbcee-2bccde+cccdd) \\
&\quad +(acdee-abdef-bbcee+bbbef)+(bbcee-bccde-abeee+acdee)] \\
&= \frac{1}{|\mathbf{A}|^2}[abdef-abeee-acddf+acdee-bbbef+bbcdf+2bbcee-3bccde+cccdd] \\
&= \frac{1}{|\mathbf{A}|^2}[abdef-abe^3-acd^2f+acde^2-b^3ef+b^2cdf+2b^2ce^2-3bc^2de+c^3d^2]
\end{aligned} \tag{7.43}$$

$$
\begin{aligned}
\mathbf{d}^{3T}\boldsymbol{\Sigma}\,\mathbf{x}^0 &= \frac{1}{|\mathbf{A}|^2}[(ce-bf)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} + (af-cc)\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1} + (bc-ae)\mathbf{1}^T\boldsymbol{\Sigma}^{-1}] \\
&\quad \boldsymbol{\Sigma}\,[(be-cd)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (bc-ae)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (ad-bb)\boldsymbol{\Sigma}^{-1}\mathbf{1}],\ \text{by (7.17)} \\
&= \frac{1}{|\mathbf{A}|^2}[(ce-bf)(be-cd)a + (af-cc)(bc-ae)d + (bc-ae)(ad-bb)f \\
&\quad + ((ce-bf)(bc-ae) + (af-cc)(be-cd))b \\
&\quad + ((ce-bf)(ad-bb) + (bc-ae)(be-cd))c \\
&\quad + ((af-cc)(ad-bb) + (bc-ae)(bc-ae))e] \\
&= \frac{1}{|\mathbf{A}|^2}[(abcee - abbef - accde + abcdf) + (abcdf - bcccd - aadef + accde) \\
&\quad + (abcdf - aadef - bbbcf + abbef) \\
&\quad + (bbcce - bbbcf - abcee + abbef) + (abbef - bbcce - abcdf + bcccd) \\
&\quad + (accde - abcdf - bbcce + bbbcf) + (bbcce - abcee - bcccd + accde) \\
&\quad + (aadef - accde - abbef + bbcce) + (bbcce - 2abcee + aaeee)] \\
&= \frac{1}{|\mathbf{A}|^2}[-aadef + aaeee + abbef + abcdf - 3abcee + accde - bbbcf + 2bbcce - bcccd] \\
&= \frac{1}{|\mathbf{A}|^2}[-a^2def + a^2e^3 + ab^2ef + abcdf - 3abce^2 + ac^2de - b^3cf + 2b^2c^2e - bc^3d]
\end{aligned}
\tag{7.44}
$$

This paraboloidal minimum-variance surface of (7.7) extends the parabolic minimum-variance frontier of (7.1). However, the expression of the surface is convoluted, with six terms and nine sub-terms in each term. Identifying minimum-variance portfolio and nondominated surface from the expression of the minimum-variance surface is difficult. Therefore weighted-sums approach will be tried to locate the minimum-variance portfolio and nondominated surface in Section 7.4. In contrast, the expression of the minimum-variance frontier (7.4) is relatively simple and apparent (see Huang and Litzenberger [43]) so that investors know its parabolic shape by viewing the expression.

Another difference is that the paraboloid is *rotated*, i.e., the major axis or minor axis of (7.34) for fixed z_1 (variance) is not parallel to either z_2 axis or z_3 axis in criterion space, if the coefficient of $z_2 z_3$, $\mathbf{d}^{2^T}\mathbf{\Sigma}\mathbf{d}^3$ is not equal to 0, where equation (7.34) for fixed z_1 can be taken as an indifference curve of (7.34) and is an ellipse. See Figure 7.7 in Section 7.7 for an example. The paraboloid is *un-tilted*, i.e., its (central) axis around which it is symmetric is parallel to z_1 axis. See Figure 7.4 in Section 7.7 for an example. In contrast, the minimum-variance frontier of (7.1) is unrotated, that is, two points are obtained by fixing z_1 (variance), as the degenerate form of an ellipse. The straight line passing through these two points is parallel to z_2 axis. Because this parabola is on two dimensional space, the concept of tilted is not applicable.

7.4 Solve by Weighted-Sums Approach

Because the three objective functions of (7.7) are convex and the feasible region S is a convex set, the weakly efficient set can be located by weighted-sums approach (see Ehrgott [23], p. 59). The definition of being weakly efficient will be introduced later. It will be proved in Theorem 7.6 that the weakly efficient set of (7.7) is the efficient set. The weighted-sums approach can be written as follows

$$\begin{aligned} \min \{\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} - \lambda_2\boldsymbol{\mu}^T\mathbf{x} - \lambda_3\boldsymbol{\delta}^T\mathbf{x} = z_w\} \quad \lambda_2 \geq 0, \lambda_3 \geq 0 \\ \text{s.t.} \qquad \mathbf{1}^T\mathbf{x} = 1 \end{aligned} \tag{7.45}$$

or equivalently by multiplying (7.45) by -1

$$\begin{aligned} -\max \{-\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} + \lambda_2\boldsymbol{\mu}^T\mathbf{x} + \lambda_3\boldsymbol{\delta}^T\mathbf{x} = -z_w\} \quad \lambda_2 \geq 0, \lambda_3 \geq 0 \\ \text{s.t.} \qquad \mathbf{1}^T\mathbf{x} = 1 \end{aligned} \tag{7.46}$$

$(1, \lambda_2, \lambda_3)^T$ is the weighting vector. The financial interpretation of (7.46) is that $-\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} + \lambda_2\boldsymbol{\mu}^T\mathbf{x} + \lambda_3\boldsymbol{\delta}^T\mathbf{x}$ can serve as expected utility. Sharpe [82] took similar approach as

> “we consider an investor whose utility function has three or more arguments – one quadratic and the others linear in the decision variables. To illustrate we use an example in which an investor associates a disutility with income yield due to

its unfavorable tax treatment relative to capital gains. Letting yp be the yield of the portfolio, utility is now:

$$up = ep + uy * yp - vp/rt$$

where uy (utility from yield) is a constant indicating the investor's attitude towards yield...we can convert the utility function to variance-equivalent terms by multiplying all terns by rt, giving:

$$vup = rt * ep + (uy * rt) * yp - vp$$

" (2004)

where up is expected utility, ep is expected return, vp is variance, rt is risk tolerance, and vup is up times rt. The connection between (7.46) and Sharpe's second model is that (a) $\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}, \boldsymbol{\mu}^T\mathbf{x}$, and $\boldsymbol{\delta}^T\mathbf{x}$ of (7.46) correspond to vp, ep, and yp of Sharpe's second model respectively (b) the weighting vector $(1, \lambda_2, \lambda_3)$ of $\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}, \boldsymbol{\mu}^T\mathbf{x}$, and $\boldsymbol{\delta}^T\mathbf{x}$ of (7.46) corresponds to the weighting vector $(1, rt, uy * rt)$ of vp, ep, and yp of Sharpe's second model. Therefore, (7.46) serves as extended maximizing expected utility approach. While Sharpe relies on quadratic utility function, (7.45) is based on the work of Geoffrion [31] so that the nondominated surface can be traced out by varying λ_2 and λ_3, i.e., obtaining an efficient solution of (7.7) for each given $\lambda_2 \geq 0, \lambda_3 \geq 0$ without assuming utility functions.

Assume (2.1). A criterion vector $\bar{\mathbf{z}} \in Z$ is *weakly nondominated* if and only if there does not exist another $\mathbf{z} \in Z$ such that $z_i > \bar{z}_i$ for all $i \in K$, where $K = \{1, 2, \ldots, k\}$. Otherwise it is weakly dominated. Consequently, a point $\bar{\mathbf{x}} \in S$ is *weakly efficient* if its *image* criterion vector $\bar{\mathbf{z}} = (f_1(\bar{\mathbf{x}}), \ldots, f_k(\bar{\mathbf{x}}))$ is weakly nondominated. Otherwise it is weakly inefficient. As the name suggests, an efficient solution is weakly efficient, but not vice versa. For example, by taking $\mathbf{c}^1 = (1, 0)^T$ and $\mathbf{c}^2 = (0, 1)^T$, feasible regions S and Z are identical and represented by the shaded region in Figure 7.2. The efficient set contains only one point $\mathbf{x}^3$. The weakly efficient set contains linear segments from $\mathbf{x}^1$ to $\mathbf{x}^3$ including $\mathbf{x}^1$ and $\mathbf{x}^3$, and from $\mathbf{x}^3$ to $\mathbf{x}^2$ including $\mathbf{x}^3$ and $\mathbf{x}^2$.

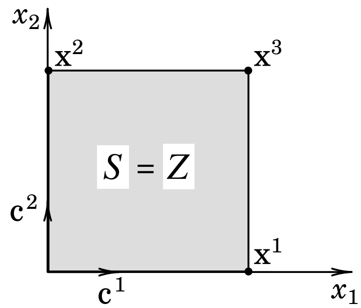

Figure 7.2: Example of weakly efficient points

Theorem 7.6. *Every weakly efficient point of* (7.7) *is also efficient.*

Proof. Want to prove that equivalently, every weakly nondominated point of (7.7) is also nondominated. Because Z is continuous by Theorem 7.2, any interior point of Z can not be weakly nondominated, i.e., all the weakly nondominated points are on the minimum-variance surface. Because the minimum-variance surface is a non-degenerate paraboloid by Theorem 7.5, the existence of weakly nondominated points which are not nondominated suggests linear segments along the direction of variance or expected return or dividend yield on the minimum-variance surface. But this possibility is eliminated by the non-existence of any linear segments by Lemma 7.3. Therefore, the only option is that every weakly nondominated point of (7.7) is also nondominated. □

To solve (7.45), apply Lagrangian method by constructing

$$L(\mathbf{x}, \lambda_4) = \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} - \lambda_2 \boldsymbol{\mu}^T \mathbf{x} - \lambda_3 \boldsymbol{\delta}^T \mathbf{x} + \lambda_4 (1 - \mathbf{1}^T \mathbf{x}) \tag{7.47}$$

where λ_4 is the only Lagrangian multiplier. The weights in (7.45) and multiplier in (7.47) are purposefully denoted by λ_2, λ_3 and λ_4 to be in the same format as (7.9) and for the formulae of general k objectives later. Because $\mathbf{x}^T\mathbf{\Sigma}\mathbf{x}$ is positive semidefinite with respect to $\mathbf{x}$, $\mathbf{x}$ is the minimizing solution of (7.45) if and only if

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{\Sigma}\mathbf{x} - \lambda_2\boldsymbol{\mu} - \lambda_3\boldsymbol{\delta} - \lambda_4\mathbf{1} = \mathbf{0} \tag{7.48}$$

$$\frac{\partial L}{\partial \lambda_4} = 1 - \mathbf{1}^T\mathbf{x} = 0 \tag{7.49}$$

Matrix $\mathbf{\Sigma}$ is invertible by Assumption 7.2, then (7.48) is

$$\mathbf{x} = \frac{1}{2}(\lambda_2\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + \lambda_3\mathbf{\Sigma}^{-1}\boldsymbol{\delta} + \lambda_4\mathbf{\Sigma}^{-1}\mathbf{1}) \tag{7.50}$$

Pre-multiply (7.50) by $\mathbf{1}^T$ and substitute in (7.49) to obtain

$$1 = \frac{1}{2}(\lambda_2\mathbf{1}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + \lambda_3\mathbf{1}^T\mathbf{\Sigma}^{-1}\boldsymbol{\delta} + \lambda_4\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1})$$

rearrange to get

$$\begin{aligned}\lambda_4 &= \frac{1}{\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1}}(2 - \lambda_2\mathbf{1}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \lambda_3\mathbf{1}^T\mathbf{\Sigma}^{-1}\boldsymbol{\delta})\text{, by (7.17)} \\ &= \frac{1}{f}(2 - \lambda_2 c - \lambda_3 e)\end{aligned} \tag{7.51}$$

Note that $f > 0$ by (7.18). Plug (7.51) into (7.50)

$$\begin{aligned}\mathbf{x} &= \frac{1}{2}[\lambda_2\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + \lambda_3\mathbf{\Sigma}^{-1}\boldsymbol{\delta} + \frac{1}{f}(2 - \lambda_2 c - \lambda_3 e)\mathbf{\Sigma}^{-1}\mathbf{1}] \\ &= \lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \mathbf{x}^1\end{aligned} \tag{7.52}$$

where

$$\mathbf{\Delta}^2 = \frac{1}{2}(\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\mathbf{\Sigma}^{-1}\mathbf{1}) \tag{7.53}$$

$$\mathbf{\Delta}^3 = \frac{1}{2}(\mathbf{\Sigma}^{-1}\boldsymbol{\delta} - \frac{e}{f}\mathbf{\Sigma}^{-1}\mathbf{1}) \tag{7.54}$$

$$\mathbf{x}^1 = \frac{1}{f}\mathbf{\Sigma}^{-1}\mathbf{1} \tag{7.55}$$

The same symbol $\mathbf{x}^1$ in (7.5) is deployed in (7.55), because they are identical. The following theorems describe the properties of $\mathbf{x}^1$, $\boldsymbol{\Delta}^2$ and $\boldsymbol{\Delta}^3$. The interpretation of $\mathbf{x}^1, \boldsymbol{\Delta}^2$, and $\boldsymbol{\Delta}^3$ will be given later.

Theorem 7.7. *The minimum-variance portfolio,* $\mathbf{x}^1$*, of* (7.7) *is also the minimum-variance portfolio of* (7.1)*.*

Proof. It follows from that the minimum-variance portfolio of (7.1) and (7.7) is the same $\mathbf{x}^1 = \frac{1}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1}$, (7.5) and (7.55). The variance= $\frac{1}{f}$, the expected return= $\frac{c}{f}$, the dividend yield= $\frac{e}{f}$. □

Lemma 7.5. *The direction (vector) of the portfolio weights of the minimum-variance frontier* (7.3)*,* $[\frac{1}{af-cc}(f\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - c\boldsymbol{\Sigma}^{-1}\mathbf{1})]$ *is parallel to* $\boldsymbol{\Delta}^2$*. The direction (vector) of the efficient set* (7.6)*,* $(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1})$ *is also parallel to* $\boldsymbol{\Delta}^2$*.*

Proof.

$$\frac{1}{af-cc}(f\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - c\boldsymbol{\Sigma}^{-1}\mathbf{1}) = \frac{2f}{af-cc}[\frac{1}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1})] = \frac{2f}{af-cc}\boldsymbol{\Delta}^2$$
$$\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1} = 2\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1} = 2\boldsymbol{\Delta}^2$$

where $f > 0$ by (7.18) and $af - cc > 0$ by (7.19). □

Theorem 7.8. *Vectors* $\boldsymbol{\Delta}^2$ *and* $\boldsymbol{\Delta}^3$ *are linearly independent.*

Proof. By (7.53) and (7.54),

$$\begin{aligned}\lambda_2\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\Delta}^3 &= \lambda_2\frac{1}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1}) + \lambda_3\frac{1}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \frac{e}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1})\\ &= \frac{\lambda_2}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\lambda_3}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \frac{c\lambda_2 + e\lambda_3}{2f}\boldsymbol{\Sigma}^{-1}\mathbf{1}\end{aligned}$$

Because $\mathbf{\Sigma}^{-1}\boldsymbol{\mu}, \mathbf{\Sigma}^{-1}\boldsymbol{\delta}$ and $\mathbf{\Sigma}^{-1}\mathbf{1}$ are linearly independent by Lemma 7.1, the necessary and sufficient condition of $\lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 = \mathbf{0}$ is

$$\begin{aligned} \frac{\lambda_2}{2} &= 0 \\ \frac{\lambda_3}{2} &= 0 \\ -\frac{c\lambda_2 + e\lambda_3}{2f} &= 0 \end{aligned}$$

The only solution of these three equations is $\lambda_2 = 0$ and $\lambda_3 = 0$, which implies the conclusion.

□

The interpretation of $\mathbf{x}^1, \mathbf{\Delta}^2$, and $\mathbf{\Delta}^3$ is as follows. Portfolio weight $\mathbf{x}^1$ is *minimum-variance portfolio* of both (7.7) and (7.1) as proved in Theorem 7.7. Similar to $\mathbf{d}^2$ (7.25) and $\mathbf{d}^3$ (7.26), $\mathbf{\Delta}^2$ and $\mathbf{\Delta}^3$ are two directions (vectors). Next introduce

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3, \lambda_2 \geq 0, \lambda_3 \geq 0\} \tag{7.56}$$

Because $\mathbf{\Delta}^2$ and $\mathbf{\Delta}^3$ are linearly independent as proved in Theorem 7.8, (7.56) is a two dimensional cone passing through the origin in decision space. Vectors $\mathbf{\Delta}^2$ and $\mathbf{\Delta}^3$ are generators (see Steuer [89]) of the cone. Next translate the cone by $\mathbf{x}^1$

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \mathbf{x}^1, \lambda_2 \geq 0, \lambda_3 \geq 0\} \tag{7.57}$$

Expression (7.57) is a translated cone. For given $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$, the minimizing solution of (7.45) and therefore an efficient solution of (7.7) is prescribed by (7.52). As both λ_2 and λ_3 change from 0 to ∞, the efficient set of (7.7) is the two dimensional translated cone (7.57).

First, $\mathbf{x}^1$ extends the minimum-variance portfolio of Huang and Litzenberger [43] in suitable-portfolio investor fashion. Second, the efficient set of model (7.7) (a two dimensional translated cone (7.57)) extends the efficient set of model (7.1) (a one dimensional translated cone (7.6)). These two cones have the common generator $\mathbf{\Delta}^2$. It will be proved in Theorem 7.9 that the efficient set of (7.1) is a subset of the efficient set of (7.7). The translated cone (7.57) is a convex set. Then any convex combination of any points belonging to the cone still belongs to the cone. The implication is that if all suitable-portfolio investors use (7.7) and hold efficient portfolios, then the market portfolio, as a convex combination of these efficient portfolios, is also efficient.

Theorem 7.9. *The efficient set* (7.6) *of model* (7.1) *is a subset of the efficient set* (7.57) *of model* (7.7).

Proof. It follows from the fact that (7.1) and (7.7) have the same minimum-variance portfolio in Theorem 7.7 and the same generator of the efficient sets in Lemma 7.5. □

Theorem 7.9 demonstrates that the efficient portfolios of model (7.1) are still efficient of model (7.7). Therefore a suitable-portfolio investor inherits all the efficient portfolios from a standard investor and enjoys greater flexibility to choose from the nondominated surface.

The nondominated surface of model (7.7) can be delineated by plugging the efficient solution (7.52) into model (7.7) as follows

$$z_1 = (\lambda_2 \boldsymbol{\Delta}^2 + \lambda_3 \boldsymbol{\Delta}^3 + \mathbf{x}^1)^T \boldsymbol{\Sigma} (\lambda_2 \boldsymbol{\Delta}^2 + \lambda_3 \boldsymbol{\Delta}^3 + \mathbf{x}^1) \tag{7.58}$$

$$z_2 = \boldsymbol{\mu}^T (\lambda_2 \boldsymbol{\Delta}^2 + \lambda_3 \boldsymbol{\Delta}^3 + \mathbf{x}^1) \tag{7.59}$$

$$z_3 = \boldsymbol{\delta}^T (\lambda_2 \boldsymbol{\Delta}^2 + \lambda_3 \boldsymbol{\Delta}^3 + \mathbf{x}^1) \tag{7.60}$$

7.5 Unify e-Constraint Approach & Weighted-Sums Approach

The similarity between (7.24) and (7.52) suggest that there are relationships between the e-constraint approach and the weighted-sums approach of (7.7). The e-constraint approach (7.8) and the weighted-sums approach (7.45) can be connected by

$$\boldsymbol{\mu}^T \mathbf{x} = z_2 \quad \text{and let} \quad z_2 = e_2$$

$$\boldsymbol{\delta}^T \mathbf{x} = z_3 \quad \text{and let} \quad z_3 = e_3$$

Pre-multiply $\mathbf{x}$ (7.52) by $\boldsymbol{\mu}^T$ and $\boldsymbol{\delta}^T$, respectively and plug the products into the two equations above.

$$\boldsymbol{\mu}^T\mathbf{x}^1 + \lambda_2\boldsymbol{\mu}^T\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\mu}^T\boldsymbol{\Delta}^3 = e_2 \tag{7.61}$$

$$\boldsymbol{\delta}^T\mathbf{x}^1 + \lambda_2\boldsymbol{\delta}^T\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\delta}^T\boldsymbol{\Delta}^3 = e_3 \tag{7.62}$$

Re-express (7.61) and (7.62) in matrix format

$$\begin{bmatrix} \boldsymbol{\mu}^T\boldsymbol{\Delta}^2 & \boldsymbol{\mu}^T\boldsymbol{\Delta}^3 \\ \boldsymbol{\delta}^T\boldsymbol{\Delta}^2 & \boldsymbol{\delta}^T\boldsymbol{\Delta}^3 \end{bmatrix}_{2\times 2} \begin{bmatrix} \lambda_2 \\ \lambda_3 \end{bmatrix}_{2\times 1} = \begin{bmatrix} e_2 - \boldsymbol{\mu}^T\mathbf{x}^1 \\ e_3 - \boldsymbol{\delta}^T\mathbf{x}^1 \end{bmatrix}_{2\times 1} \tag{7.63}$$

Denote

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\mu}^T\boldsymbol{\Delta}^2 & \boldsymbol{\mu}^T\boldsymbol{\Delta}^3 \\ \boldsymbol{\delta}^T\boldsymbol{\Delta}^2 & \boldsymbol{\delta}^T\boldsymbol{\Delta}^3 \end{bmatrix}_{2\times 2}$$

$$\begin{aligned}
|\mathbf{B}| &= (\boldsymbol{\mu}^T\boldsymbol{\Delta}^2)(\boldsymbol{\delta}^T\boldsymbol{\Delta}^3) - (\boldsymbol{\mu}^T\boldsymbol{\Delta}^3)(\boldsymbol{\delta}^T\boldsymbol{\Delta}^2) \quad \text{, by (7.53) and (7.54)} \\
&= \frac{1}{4}[(\boldsymbol{\mu}^T(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1}))(\boldsymbol{\delta}^T(\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \frac{e}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1})) \\
&\quad - (\boldsymbol{\mu}^T(\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \frac{e}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1}))(\boldsymbol{\delta}^T(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1}))] \\
&= \frac{1}{4}[(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})(\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \frac{e}{f}\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) \\
&\quad - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \frac{e}{f}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \frac{c}{f}\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})] \text{ , by (7.17)} \\
&= \frac{1}{4}[(a - \frac{c}{f}c)(d - \frac{e}{f}e) - (b - \frac{e}{f}c)(b - \frac{c}{f}e)] \\
&= \frac{1}{4}[(ad - \frac{ccd}{f} - \frac{aee}{f} + \frac{ccee}{ff}) - (bb - \frac{2bce}{f} + \frac{ccee}{ff})] \\
&= \frac{1}{4f}[adf + 2bce - aee - bbf - ccd] \quad \text{, by (7.21)} \\
&= \frac{1}{4f}|\mathbf{A}|
\end{aligned}$$

Because $|\mathbf{A}| \neq 0$ ($\mathbf{A}$ is positive definite) by Lemma 7.2, and $f > 0$ by (7.18), $|\mathbf{B}| \neq 0$ ($\mathbf{B}$ is invertible).

$$\mathbf{B}^{-1} = \frac{4f}{|\mathbf{A}|}\begin{bmatrix} \boldsymbol{\delta}^T\boldsymbol{\Delta}^3 & -\boldsymbol{\mu}^T\boldsymbol{\Delta}^3 \\ -\boldsymbol{\delta}^T\boldsymbol{\Delta}^2 & \boldsymbol{\mu}^T\boldsymbol{\Delta}^2 \end{bmatrix}$$

Equation (7.63) is re-expressed as $\mathbf{B}\begin{bmatrix}\lambda_2\\ \lambda_3\end{bmatrix} = \begin{bmatrix}e_2 - \boldsymbol{\mu}^T\mathbf{x}^1\\ e_3 - \boldsymbol{\delta}^T\mathbf{x}^1\end{bmatrix}$. Then

$$\begin{aligned}\begin{bmatrix}\lambda_2\\ \lambda_3\end{bmatrix} &= \mathbf{B}^{-1}\begin{bmatrix}e_2 - \boldsymbol{\mu}^T\mathbf{x}^1\\ e_3 - \boldsymbol{\delta}^T\mathbf{x}^1\end{bmatrix}\\ &= \frac{4f}{|\mathbf{A}|}\begin{bmatrix}\boldsymbol{\delta}^T\boldsymbol{\Delta}^3(e_2 - \boldsymbol{\mu}^T\mathbf{x}^1) - \boldsymbol{\mu}^T\boldsymbol{\Delta}^3(e_3 - \boldsymbol{\delta}^T\mathbf{x}^1)\\ \boldsymbol{\mu}^T\boldsymbol{\Delta}^2(e_3 - \boldsymbol{\delta}^T\mathbf{x}^1) - \boldsymbol{\delta}^T\boldsymbol{\Delta}^2(e_2 - \boldsymbol{\mu}^T\mathbf{x}^1)\end{bmatrix}_{2\times 1}\end{aligned} \tag{7.64}$$

By (7.63) and (7.64), the initial connection between the e-constraint approach (7.8) and the weighted-sums approach (7.45) is that there is a one-to-one correspondence between the pair of e_2 and e_3 and the pair of λ_2 and λ_3, although the condition that $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$ may not be satisfied.

As a step further, the following theorems verify the subset relationship between the efficient set of (7.7) and the portfolio weights of the minimum-variance surface of (7.7). These theorems demonstrate the quantitative relationship, as well.

Theorem 7.10. *The minimum variance portfolio* $\mathbf{x}^1$ (7.55) *belongs to the portfolio weights of the minimum-variance surface* (7.32) *as follows*

$$\mathbf{x}^1 = \frac{c}{f}\mathbf{d}^2 + \frac{e}{f}\mathbf{d}^3 + \mathbf{x}^0 \tag{7.65}$$

Proof. Want to prove that there exist e_2 and e_3 such that $\mathbf{x}^1 = e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \mathbf{x}^0$. Because $\mathbf{x}^1$ is obtained by setting $\lambda_2 = 0$ and $\lambda_3 = 0$, plug $\lambda_2 = 0$ and $\lambda_3 = 0$ into (7.61) and (7.62).

$$\begin{aligned}
e_2 &= \boldsymbol{\mu}^T\mathbf{x}^1 + \lambda_2\boldsymbol{\mu}^T\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\mu}^T\boldsymbol{\Delta}^3 \\
&= \boldsymbol{\mu}^T\mathbf{x}^1 + 0 \quad \text{, by (7.55)} \\
&= \frac{1}{f}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} \quad \text{, by (7.17)} \\
&= \frac{c}{f} \\
e_3 &= \boldsymbol{\delta}^T\mathbf{x}^1 + \lambda_2\boldsymbol{\delta}^T\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\delta}^T\boldsymbol{\Delta}^3 \\
&= \boldsymbol{\delta}^T\mathbf{x}^1 + 0 \quad \text{, by (7.55)} \\
&= \frac{1}{f}\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} \quad \text{, by (7.17)} \\
&= \frac{e}{f}
\end{aligned}$$

Plug $e_2 = \frac{c}{f}$ and $e_3 = \frac{e}{f}$ into (7.24)

$$\begin{aligned}
e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \mathbf{x}^0 &= \frac{1}{|\mathbf{A}|}[(e_2(df-ee) + e_3(ce-bf) + (be-cd))\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\quad + (e_2(ce-bf) + e_3(af-cc) + (bc-ae))\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} \\
&\quad + (e_2(be-cd) + e_3(bc-ae) + (ad-bb))\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\
&= \frac{1}{|\mathbf{A}|}[(\frac{c}{f}(df-ee) + \frac{e}{f}(ce-bf) + (be-cd))\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\quad + (\frac{c}{f}(ce-bf) + \frac{e}{f}(af-cc) + (bc-ae))\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} \\
&\quad + (\frac{c}{f}(be-cd) + \frac{e}{f}(bc-ae) + (ad-bb))\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\
&= \frac{1}{|\mathbf{A}|}[((cd - \frac{cee}{f}) + (\frac{cee}{f} - be) + (be-cd))\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\quad + ((\frac{cce}{f} - bc) + (ae - \frac{cce}{f}) + (bc-ae))\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} \\
&\quad + (\frac{bce-ccd}{f} + \frac{bce-aee}{f} + (ad-bb))\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\
&= \frac{1}{f|\mathbf{A}|}[\mathbf{0} + (adf + 2bce - aee - bbf - ccd))\boldsymbol{\Sigma}^{-1}\mathbf{1}]\text{, by (7.21)} \\
&= \frac{1}{f|\mathbf{A}|}[|\mathbf{A}|\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\
&= \frac{1}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1}
\end{aligned}$$

$\frac{1}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1}$ is exactly the expression of $\mathbf{x}^1$ (7.55). Therefore, $\mathbf{x}^1$ belongs to the portfolio weights of the minimum-variance surface. □

Theorem 7.11. *Vector* $\boldsymbol{\Delta}^2$ (7.53) *is a linear combination of* $\mathbf{d}^2$ *and* $\mathbf{d}^3$ *as follows*

$$\boldsymbol{\Delta}^2 = \frac{af-cc}{2f}\mathbf{d}^2 + \frac{bf-ce}{2f}\mathbf{d}^3 \tag{7.66}$$

Proof. Plug in the expression of $\mathbf{d}^2$ (7.25) and $\mathbf{d}^3$ (7.26).

$$\begin{aligned}
e_2\mathbf{d}^2 + e_3\mathbf{d}^3 &= \frac{e_2}{|\mathbf{A}|}[(df - ee)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (ce - bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (be - cd)\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\
&\quad + \frac{e_3}{|\mathbf{A}|}[(ce - bf)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (af - cc)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} + (bc - ae)\boldsymbol{\Sigma}^{-1}\mathbf{1}] \\
&= \frac{1}{|\mathbf{A}|}\{[(df - ee)e_2 + (ce - bf)e_3]\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\quad + [(ce - bf)e_2 + (af - cc)e_3]\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} \\
&\quad + [(be - cd)e_2 + (bc - ae)e_3]\boldsymbol{\Sigma}^{-1}\mathbf{1}\}
\end{aligned}$$

By (7.53),

$$\boldsymbol{\Delta}^2 = \frac{1}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1})$$

Because $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}\mathbf{1}$ are linearly independent by Lemma 7.1, the necessary and sufficient condition of $\boldsymbol{\Delta}^2 = e_2\mathbf{d}^2 + e_3\mathbf{d}^3$ is

$$\frac{1}{|\mathbf{A}|}[(df - ee)e_2 + (ce - bf)e_3] = \frac{1}{2} \tag{7.67}$$

$$\frac{1}{|\mathbf{A}|}[(ce - bf)e_2 + (af - cc)e_3] = 0 \tag{7.68}$$

$$\frac{1}{|\mathbf{A}|}[(be - cd)e_2 + (bc - ae)e_3] = -\frac{c}{2f} \tag{7.69}$$

By (7.68),

$$e_3 = \frac{bf - ce}{af - cc}e_2$$

Because $af - cc > 0$ by (7.19), $\frac{1}{af-cc}$ is well-defined. Plug this expression into (7.67).

$$\begin{aligned}
\frac{1}{|\mathbf{A}|}[(df - ee)e_2 + (ce - bf)\frac{bf - ce}{af - cc}e_2] &= \frac{1}{2} \\
[adff - aeef - ccdf + ccee - bbff - ccee + 2bcef]e_2 &= \frac{|\mathbf{A}|(af - cc)}{2} \\
f[adf + 2bce - aee - bbf - ccd]e_2 &= \frac{|\mathbf{A}|(af - cc)}{2}, \text{ by (7.21)} \\
f|\mathbf{A}|e_2 &= \frac{|\mathbf{A}|(af - cc)}{2} \\
e_2 &= \frac{af - cc}{2f}
\end{aligned}$$

$f > 0$ by (7.18), so $\frac{1}{f}$ is well-defined. Then

$$e_3 = \frac{bf - ce}{2f}$$

Plug the expressions of e_2 and e_3 into (7.69).

$$\begin{aligned}
&\frac{1}{|\mathbf{A}|}[(be - cd)e_2 + (bc - ae)e_3] \\
&= \frac{1}{|\mathbf{A}|}[(be - cd)\frac{(af - cc)}{2f} + (bc - ae)\frac{bf - ce}{2f}] \\
&= \frac{1}{2f|\mathbf{A}|}[abef - acdf - bcce + cccd + bbcf - abef - bcce + acee] \\
&= \frac{1}{2f|\mathbf{A}|}(-c)[adf + 2bce - aee - bbf - ccd], \text{ by (7.21)} \\
&= \frac{1}{2f|\mathbf{A}|}(-c)|\mathbf{A}| \\
&= -\frac{c}{2f}
\end{aligned}$$

This is exactly (7.69). To summarize, one computes a pair of e_2 and e_3 from (7.68) and (7.67), verifies the pair also satisfies (7.69), and obtains the combination. □

Theorem 7.12. *Vector* $\mathbf{\Delta}^3$ (7.54) *is a linear combination of* $\mathbf{d}^2$ *and* $\mathbf{d}^3$ *as follows*

$$\mathbf{\Delta}^3 = \frac{bf - ce}{2f}\mathbf{d}^2 + \frac{df - ee}{2f}\mathbf{d}^3 \tag{7.70}$$

Proof. Plug in the expression of $\mathbf{d}^2$ (7.25) and $\mathbf{d}^3$ (7.26).

$$\begin{aligned}
e_2\mathbf{d}^2 + e_3\mathbf{d}^3 &= \frac{e_2}{|\mathbf{A}|}[(df - ee)\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + (ce - bf)\mathbf{\Sigma}^{-1}\boldsymbol{\delta} + (be - cd)\mathbf{\Sigma}^{-1}\mathbf{1}] \\
&\quad + \frac{e_3}{|\mathbf{A}|}[(ce - bf)\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + (af - cc)\mathbf{\Sigma}^{-1}\boldsymbol{\delta} + (bc - ae)\mathbf{\Sigma}^{-1}\mathbf{1}] \\
&= \frac{1}{|\mathbf{A}|}\{[(df - ee)e_2 + (ce - bf)e_3]\mathbf{\Sigma}^{-1}\boldsymbol{\mu} \\
&\quad + [(ce - bf)e_2 + (af - cc)e_3]\mathbf{\Sigma}^{-1}\boldsymbol{\delta} \\
&\quad + [(be - cd)e_2 + (bc - ae)e_3]\mathbf{\Sigma}^{-1}\mathbf{1}\}
\end{aligned}$$

By (7.54),

$$\mathbf{\Delta}^3 = \frac{1}{2}(\mathbf{\Sigma}^{-1}\boldsymbol{\delta} - \frac{e}{f}\mathbf{\Sigma}^{-1}\mathbf{1})$$

Because $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}\mathbf{1}$ are linearly independent by Lemma 7.1, the necessary and sufficient condition of $\boldsymbol{\delta}^3 = e_2\mathbf{d}^2 + e_3\mathbf{d}^3$ is

$$\frac{1}{|\mathbf{A}|}[(df - ee)e_2 + (ce - bf)e_3] = 0 \tag{7.71}$$
$$\frac{1}{|\mathbf{A}|}[(ce - bf)e_2 + (af - cc)e_3] = \frac{1}{2} \tag{7.72}$$
$$\frac{1}{|\mathbf{A}|}[(be - cd)e_2 + (bc - ae)e_3] = -\frac{e}{2f} \tag{7.73}$$

By (7.71),

$$e_2 = \frac{bf - ce}{df - ee}e_3$$

Because $df - ee > 0$ by (7.19), $\frac{1}{df-ee}$ is well-defined. Plug this expression into (7.72).

$$\begin{aligned}
\frac{1}{|\mathbf{A}|}[(ce - bf)\frac{bf - ce}{df - ee}e_3 + (af - cc)e_3] &= \frac{1}{2} \\
[-bbff - ccee + 2bcef + adff - aeef - ccdf + ccee]e_3 &= \frac{(df - ee)|\mathbf{A}|}{2} \\
f[adf + 2bce - aee - bbf - ccd]e_3 &= \frac{(df - ee)|\mathbf{A}|}{2}, \text{ by (7.21)} \\
f|\mathbf{A}|e_3 &= \frac{(df - ee)|\mathbf{A}|}{2} \\
e_3 &= \frac{df - ee}{2f}
\end{aligned}$$

$f > 0$ by (7.18), so $\frac{1}{f}$ is well-defined. Then

$$e_2 = \frac{bf - ce}{2f}$$

Plug the expressions of e_2 and e_3 into (7.73).

$$
\begin{aligned}
&\frac{1}{|\mathbf{A}|}[(be-cd)e_2+(bc-ae)e_3] \\
&=\frac{1}{|\mathbf{A}|}[(be-cd)\frac{bf-ce}{2f}+(bc-ae)\frac{df-ee}{2f}] \\
&=\frac{1}{2f|\mathbf{A}|}[bbef-bcdf-bcee+ccde+bcdf-adef-bcee+aeee] \\
&=\frac{1}{2f|\mathbf{A}|}(-e)[adf+2bce-aee-bbf-ccd], \text{ by (7.21)} \\
&=\frac{-e}{2f|\mathbf{A}|}|\mathbf{A}| \\
&=\frac{-e}{2f} \qquad (7.74)
\end{aligned}
$$

This is exactly (7.73). To summarize, one computes a pair of e_2 and e_3 from (7.71) and (7.72), verifies the pair also satisfies (7.73), and obtains the combination. □

Corollary 7.3. *The efficient set* (7.57) *of model* (7.7) *is a subset of the portfolio weights* (7.32) *of the minimum-variance surface of model* (7.7)*. Therefore the nondominated surface is a subset of the minimum-variance surface.*

Proof. It follows from the fact that $\mathbf{x}^1$ belongs to the portfolio weights of the minimum-variance surface and $\boldsymbol{\Delta}^2$ and $\boldsymbol{\Delta}^3$ are linear combinations of $\mathbf{d}^2$ and $\mathbf{d}^3$ by Theorem 7.10, Theorem 7.11 and Theorem 7.12. □

Theorem 7.13. *The portfolio weights of the minimum-variance frontier* (7.3) *is a subset of the portfolio weights of the minimum-variance surface* (7.32)*, i.e.,*

$$
\begin{aligned}
&\{\mathbf{x}\in\mathbb{R}^n \mid \mathbf{x}=[\frac{1}{af-cc}(-c\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}+a\boldsymbol{\Sigma}^{-1}\mathbf{1})]+e_2[\frac{1}{af-cc}(f\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-c\boldsymbol{\Sigma}^{-1}\mathbf{1})], e_2\in\mathbb{R}\} \\
&\subseteq \{\mathbf{x}\in\mathbb{R}^n \mid \mathbf{x}=e_2\mathbf{d}^2+e_3\mathbf{d}^3+\mathbf{x}^0,\ e_2,e_3\in\mathbb{R}\}
\end{aligned}
$$

Proof. It follows from the fact that (a) models (7.1) and (7.7) have the same minimum-variance portfolio $\mathbf{x}^1$ of by Theorem 7.7, (b) $\mathbf{x}^1$ belongs to the portfolio weights of the minimum-variance frontier, (c) $\mathbf{x}^1$ also belongs to the inverse images (portfolio weights) of the minimum-variance surface by Theorem 7.10, (d) the direction (vector) of the portfolio weights of the minimum-variance frontier is parallel to $\mathbf{\Delta}^2$ by Lemma 7.5, (e) $\mathbf{\Delta}^2$ is a linear combination of $\mathbf{d}^2$ and $\mathbf{d}^3$ by Theorem 7.11, and (f) as follows

$$\begin{aligned}
&\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = [\frac{1}{af - cc}(-c\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + a\mathbf{\Sigma}^{-1}\mathbf{1})] + e_2[\frac{1}{af - cc}(f\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - c\mathbf{\Sigma}^{-1}\mathbf{1})], e_2 \in \mathbb{R}\} \\
&= \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{x}^1 + e_2[\frac{1}{af - cc}(f\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - c\mathbf{\Sigma}^{-1}\mathbf{1})], e_2 \in \mathbb{R}\} \\
&= \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{x}^1 + e_2\mathbf{\Delta}^2, e_2 \in \mathbb{R}\}
\end{aligned}$$

□

This theorem demonstrates that the portfolio weights of the minimum-variance frontier will be part of the portfolio weights of the minimum-variance surface. Therefore deploying (7.7) can keep the virtue of (7.1) and bring more flexibility by choosing from the minimum-variance surface.

The connection between the e-constraint approach and the weighted-sums approach is that for given $\lambda_2 \geq 0, \lambda_3 \geq 0$, a pair of e_2 and e_3 can be located by (7.61) and (7.62). However, the other way is not that straightforward. For given e_2 and e_3, a pair of λ_2 and λ_3 can be calculated by (7.64), but the condition that $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$ may not be satisfied. Next this other way is analyzed based on the following conditions.

1. If $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$, then an efficient solution of (7.7) is located by $\mathbf{x} = \lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \mathbf{x}^1$ (7.52).

2. If $\lambda_2 \leq 0$ and $\lambda_3 \geq 0$, then by (7.52)

$$\begin{aligned}
\mathbf{x} &= \lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \mathbf{x}^1 \\
&= \mathbf{x}^1 + (-\lambda_2)(-\mathbf{\Delta}^2) + \lambda_3\mathbf{\Delta}^3 \quad , -\lambda_2 \geq 0
\end{aligned}$$

Then, the $\mathbf{x}$ is a nonnegative combination of $-\mathbf{\Delta}^2$ and $\mathbf{\Delta}^3$. Thus, the $\mathbf{x}$ belongs to the cone generated by $-\mathbf{\Delta}^2$ and $\mathbf{\Delta}^3$ and then translated by $\mathbf{x}^1$. Note that (7.47)-(7.51) do not require $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$. That is, (7.47)-(7.51) hold for any λ_2 and λ_3. Therefore, the $\mathbf{x}$ is still a minimizing solution of (7.47). Next re-express (7.47)

$$L(\mathbf{x}, \lambda_4) = \mathbf{x}^T\mathbf{\Sigma}\mathbf{x} - \lambda_2\boldsymbol{\mu}^T\mathbf{x} - \lambda_3\boldsymbol{\delta}^T\mathbf{x} + \lambda_4(1 - \mathbf{1}^T\mathbf{x})$$

as

$$L(\mathbf{x}, \lambda_4) = \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} + (-\lambda_2)\boldsymbol{\mu}^T \mathbf{x} - \lambda_3 \boldsymbol{\delta}^T \mathbf{x} + \lambda_4(1 - \mathbf{1}^T \mathbf{x})$$

This is the Lagrangian function of the model as follows

$$\begin{aligned} \min \{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} + (-\lambda_2)\boldsymbol{\mu}^T \mathbf{x} - \lambda_3 \boldsymbol{\delta}^T \mathbf{x} = z'_w\} \quad & -\lambda_2 \geq 0, \lambda_3 \geq 0 \\ \text{s.t.} \quad \mathbf{1}^T \mathbf{x} = 1 & \end{aligned}$$

This is the weighted-sums of the model as follows

$$\begin{aligned} \min \{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} &= z_1 \ \text{variance}\} \\ \min \{\boldsymbol{\mu}^T \mathbf{x} &= z_2 \ \text{expected return}\} \\ \max \{\boldsymbol{\delta}^T \mathbf{x} &= z_3 \ \text{dividend yield}\} \\ \text{s.t.} \quad \mathbf{1}^T \mathbf{x} &= 1 \end{aligned} \tag{7.75}$$

Therefore, the $\mathbf{x}$ is an efficient solution of (7.75) which traces out the minimizing expected return-maximizing dividend yield portion of the minimum-variance surface. This is approximately the same method deployed by Markowitz and Todd [66] to obtain the lower portion of minimum-variance frontier (discussed in Section 4.3). The efficient set of (7.75) is the cone generated by $-\boldsymbol{\Delta}^2$ and $\boldsymbol{\Delta}^3$ and then translated by $\mathbf{x}^1$.

3. Similarly, if $\lambda_2 \leq 0$ and $\lambda_3 \leq 0$, then the $\mathbf{x}$ obtained by (7.52) is a nonnegative combination of $-\boldsymbol{\Delta}^2$ and $-\boldsymbol{\Delta}^3$. Moreover, the $\mathbf{x}$ is an efficient solution of the model as follows

$$\begin{aligned} \min \{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} &= z_1 \ \text{variance}\} \\ \min \{\boldsymbol{\mu}^T \mathbf{x} &= z_2 \ \text{expected return}\} \\ \min \{\boldsymbol{\delta}^T \mathbf{x} &= z_3 \ \text{dividend yield}\} \\ \text{s.t.} \quad \mathbf{1}^T \mathbf{x} &= 1 \end{aligned} \tag{7.76}$$

The efficient set of (7.76) is a cone generated by $-\mathbf{\Delta}^2$ and $-\mathbf{\Delta}^3$ and then translated by $\mathbf{x}^1$.

4. Similarly, if $\lambda_2 \geq 0$ and $\lambda_3 \leq 0$, then the $\mathbf{x}$ obtained by (7.52) is a nonnegative combination of $\mathbf{\Delta}^2$ and $-\mathbf{\Delta}^3$. Moreover the $\mathbf{x}$ is an efficient solution of the model as follows

$$\begin{aligned} \min \{\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} &= z_1 \ \text{ variance}\} \\ \max \{\boldsymbol{\mu}^T\mathbf{x} &= z_2 \ \text{ expected return}\} \\ \min \{\boldsymbol{\delta}^T\mathbf{x} &= z_3 \ \text{ dividend yield}\} \\ \text{s.t.} \quad \mathbf{1}^T\mathbf{x} &= 1 \end{aligned} \tag{7.77}$$

The efficient set of (7.77) is a cone generated by $\mathbf{\Delta}^2$ and $-\mathbf{\Delta}^3$ and then translated by $\mathbf{x}^1$.

These four cases are exhaustive and mutually exclusive except when $\lambda_2 = 0$ or $\lambda_3 = 0$. They divide the inverse images (portfolio weights) of the minimum-variance surface (7.32) into four parts as in Figure 7.3. The affine subspace is depicted as the whole graph in Figure 7.3. It is spanned by $\mathbf{d}^2$ and $\mathbf{d}^3$ and translated by $\mathbf{x}^0$. In the figure, $\mathbf{x}^0, \mathbf{d}^2$ and $\mathbf{d}^3$ are represented at the lower right corner. The minimum-variance portfolio $\mathbf{x}^1$ belongs to the portfolio weights of the minimum-variance surface. In the figure, $\mathbf{x}^1, \mathbf{\Delta}^2, \mathbf{\Delta}^3, -\mathbf{\Delta}^2$ and $-\mathbf{\Delta}^3$ are represented in the center. Furthermore, $\mathbf{x}^1, \mathbf{\Delta}^2, \mathbf{\Delta}^3, -\mathbf{\Delta}^2$ and $-\mathbf{\Delta}^3$ divide the portfolio weights of the minimum-variance surface into four translated cones marked by four different colors. $\mathbf{x}^1$ serves as the common extreme point (tip) of the cones. Each cone is the efficient set of one of the four models (7.7), (7.75), (7.76) and (7.77), determined by the condition of λ_2 and λ_3. Any point on the affine subspace belongs to one of the cones. The images of the four cones are the four portions of the minimum-variance surface.

This structure extends the simpler relationship between the portfolio weights of minimum-variance frontier (7.3) of model (7.1) and the efficient set (7.6) of model (7.1). The simpler

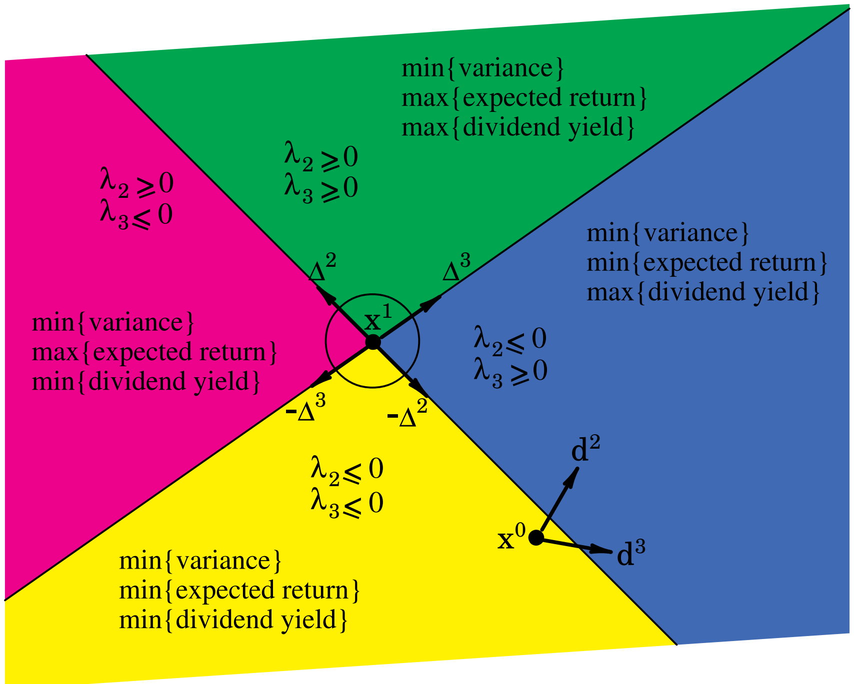

Figure 7.3: Relationship between the e-constraint approach and the weighted-sums approach

relationship is that both the portfolio weights of minimum-variance frontier (7.3) and the efficient set (7.6) are spanned by the same direction $(f\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{c}{f}\boldsymbol{\Sigma}^{-1}\mathbf{1})$, and are one dimensional.

7.6 Back to Risk-return Space by Projection

Because a suitable-portfolio investor takes standard portfolio selection as a projection from high-dimensional space to (variance, expected return) space, it is interesting to project the minimum-variance surface (7.34) of model (7.7) and the nondominated surface along the direction of dividend yield to (variance, expected return) space. First, introduce a general result about projecting a feasible region Z.

Theorem 7.14. *Suppose the two models below have the same feasible region S in decision space.*

$$\begin{aligned} &\max\{f_1(\mathbf{x}) = z_1\} \\ &\max\{f_2(\mathbf{x}) = z_2\} \\ &s.t. \quad \mathbf{x} \in S \end{aligned} \tag{7.78}$$

$$\begin{aligned} &\max\{f_1(\mathbf{x}) = z_1\} \\ &\max\{f_2(\mathbf{x}) = z_2\} \\ &\vdots \\ &\max\{f_k(\mathbf{x}) = z_k\} \\ &s.t. \quad \mathbf{x} \in S \end{aligned} \tag{7.79}$$

Denote the feasible region in criterion space of (7.78) *as Z^2. Denote the feasible region in criterion space of* (7.79) *as Z^k. Projection along the axes of $z_3, z_4, \ldots, z_k$ to z_1-z_2 space is performed in Z^k and denoted as $Proj\,(z_1, z_2, \ldots, z_k) = (z_1, z_2)$. Denote the projection of Z^k as $Proj\ Z^k = \{(z_1, z_2) \mid (z_1, z_2, \ldots, z_k) \in Z^k\}$. Then $Proj\ Z^k = Z^2$.*

Proof. Want $Proj\ Z^k \subseteq Z^2$. For any $(z_1, z_2, \ldots, z_k) \in Z^k$, there exists $\mathbf{x} \in S$ such that $(z_1, z_2, \ldots, z_k)$ and (z_1, z_2) are the images (criterion vectors) of $\mathbf{x}$ of (7.79) and (7.78), respectively. Then, $Proj\ (z_1, z_2, \ldots, z_k) = (z_1, z_2)$ of (7.79) $= (z_1, z_2)$ of (7.78). Then, $Proj\ (z_1, z_2, \ldots, z_k) \in Z^2$. Therefore, $Proj\ Z^k \subseteq Z^2$.

Want $Z^2 \subseteq Proj\ Z^k$. For any $(z_1, z_2) \in Z^2$, there exists $\mathbf{x} \in S$ such that (z_1, z_2) and $(z_1, z_2, \ldots, z_k)$ are the images (criterion vectors) of $\mathbf{x}$ of (7.78) and (7.79), respectively.Then, (z_1, z_2) of (7.78) $= (z_1, z_2)$ of (7.79) $= Proj\,(z_1, z_2, \ldots, z_k)$. Then, $(z_1, z_2) \in Proj\ Z^k$. Therefore $Z^2 \subseteq Proj\ Z^k$.

Because $Proj\ Z^k \subseteq Z^2$ and $Z^2 \subseteq Proj\ Z^k$, $Proj\ Z^k = Z^2$. □

Theorem 7.14 demonstrates that the projection of a suitable-portfolio investor's Z to (variance, expected return) space is exactly a standard investor's Z. The following theorem describes the projection of the minimum-variance surface (7.34).

Theorem 7.15. *The boundary of the projection of the minimum-variance surface* (7.34) *of model* (7.7) *on* (*variance, expected return*) *space is the minimum-variance frontier of model* (7.1)*.*

Proof. It follows from the facts that (a) the projection of the minimum-variance surface is the projection of the Z of (7.7) by Theorem 7.5 and Corollary 7.2, (b) the projection of the Z of model (7.7) is the Z of model (7.1) by Theorem 7.14, (c) the minimum-variance frontier is the boundary of the Z of (7.1), and (d) the portfolio weights of the minimum-variance frontier is a subset of the portfolio weights of the minimum-variance surface by Theorem 7.13. □

Theorem 7.15 demonstrates that a suitable-portfolio investor's model (7.7) can keep the virtue of a standard investor's model (7.1) and bring more flexibility by choosing from the minimum-variance surface.

For general models like (7.78) and (7.79), there are no conclusive results about the relationship between their efficient sets. A general trend is that an efficient set can grow as the number of objectives increases. However, see Steuer [89], p. 179 for counter-examples. Models (7.1) and (7.7) fit the general trend. The following theorem describes the projection of the nondominated surface.

Theorem 7.16. *The nondominated frontier of model* (7.1) *is the "minimizing variance-maximizing expected return" boundary of the projection of the nondominated surface of model* (7.7)*.*

Proof. The projection of the nondominated surface is a subset of the projection of the Z of (7.7) and therefore a subset of the Z of (7.1) by Theorem 7.14. Therefore, the projection is dominated by the nondominated frontier of (7.1). Because the efficient set of model (7.1) is a subset of the efficient set of (7.7) by Theorem 7.9, some points on the nondominated surface of (7.7) are projected as the nondominated frontier. □

Corollary 7.4. *The minimum-variance portfolio,* $\mathbf{x}^1$ *of* (7.7) *is projected as the minimum-variance portfolio of* (7.1).

Proof. It follows from the fact that the variance and expected return of the minimum-variance portfolio of (7.7) match those of the minimum-variance portfolio of (7.1) in the proof of Theorem 7.7. □

Next introduce the concept of *projection boundary curve*. It is a curve on the minimum-variance surface (7.34) whose projection forms the boundary of the projection of the minimum-variance surface.

The minimum-variance surface, (7.34) is

$$z_1 = \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 z_2^2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_3^2 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 z_3 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_2 + 2\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_3 + \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0$$

Re-express the surface as $F(z_1, z_2, z_3) = 0$, where

$$F(z_1, z_2, z_3) = -z_1 + \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 z_2^2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_3^2 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 z_3 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_2 + 2\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_3 + \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0$$

Because the surface is a non-degenerate paraboloid and the projection is executed in the direction of z_3 (dividend yield), $(0, 0, 1)^T$, the normal of projection boundary curve is orthogonal to $(0, 0, 1)^T$, which means $\frac{\partial F(z_1,z_2,z_3)}{\partial z_3} = 0$, i.e.,

$$0 + 0 + 2\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_3 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 + 0 + 2\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0 + 0 = 0$$

Then $z_3 = -\frac{\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3}$. Plug this equation of z_3 back to the expression of the minimum-variance surface (7.34) to get projection boundary curve.

$$
\begin{aligned}
z_1 &= \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 z_2^2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3\Big(-\frac{\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3}\Big)^2 + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2\Big(-\frac{\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3}\Big) + 2\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_2 \\
&\quad + 2\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0\Big(-\frac{\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3}\Big) + \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0 \\
&= \frac{(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3) - (\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3)}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3} z_2^2 + 2\frac{(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3) - (\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0)}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3} z_2 \\
&\quad + \frac{(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0) - (\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0)}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3} \qquad (7.80)
\end{aligned}
$$

$$
z_3 = -\frac{\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 z_2 + \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3} \qquad (7.81)
$$

(7.80) and (7.81) delineate the projection boundary curve. The projection of this curve to (variance, expected return) space is

$$
\begin{aligned}
z_1 = &\frac{(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3) - (\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3)}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3} z_2^2 + 2\frac{(\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3) - (\mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0)}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3} z_2 \\
&+ \frac{(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3)(\mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0) - (\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0)(\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0)}{\mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3}
\end{aligned}
$$

By Theorem 7.15, this projection is the minimum-variance frontier (7.4).

7.7 Illustrations

A suitable-portfolio investor has the following information for four securities ($n = 4$) and applies model (7.7)

$$
\boldsymbol{\mu} = \begin{bmatrix} 0.0455 \\ 0.0796 \\ 0.0588 \\ 0.0391 \end{bmatrix} \qquad \boldsymbol{\delta} = \begin{bmatrix} 0.0353 \\ 0.0113 \\ 0.0454 \\ 0.0445 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 0.0078 & 0.0039 & 0.0052 & 0.0010 \\ 0.0039 & 0.0130 & 0.0019 & 0.0009 \\ 0.0052 & 0.0019 & 0.0081 & 0.0015 \\ 0.0010 & 0.0009 & 0.0015 & 0.0044 \end{bmatrix}
$$

First compute (7.17)

$$a = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \qquad b = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} \qquad c = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$$

$$= 0.9371 \qquad = 0.5459 \qquad = 15.7893$$

$$d = \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} \qquad e = \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \qquad f = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$$

$$= 0.5835 \qquad = 12.8834 \qquad = 330.4362$$

Then by (7.16)

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} = \begin{bmatrix} 0.9371 & 0.5459 & 15.7893 \\ 0.5459 & 0.5835 & 12.8834 \\ 15.7893 & 12.8834 & 330.4362 \end{bmatrix}$$

$$|\mathbf{A}| = 3.2931$$

Then by (7.25)-(7.27)

$$\mathbf{d}^2 = \begin{bmatrix} -35.7890 \\ 10.9959 \\ 39.7826 \\ -14.9895 \end{bmatrix} \qquad \mathbf{d}^3 = \begin{bmatrix} -35.3701 \\ -18.9515 \\ 50.4519 \\ 3.8696 \end{bmatrix} \qquad \mathbf{x}^0 = \begin{bmatrix} 3.2525 \\ 0.3378 \\ -3.7359 \\ 1.1456 \end{bmatrix}$$

Then by 7.32, the inverse images of the minimum-variance surface is

$$\{\mathbf{x} \in \mathbb{R}^4 \mid \mathbf{x} = e_2 \mathbf{d}^2 + e_3 \mathbf{d}^3 + \mathbf{x}^0,\ e_2, e_3 \in \mathbb{R}\}$$

Then by (7.35)

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^2 & \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 & \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0 \\ \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{d}^3 & \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{d}^3 & \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0 \\ \mathbf{d}^{2^T}\boldsymbol{\Sigma}\mathbf{x}^0 & \mathbf{d}^{3^T}\boldsymbol{\Sigma}\mathbf{x}^0 & \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0 \end{bmatrix} = \begin{bmatrix} 8.1433 & 6.9929 & -0.6618 \\ 6.9929 & 18.3272 & -1.0487 \\ -0.6618 & -1.0487 & 0.0755 \end{bmatrix}$$

Then by (7.34), the minimum-variance surface is

$$z_1 = 8.1433z_2^2 + 18.3272z_3^2 + 13.9859z_2z_3 - 1.3235z_2 - 2.0974z_3 + 0.0755$$

By (7.53)-(7.55), the directions $\boldsymbol{\Delta}^2$ and $\boldsymbol{\Delta}^3$, the minimum-variance portfolio $\mathbf{x}^1$, and the image (criterion vector) of $\mathbf{x}^1$ are

$$\boldsymbol{\Delta}^2 = \begin{bmatrix} -2.0359 \\ 1.6645 \\ 1.8751 \\ -1.5037 \end{bmatrix} \quad \boldsymbol{\Delta}^3 = \begin{bmatrix} -0.1881 \\ -1.1522 \\ 0.6610 \\ 0.6793 \end{bmatrix} \quad \mathbf{x}^1 = \begin{bmatrix} 0.1634 \\ 0.1243 \\ 0.1321 \\ 0.5802 \end{bmatrix} \quad \mathbf{z}^1 = \begin{bmatrix} 0.0030 \\ 0.0478 \\ 0.0390 \end{bmatrix}$$

By (7.57), the efficient set of (7.7) is

$$\{\mathbf{x} \in \mathbb{R}^4 \mid \mathbf{x} = \lambda_2\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\Delta}^3 + \mathbf{x}^1, \lambda_2 \geq 0, \lambda_3 \geq 0\}$$

The connection between the e-constraint approach (7.8) and the weighted-sums approach (7.45) is (7.63) and (7.64) as follows

$$\begin{bmatrix} 0.0913 & -0.0348 \\ -0.0348 & 0.0406 \end{bmatrix} \begin{bmatrix} \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} e_2 - 0.0478 \\ e_3 - 0.0390 \end{bmatrix}$$

$$\begin{bmatrix} 16.2866 & 13.9859 \\ 13.9859 & 36.6544 \end{bmatrix} \begin{bmatrix} e_2 - 0.0478 \\ e_3 - 0.0390 \end{bmatrix} = \begin{bmatrix} \lambda_2 \\ \lambda_3 \end{bmatrix}$$

The minimum-variance surface is in Figure 7.4 from different viewing angles (tilt and rotate in SAS) [6] in (variance, expected return, and dividend yield) space. A rectangular grid [7] on (expected return, dividend yield) space is needed to create the graphs by SAS. The rectangular grid is centered at $\mathbf{z}^1$ (the image of the minimum-variance portfolio), with $z_2^1 = 0.0478, z_3^1 = 0.0390$ and with a range of $4 \times 0.0478, 4 \times 0.0390$ in the order of expected return and dividend yield. The surface is a paraboloid. The top the surface is painted black. The bottom of the surface is painted grey (the true color is red in color prints). Figure 7.4 *top* is from the default view angle (tilt=70 degrees, rotate=70 degrees). Figure 7.4 *middle* is still in the 3-dimensional space by visualizing along (horizontal to) the $z_1 = 0$ plane (tilt=90 degrees, rotate=70 degrees). Figure 7.4 *bottom* is back to two dimensional (variance, expected return) space by projecting the surface along the axis of dividend yield (tilt=90 degrees, rotate=90 degrees). The boundary of the projection is represented by a thick line in graph. This thick line is the minimum-variance frontier of model (7.1) by Theorem 7.15. The image of the minimum-variance portfolio is not conspicuous in Figure 7.4 *top* and *middle*, but it is noticeable and marked as $\mathbf{z}^1$ in Figure 7.4 *bottom*.

Figure 7.5 is the same as Figure 7.4 except in (standard deviation, expected return, and dividend yield) space. The minimum-variance surface is a hyperboloid.

The nondominated surface in (variance, expected return, and dividend yield) space is in Figure 7.6 *top*. The rectangular grid on (expected return, dividend yield) space is starting from the image of minimum-variance portfolio, $z_2^1 = 0.0478, z_3^1 = 0.0390$ and with a range of $2 \times 0.0478, 2 \times 0.0390$ in the order of expected return and dividend yield. This surface is part of the minimum-variance surface. The projection of the nondominated surface onto (variance, expected return) space is in Figure 7.6 *middle*. A boundary of the projection is represented by the gray thick line in graph. This thick line is the nondominated frontier of

[6] "Tilt specifies one or more angles at which to tilt the picture about the y axis in z-x-y space... Rotate specifies one or more angles at which to rotate the picture about the z axis in z-x-y space", SAS [1], p. 404

[7] "The x, y data should form a rectangular grid with a z-value present for every possible (x, y) combination", SAS [1], p. 409

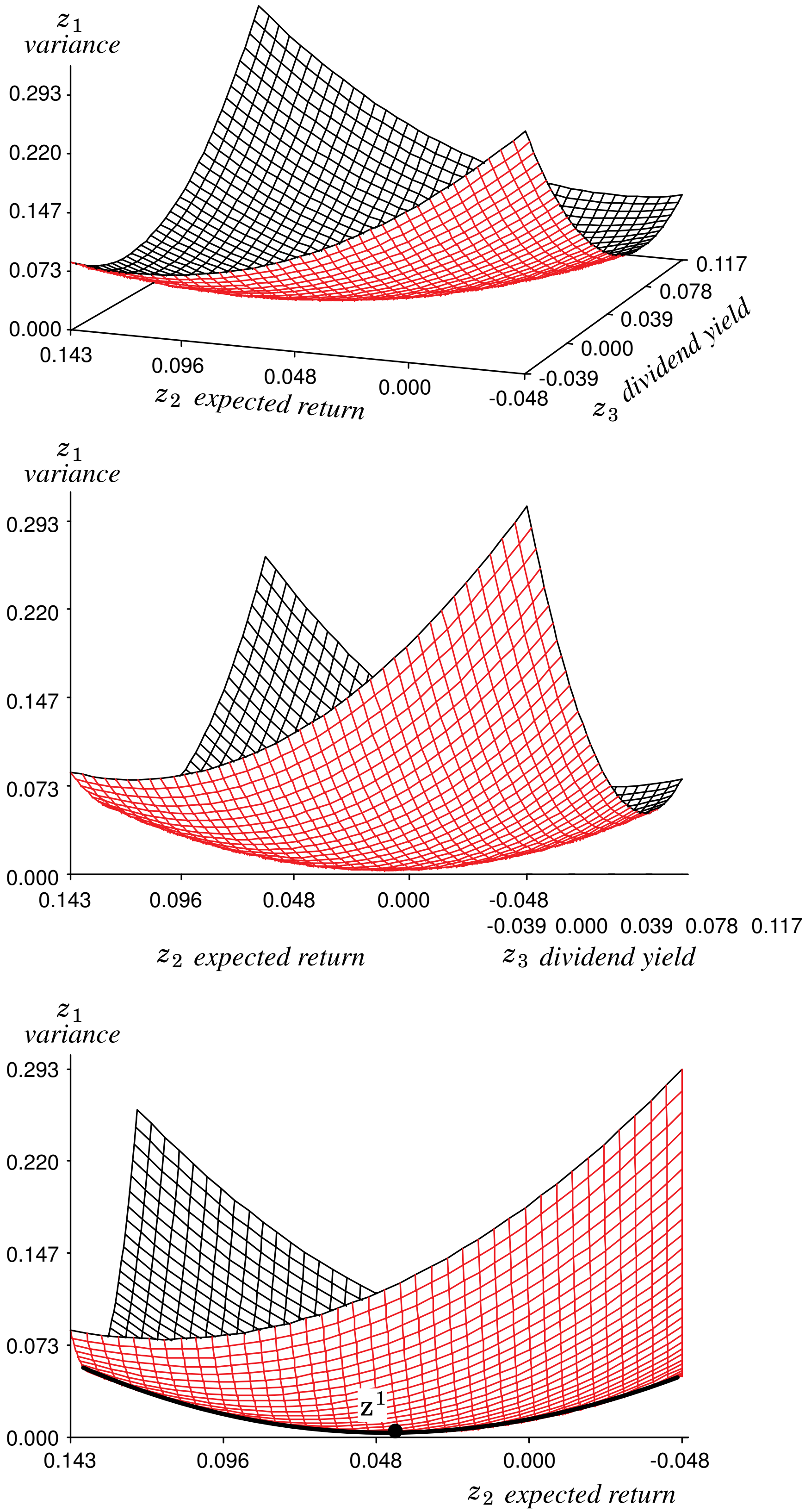

Figure 7.4: The minimum-variance surface (paraboloid) from different viewing angles and the projection onto (variance, expected return) space

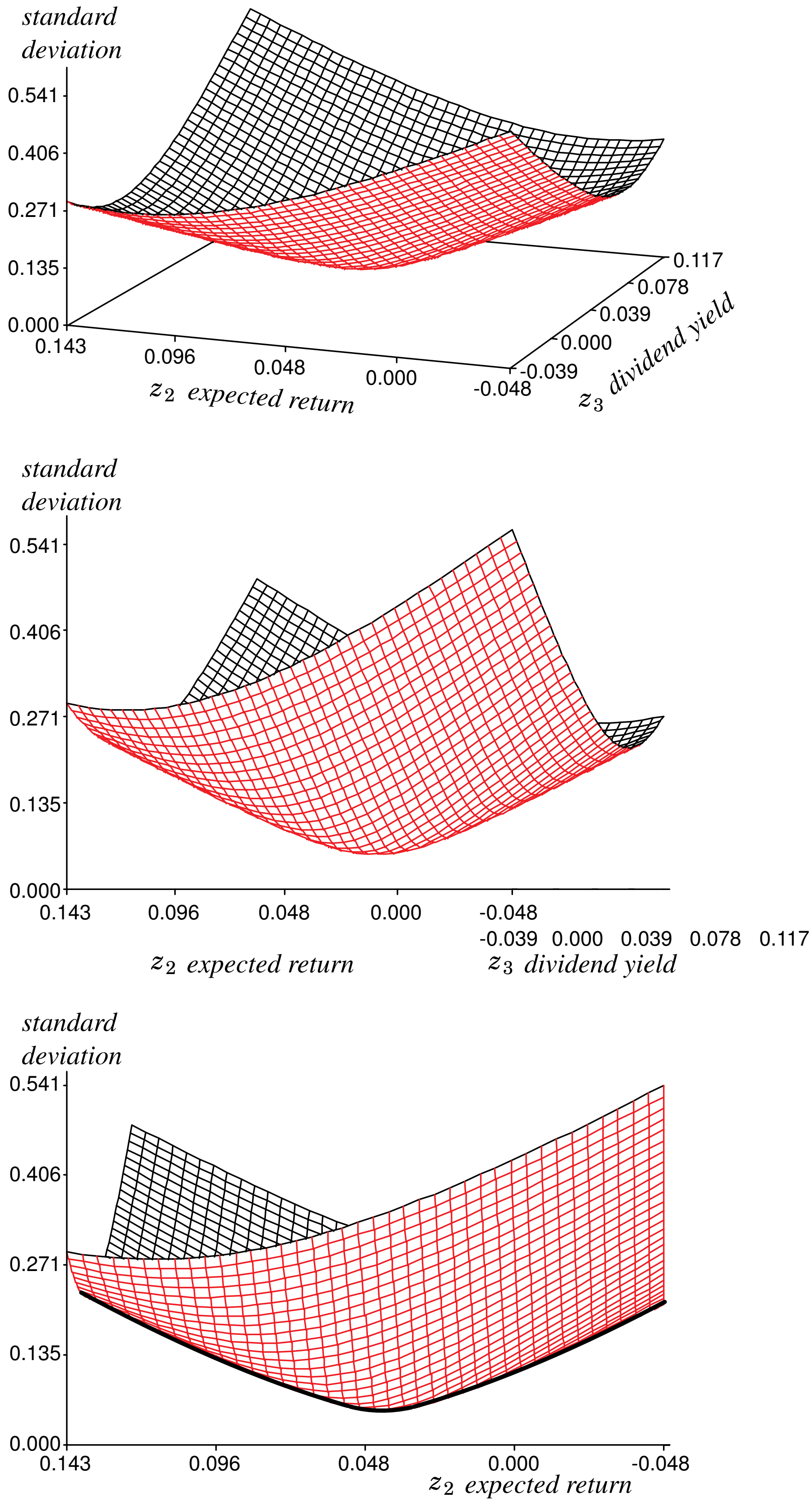

Figure 7.5: The minimum-variance surface (hyperboloid) from different viewing angles and the projection onto (standard deviation, expected return) space

model (7.1) by Theorem 7.16. This nondominated surface in (standard deviation, expected return, and dividend yield) space is in Figure 7.6 *bottom*.

The indifference curve of the minimum-variance surface with variance= 0.0500 is in Figure 7.7. The curve is an ellipse. The ellipse is rotated. That is, its major axis or minor axis are not parallel to either of the coordinate axes, because the coefficient of $z_2 z_3$, $\mathbf{d}^{2^T}\mathbf{\Sigma}\mathbf{d}^3 = 13.9859$ is not 0 in the expression of the minimum-variance surface.

7.8 Adding a Risk-free Asset and Implication to CAPM

Model (7.7) can be extended by adding a risk-free asset which has zero variance, a deterministic return r_f and zero dividend yield as follows

$$\begin{aligned} \min\{ \qquad\qquad\qquad \mathbf{x}^T\mathbf{\Sigma}\mathbf{x} &= z_1 \text{ variance}\} \\ \max\{\boldsymbol{\mu}^T\mathbf{x} + r_f(1-\mathbf{1}^T\mathbf{x}) &= z_2 \text{ expected return}\} \\ \max\{ \qquad\qquad\qquad \boldsymbol{\delta}^T\mathbf{x} &= z_3 \text{ dividend yield}\} \end{aligned} \tag{7.82}$$

where $\mathbf{x} \in \mathbb{R}^n$ and $1 - \mathbf{1}^T\mathbf{x}$ denote the portfolio weight in the n risky assets and the risk-free asset. The constraint $\mathbf{1}^T\mathbf{x} = 1$ drops out. Therefore (7.82) has no constraints. The objective function $\boldsymbol{\mu}^T\mathbf{x} + r_f(1-\mathbf{1}^T\mathbf{x}) = z_2$ can be simplified as $(\boldsymbol{\mu} - r_f\mathbf{1})^T\mathbf{x} + r_f = z_2$ Similar to (7.45), a weighted-sums approach can be applied to (7.82) as follows

$$\min\{\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} - \lambda_2((\boldsymbol{\mu} - r_f\mathbf{1})^T\mathbf{x} + r_f) - \lambda_3\boldsymbol{\delta}^T\mathbf{x} = z_w\} \quad \lambda_2 \geq 0, \lambda_3 \geq 0 \tag{7.83}$$

Model (7.83) will locate all the weakly efficient solutions of (7.82) as both λ_2 and λ_3 change continuously in $[0, \infty)$, and most of the efficient solutions as both λ_2 and λ_3 change continuously in $(0, \infty)$. Model (7.83) can be solved by taking partial derivatives with respect to $\mathbf{x}$ as follows

$$\begin{aligned} 2\mathbf{\Sigma}\mathbf{x} - \lambda_2(\boldsymbol{\mu} - r_f\mathbf{1}) - \lambda_3\boldsymbol{\delta} &= \mathbf{0} \\ \mathbf{x} &= \frac{1}{2}\mathbf{\Sigma}^{-1}[\lambda_2(\boldsymbol{\mu} - r_f\mathbf{1}) + \lambda_3\boldsymbol{\delta}] \end{aligned} \tag{7.84}$$

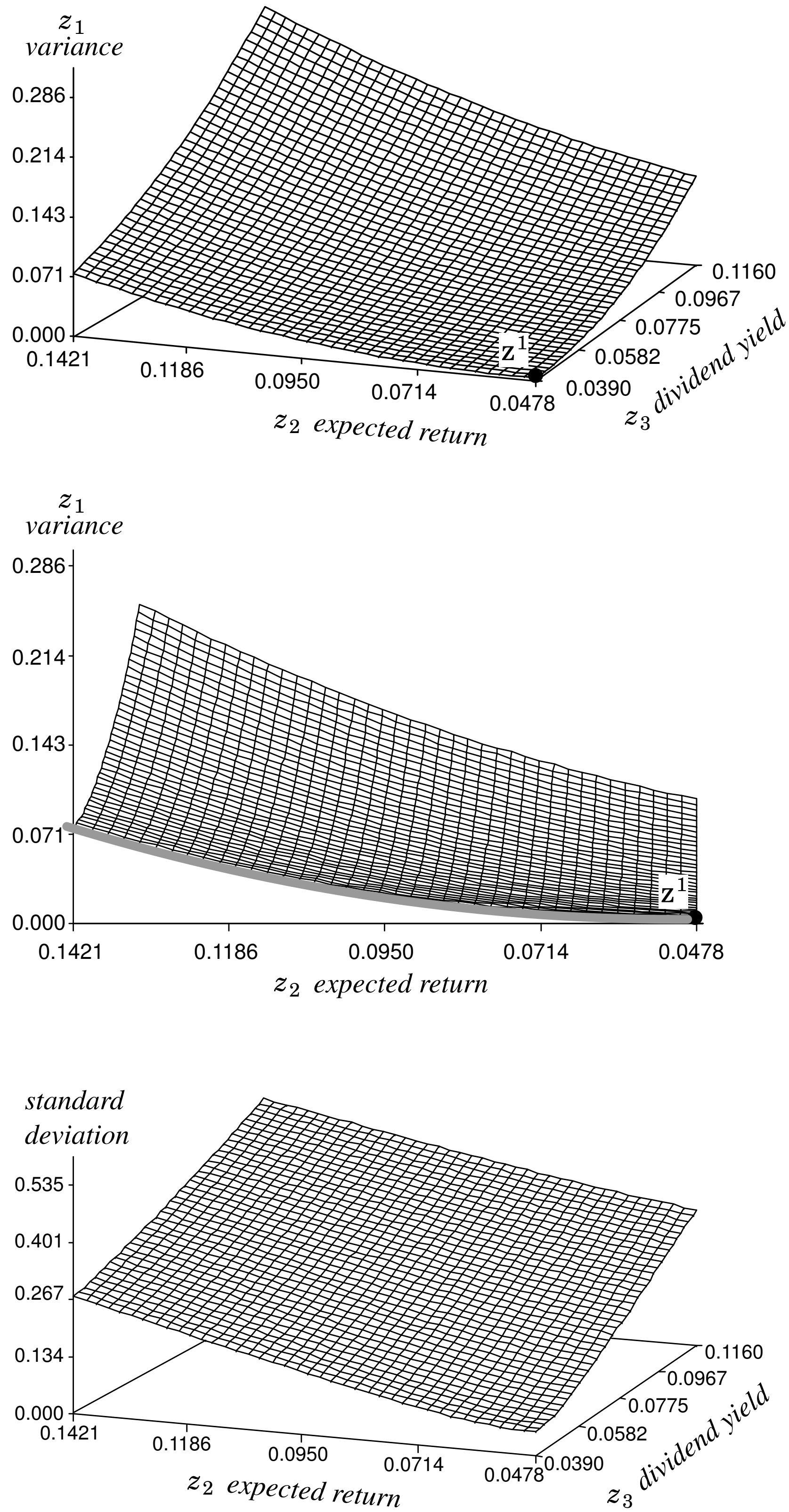

Figure 7.6: The Nondominated surface and the projection onto (variance, expected return) space

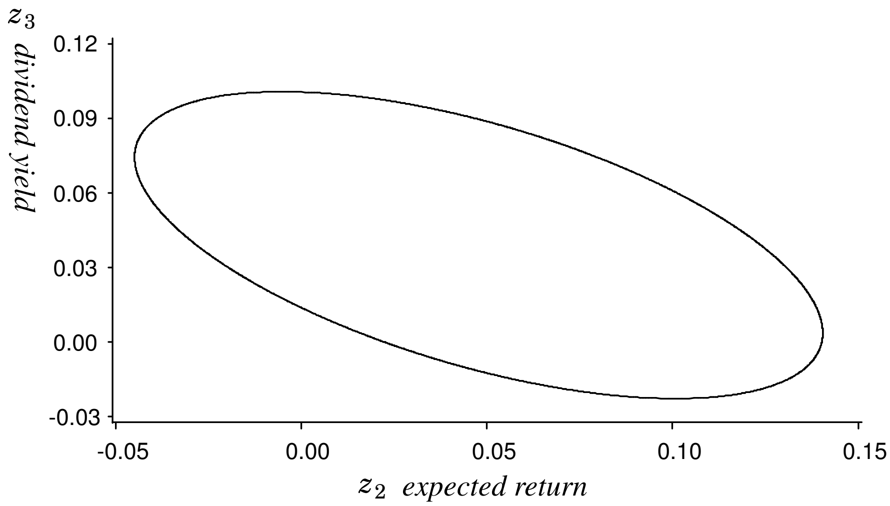

Figure 7.7: An indifference curve of the minimum-variance surface with variance= 0.0500

Merton [70] proved that if a risk-free asset with zero variance and deterministic return r_f exists and $r_f < \frac{c}{f}$, then there is a straight line (capital allocation line) passing through $(0, r_f)$ and tangent to the nondominated frontier of (7.1) in (standard deviation, expected return) space. Part of this straight line serves as the new nondominated frontier. If this were true in (standard deviation, expected return, and dividend yield) space of (7.82), a suitable-portfolio investor would observe a linear expression of dividend yield of the nondominated surface of (7.82), for example, $z_3 = a_0 + a_1 t, t \geq 0$ where a_0 and a_1 are fixed coefficients. However, pre-multiply (7.84) by $\boldsymbol{\delta}$ as

$$\begin{aligned} z_3 &= \boldsymbol{\delta}^T \frac{1}{2} \boldsymbol{\Sigma}^{-1} [\lambda_2(\boldsymbol{\mu} - r_f \mathbf{1}) + \lambda_3 \boldsymbol{\delta}] \\ &= \frac{1}{2} [\lambda_2 \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - r_f \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} + \lambda_3 \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}] \quad \text{, by (7.17)} \\ &= \frac{1}{2} [-r_f e + \lambda_2 b + \lambda_3 d] \, , \lambda_2 \geq 0, \lambda_3 \geq 0 \end{aligned}$$

This expression is not for a straight line, because besides λ_2, λ_3 also works as a parameter and λ_3's coefficient $d > 0$ by (7.18). This suggests that instead of only one tangency portfolio for (7.1), there can be a whole set of tangency portfolios for (7.7). Passing through each tangency portfolio and starting from the risk-free asset $(0, r_f, 0)$ can be a capital allocation line in (standard deviation, expected return, and dividend yield) space. These lines form a surface which is the nondominated surface of (7.82). This raises the question that among the set of tangency portfolios how a suitable-portfolio investor picks the one which can serve as the market portfolio and play an important role in extended Capital Asset Pricing Model.

One way is that a suitable-portfolio investor selects one tangency portfolio such that the corresponding capital allocation line has the greatest angle between the axis of standard deviation among all the capital allocation lines on the nondominated surface in (standard deviation, expected return, and dividend yield) space. This angle is the extended reward-to-variability ratio by Bodie, Kane and Marcus [10], p. 189 of a standard investor. The derivation of the tangency portfolio by Merton [70] is an instance of this way in in standard deviation-expected return space. Future research will be pursued to realize this way, check whether closed-form results can be obtained, and try to extend Sharpe's Capital Asset Pricing Model.

An alternative of (7.82) is to add two risk-free assets $(0, r_{fr}, 0)$ and $(0, 0, r_{fd})$, i.e., return risk-free asset and dividend yield risk-free asset. This can leave it questionable that these two risk-free assets can be taken as only one risk-free asset, because risk-free dividend yield can be treated as expected return. However, the formulation is a three objective model of variance, expected return and one other linear objective. This linear objective can be, for example, social responsibility instead of dividend yield. A suitable-portfolio investor will not add social responsibility to expected return directly. Then it is admissible to have two risk-free assets. This alternative is as follows

$$\begin{aligned} \min \{ \quad & \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} = z_1 \ \text{ variance}\} \\ \max \{ \quad & \boldsymbol{\mu}^T \mathbf{x} + r_{fr} x_{n+1} = z_2 \ \text{ expected return}\} \\ \max \{ & \boldsymbol{\delta}^T \mathbf{x} + r_{fd}(1 - \mathbf{1}^T \mathbf{x} - x_{n+1}) = z_3 \ \text{ dividend yield}\} \end{aligned} \tag{7.85}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n), x_{n+1}$ and $1 - \mathbf{1}^T \mathbf{x} - x_{n+1}$ denote the portfolio weight in the n risky assets, the return risk-free asset and the dividend yield risk-free asset, respectively. Similar to (7.45), a weighted-sums approach can be applied to (7.85) as follows

$$\min\left\{\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} - \lambda_2(\boldsymbol{\mu}^T\mathbf{x} + r_{fr}x_{n+1}) - \lambda_3[\boldsymbol{\delta}^T\mathbf{x} + r_{fd}(1 - \mathbf{1}^T\mathbf{x} - x_{n+1})] = z_w\right\} \quad \lambda_2 \geq 0, \lambda_3 \geq 0 \tag{7.86}$$

Model (7.86) will locate all the weakly efficient solutions of (7.85) as both λ_2 and λ_3 change continuously in $[0, \infty)$, and most of the efficient solutions as both λ_2 and λ_3 change continuously in $(0, \infty)$. Model (7.86) can be solved by taking partial derivatives with respect to $\mathbf{x}$ and x_{n+1} as follows

$$2\boldsymbol{\Sigma}\mathbf{x} - \lambda_2\boldsymbol{\mu} - \lambda_3(\boldsymbol{\delta} - r_{fd}\mathbf{1}) = \mathbf{0}$$

$$-\lambda_2 r_{fr} + \lambda_3 r_{fd} = 0$$

These two equations determine

$$\mathbf{x} = \frac{1}{2}\boldsymbol{\Sigma}^{-1}[\lambda_2\boldsymbol{\mu} + \frac{r_{fr}}{r_{fd}}\lambda_2(\boldsymbol{\delta} - r_{fd}\mathbf{1})]\,, \ \lambda_2 \geq 0, \quad x_{n+1} \text{ unrestricted} \tag{7.87}$$

Future research will be pursued to check whether (7.85) can determine a unique tangency portfolio on the nondominated surface of (7.7), so that there exists a capital allocation *surface* passing through this portfolio and the two risk-free assets. This portfolio can serve as the market portfolio in equilibrium and assume an important role in extending Sharpe's Capital Asset Pricing Model.

Black [8] and Huang and Litzenberger [43] proved that for any portfolio r_p on the minimum-variance frontier (7.5) except the minimum-variance portfolio of model (7.1), there exists another portfolio r_{pz} on this frontier having zero covariance with r_p. The portfolio r_{pz} can be geometrically determined. The expected return of any asset $E(r_i)$ can be determined by these two portfolio as $E(r_i) = E(r_{pz}) + \frac{Cov(r_i,r_p)}{\sigma_p^2}[E(r_p) - E(r_{pz})]$. Future research will be pursued to check for model (7.7) for any given portfolio r_p on the minimum-variance surface (7.34) the existence of a zero-covariance portfolio r_{pz} also on the surface. That is, the portfolio weight of r_p is $\mathbf{x}^p = e_2^p\mathbf{d}^2 + e_3^p\mathbf{d}^3 + \mathbf{x}^0$ with e_2^p and e_3^p known. Do e_2^{pz} and e_3^{pz} exist such that the portfolio weight of r_{pz} is $\mathbf{x}^{pz} = e_2^{pz}\mathbf{d}^2 + e_3^{pz}\mathbf{d}^3 + \mathbf{x}^0$ and the following equation holds?

$$Cov(r_p, r_{pz}) = (e_2^p\mathbf{d}^2 + e_3^p\mathbf{d}^3 + \mathbf{x}^0)^T\mathbf{\Sigma}(e_2^{pz}\mathbf{d}^2 + e_3^{pz}\mathbf{d}^3 + \mathbf{x}^0) = 0$$

Initial look reveals that there may exist a whole set of r_{pz} in the term of e_2^{pz} and e_3^{pz}. The geometrical properties of this set and the pricing expression based on r_p and r_{pz} need to be analyzed.

7.9 General k Objective models

Model (7.7) can be extended into k-objective formulation by bringing in other linear criteria as follows

$$\begin{aligned}
&\min\{\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} = z_1 \text{ variance}\} \\
&\max\{\ \mathbf{c}^{2^T}\mathbf{x} = z_2 \text{ expected return}\} \\
&\max\{\ \mathbf{c}^{3^T}\mathbf{x} = z_3 \text{ dividend yield}\} \\
&\max\{\ \mathbf{c}^{4^T}\mathbf{x} = z_4 \text{ R \& D}\} \\
&\max\{\ \mathbf{c}^{5^T}\mathbf{x} = z_5 \text{ social responsibility}\} \\
&\qquad\qquad\vdots \\
&\max\{\ \mathbf{c}^{k^T}\mathbf{x} = z_k \text{ other linear objective}\} \\
&\quad \text{s.t.} \quad \mathbf{1}^T\mathbf{x} = 1
\end{aligned} \tag{7.88}$$

where $\mathbf{c}^2$ and $\mathbf{c}^3$ are the expected return vector and the dividend yield vector of the n assets. $\mathbf{c}^3, \mathbf{c}^4, \ldots, \mathbf{c}^k$ are the vectors of the other linear objectives. The feasible region Z of (7.88) is in $\mathbb{R}^k$. Assumption 7.2 is kept. Assumptions 7.1 and 7.3 are generalized as follows.

Assumption 7.4. *The number of securities, n in asset universe is greater than the number of criteria, i.e., $n \geq k+1$.*

Assumption 7.5. *The vectors $\mathbf{c}^2, \mathbf{c}^3, \ldots, \mathbf{c}^k$ and vector one $\mathbf{1}$ are linearly independent.*

Similar to Lemma 7.1 and its proof. A suitable-portfolio investor has the following lemma.

Lemma 7.6. $\mathbf{\Sigma}^{-1}\mathbf{c}^2, \mathbf{\Sigma}^{-1}\mathbf{c}^3, \ldots, \mathbf{\Sigma}^{-1}\mathbf{c}^k, \mathbf{\Sigma}^{-1}\mathbf{1}$ *are linearly independent.*

A standard investor treats (7.88) by an e-constraint approach and obtains the following model

$$\begin{aligned} \min \{\mathbf{x}^T \mathbf{\Sigma} \mathbf{x} = z_1 \ \text{ variance}\} \\ \text{s.t.} \quad {\mathbf{c}^2}^T \mathbf{x} = e_2 \\ {\mathbf{c}^3}^T \mathbf{x} = e_3 \\ \vdots \\ {\mathbf{c}^k}^T \mathbf{x} = e_k \\ \mathbf{1}^T \mathbf{x} = 1 \end{aligned} \tag{7.89}$$

where $e_2, e_3, \ldots, e_k$ are target values of objective 2, objective 3, ... objective k, respectively. Similar to the geometrical interpretation of (7.8), (7.89) will uncover the minimum-variance surface of (7.88) by varying $e_2, e_3, \ldots, e_k$ continuously. Denote

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}^2 & \mathbf{c}^3 & \ldots \mathbf{c}^k & \mathbf{1} \end{bmatrix}_{n \times k} \qquad \mathbf{e} = \begin{bmatrix} e^2 \\ e^3 \\ \vdots \\ e^k \\ 1 \end{bmatrix}_{k \times 1} \tag{7.90}$$

Then (7.89) is re-expressed in matrix format as

$$\begin{aligned} \min \{\mathbf{x}^T \mathbf{\Sigma} \mathbf{x} = z_1 \ \text{ variance}\} \\ \text{s.t.} \quad \mathbf{C}^T \mathbf{x} = \mathbf{e} \end{aligned}$$

Similar to Theorems 7.1 and 7.2 and their proofs, a suitable-portfolio investor has the following theorems to back up (7.89).

Theorem 7.17. *For any given* $\mathbf{e} = (e_2, e_3, \ldots, e_k, 1)$*, let* $S_{e_2e_3\ldots e_k}$ *be the feasible region of* (7.89),

$$S_{e_2e_3\ldots e_k} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{C}^T\mathbf{x} = \mathbf{e}\}$$

Let $Z_{e_2e_3\ldots e_k}$ *be the range of* (7.89)

$$Z_{e_2e_3\ldots e_k} = \{z_1 \in \mathbb{R} \mid z_1 = \mathbf{x}^T\mathbf{\Sigma}\mathbf{x}, \mathbf{x} \in S_{e_2e_3\ldots e_k}\}$$

Then

(i) $S_{e_2e_3\ldots e_k}$ *is unbounded with respect to at least one of its components, i.e., there exists* $i \in \{1, 2, \ldots, n\}$ *such that for any* $y > 0$*, there exists* $(x_1, x_2, \ldots, x_n) \in S_{e_2e_3\ldots e_k}$ *with* $|x_i| > y$.

(ii) $Z_{e_2e_3\ldots e_k}$ *is lower bounded by* 0*, i.e., for any* $z_1 \in Z_{e_2e_3\ldots e_k}, z_1 \geq 0$*, but upper unbounded, i.e., for any* $y > 0$*, there exists* $z_1 \in Z_{e_2e_3\ldots e_k}$ *with* $z_1 > y$.

Theorem 7.18. *The feasible region* Z *of* (7.88) *has the following properties.*

(i) *For any given* $e_2, e_3, \ldots, e_k$*, if* $(\sigma_a^2, e_2, e_3, \ldots, e_k)$ *and* $(\sigma_c^2, e_2, e_3, \ldots, e_k)$ *are images of two feasible portfolios of* (7.88)*, with* $\sigma_a^2 < \sigma_c^2$*, then for any* $\sigma_b^2 \in [\sigma_a^2, \sigma_c^2]$ *there exists a feasible portfolio of* (7.88) *such that its image is* $(\sigma_b^2, e_2, e_3, \ldots, e_k)$.

(ii) Z *is continuous.*

To solve (7.89) and obtain the minimum-variance surface, apply Lagrangian method by constructing

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T\mathbf{\Sigma}\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{e} - \mathbf{C}^T\mathbf{x})$$

where $\boldsymbol{\lambda} = (\lambda_2, \lambda_3, \ldots, \lambda_k, \lambda_{k+1})$ is Lagrangian multipliers. Because $\mathbf{x}^T\mathbf{\Sigma}\mathbf{x}$ is positive semidefinite with respect to $\mathbf{x}$, $\mathbf{x}$ is the minimizing solution of (7.89) if and only if

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{\Sigma}\mathbf{x} - \mathbf{C}\boldsymbol{\lambda} = \mathbf{0}$$
$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = \mathbf{e} - \mathbf{C}^T\mathbf{x} = 0$$

Because $\mathbf{\Sigma}$ is invertible, one can get $\mathbf{x} = \frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{C}\boldsymbol{\lambda}$ from the first equation and plug it into the second equation as $\mathbf{e} - \mathbf{C}^T(\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{C}\boldsymbol{\lambda}) = 0$, $(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})\boldsymbol{\lambda} = 2\mathbf{e}$. Matrix $(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})_{k\times k}$ generalizes matrix $\mathbf{A}$ (7.16). Lemma 7.2 is extended as follows.

Lemma 7.7. *Matrix $(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})_{k\times k}$ is a covariance matrix. Furthermore, it is positive definite and thus invertible.*

Proof. Similar to the proof of Lemma 7.2, $(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})$ is the covariance matrix of $\mathbf{C}^T\mathbf{v}$. □

By Lemma 7.7, one can get $\boldsymbol{\lambda} = 2(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e}$ from $(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})\boldsymbol{\lambda} = 2\mathbf{e}$. Note that one can not have $(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1} = \mathbf{C}^{-1}\mathbf{\Sigma}(\mathbf{C}^T)^{-1}$, because $\mathbf{C}_{n\times k}$ is not a square matrix. After plugging $\boldsymbol{\lambda} = 2(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e}$ into $\mathbf{x} = \frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{C}\boldsymbol{\lambda}$, one can get the minimizing solution of (7.89) as follows

$$\mathbf{x} = \mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e} \tag{7.91}$$

The following theorem generalizes Theorem 7.3 and reveals the $k-1$ dimension of the inverse images (portfolio weights) of the minimum-variance surface of (7.88). The interpretation of this minimizing solution will be given later.

Theorem 7.19. *The rank of matrix $(\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1})_{n\times k}$ is k,*
i.e., $rank(\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}) = k$.

Proof. Want $rank(\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1})_{n\times k} \le k$. This is guaranteed by $n \ge k+1$ by Assumption 7.4.
Want $rank(\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}) \ge k$. The product of $\mathbf{C}^T$ and $\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}$ is

$$\mathbf{C}^T(\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}) = (\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1} = I_k$$

Therefore the rank of the product is k. In linear algebra, $rank(\mathbf{A}_1\mathbf{A}_2) \le min\{rank(\mathbf{A}_1), rank(\mathbf{A}_2)\}$. $rank(\mathbf{C}) = k$ by Assumption 7.5, so $rank(\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}) \ge k$.
Then the only possibility is $rank(\mathbf{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{\Sigma}^{-1}\mathbf{C})^{-1}) = k$. □

Denote the k columns of $(\boldsymbol{\Sigma}^{-1}\boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C})^{-1})_{n\times k}$ as $\mathbf{d}^2, \mathbf{d}^3, \ldots, \mathbf{d}^k, \mathbf{x}^0$. This may be an awkward notation. But this notation can be admissible, because $\mathbf{d}^2, \mathbf{d}^3, \ldots, \mathbf{d}^k$ are extensions of $\mathbf{d}^2$ (7.25) and $\mathbf{d}^3$ (7.26). As the extension of $\mathbf{x}^0$ (7.27), $\mathbf{x}^0$ as the kth column of $(\boldsymbol{\Sigma}^{-1}\boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C})^{-1})$ is the minimizing solution of (7.89) when $e_2 = 0, e_3 = 0, \ldots, e_k = 0$. Then for any given $e_2, e_3, \ldots, e_k$ re-express (7.91) in this notation as $\mathbf{x} = e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \ldots + e_k\mathbf{d}^k + \mathbf{x}^0$. As $e_2, e_3, \ldots, e_k$ vary, the minimizing solutions of (7.89) are as follows

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \ldots + e_k\mathbf{d}^k + \mathbf{x}^0,\ e_2, e_3, \ldots, e_k \in \mathbb{R}\} \tag{7.92}$$

Expression (7.92) is the portfolio weights (inverse images) of the minimum-variance surface of (7.88). Because $\mathbf{d}^2, \mathbf{d}^3, \ldots, \mathbf{d}^k$ are linearly independent as implied by Theorem 7.19, (7.92) is a $k-1$ dimensional affine subspace in decision space and generalizes the inverse images (portfolio weights) of the minimum-variance surface (7.32). Theorem 7.4 is extended as

Theorem 7.20. *The affine subspace* (7.92) *has the following properties.*

- **(i)** *Any k points $\mathbf{x}^a, \mathbf{x}^b, \ldots, \mathbf{x}^k$ on it can span* (*generate*) *this affine subspace, as long as $\mathbf{x}^a, \mathbf{x}^b, \ldots, \mathbf{x}^k$ are affinely independent.*
- **(ii)** *Any convex combination of any m points on this affine subspace is still on this affine subspace.*

Proof. Similar to the proof of Theorem 7.4 except that the portfolio weights of the minimum-variance surface of (7.88) can be expressed as

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{x}^a + t_b(\mathbf{x}^b - \mathbf{x}^a) + t_c(\mathbf{x}^c - \mathbf{x}^a) + \ldots + t_k(\mathbf{x}^k - \mathbf{x}^a),\ t_b, t_c, \ldots, t_k \in \mathbb{R}\}$$

□

Theorem 7.20 implies that if all investors hold portfolio weights (inverse images) of the minimum-variance surface of (7.88), then the market portfolio as a convex combination of all the investors' portfolios also belongs to the portfolio weights of the minimum-variance surface. The minimum-variance surface of (7.88) is obtained by plugging in the expression of the affine subspace (7.92).

$$
\begin{aligned}
V[R(\mathbf{x},\mathbf{r})] &= \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} \\
&= (e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \ldots + e_k\mathbf{d}^k + \mathbf{x}^0)^T\boldsymbol{\Sigma}(e_2\mathbf{d}^2 + e_3\mathbf{d}^3 + \ldots + e_k\mathbf{d}^k + \mathbf{x}^0) \\
&= \sum_{i=2}^{k}\mathbf{d}^{i^T}\boldsymbol{\Sigma}\mathbf{d}^i e_i^2 + \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0 + \sum_{i\neq j}\mathbf{d}^{i^T}\boldsymbol{\Sigma}\mathbf{d}^j e_i e_j + 2\sum_{i=2}^{k}\mathbf{d}^{i^T}\boldsymbol{\Sigma}\mathbf{x}^0 e_i
\end{aligned}
$$

or re-express in the symbols of $z_1, z_2, \ldots, z_k$ of criterion space as

$$
z_1 = \sum_{i=2}^{k}\mathbf{d}^{i^T}\boldsymbol{\Sigma}\mathbf{d}^i z_i^2 + \mathbf{x}^{0^T}\boldsymbol{\Sigma}\mathbf{x}^0 + \sum_{i\neq j}\mathbf{d}^{i^T}\boldsymbol{\Sigma}\mathbf{d}^j z_i z_j + 2\sum_{i=2}^{k}\mathbf{d}^{i^T}\boldsymbol{\Sigma}\mathbf{x}^0 z_i \tag{7.93}
$$

Theorem 7.5 and Corollary 7.2 are generalized as follows.

Theorem 7.21. *The minimum-variance surface* (7.93) *of model* (7.88) *is a non-degenerate paraboloid in the criterion space of model* (7.88).

Proof. Plug $\mathbf{x} = \boldsymbol{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e}$ (7.91) in the expression of the affine subspace (7.92). Because $\boldsymbol{\Sigma}^{-1}$ is symmetric and easy to check $(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1^T} = (\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{T^{-1}} = (\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}$,

$$
\begin{aligned}
V[R(\mathbf{x},\mathbf{r})] &= \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} \\
&= [\boldsymbol{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e}]^T\,\boldsymbol{\Sigma}\,[\boldsymbol{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e}] \\
&= [\mathbf{e}^T(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{C}^T\boldsymbol{\Sigma}^{-1}]\,\boldsymbol{\Sigma}\,[\boldsymbol{\Sigma}^{-1}\mathbf{C}(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e}] \\
&= \mathbf{e}^T(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e} \\
&= \mathbf{e}^T(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{e}
\end{aligned}
$$

Matrix $(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}$ is a covariance matrix and positive definite because matrix $(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C})$ is by Lemma 7.7. Then similar to the proof of Theorem 7.5, the expression above is a non-degenerate paraboloid. □

Corollary 7.5. *The feasible region Z of* (7.88) *is an unbounded continuous set in criterion space, with a non-degenerate paraboloidal boundary.*

Proof. Based on Theorem 7.21 and similar to the proof of Corollary 7.2. □

Similar to the convoluted expression of the minimum-variance surface (7.34) of model (7.7), identifying the minimum-variance portfolio and nondominated surface from (7.93) is difficult. Therefore a weighted-sums approach is introduced as follows

$$\min \{\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} - \lambda_2 {\mathbf{c}^2}^T\mathbf{x} - \lambda_3 {\mathbf{c}^3}^T\mathbf{x} - \ldots - \lambda_k {\mathbf{c}^k}^T\mathbf{x} = z_w\} \quad \lambda_2 \geq 0, \ldots, \lambda_k \geq 0 \tag{7.94}$$
$$\text{s.t.} \qquad \mathbf{1}^T\mathbf{x} = 1$$

The financial interpretation of (7.94) is similar to that of (7.45). Similar to Theorem 7.6, the following theorem demonstrates that (7.94) will locate all the efficient solutions of (7.88).

Theorem 7.22. *Every weakly efficient point of* (7.88) *is also efficient.*

Proof. Similar to the proof of Theorem 7.6. □

Denote

$$\mathbf{C}_{2,3,\ldots,k} = \begin{bmatrix}\mathbf{c}^2 & \mathbf{c}^3 & \ldots \mathbf{c}^k\end{bmatrix}_{n\times(k-1)} \qquad \boldsymbol{\lambda}_{2,3,\ldots,k} = \begin{bmatrix}\lambda_2\\ \lambda_3\\ \vdots,\\ \lambda_k\end{bmatrix}_{(k-1)\times 1} \tag{7.95}$$

To solve (7.94), apply Lagrangian method by constructing

$$L(\mathbf{x}, \lambda_{k+1}) = \mathbf{x}^T\mathbf{\Sigma}\mathbf{x} - \boldsymbol{\lambda}^T_{2,3,\ldots,k}({\mathbf{C}_{2,3,\ldots,k}}^T\mathbf{x}) + \lambda_{k+1}(1 - \mathbf{1}^T\mathbf{x})$$

where λ_{k+1} is the only Lagrangian multiplier. Because $\mathbf{x}^T\mathbf{\Sigma}\mathbf{x}$ is positive semidefinite with respect to $\mathbf{x}$, $\mathbf{x}$ is the minimizing solution of (7.94) if and only if

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{\Sigma}\mathbf{x} - \mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} - \lambda_{k+1}\mathbf{1} = \mathbf{0}$$
$$\frac{\partial L}{\partial \lambda_{k+1}} = 1 - \mathbf{1}^T\mathbf{x} = 0$$

One can get $\mathbf{x} = \frac{1}{2}\mathbf{\Sigma}^{-1}(\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} + \lambda_{k+1}\mathbf{1})$ from the first equation and plug into the second equation as $1 - \mathbf{1}^T\frac{1}{2}\mathbf{\Sigma}^{-1}(\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} + \lambda_{k+1}\mathbf{1}) = 0$. Then

$$\begin{aligned}\lambda_{k+1} &= \frac{1}{\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1}}(2 - \mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k}) \quad \text{, by (7.2)}\\ &= \frac{1}{f}(2 - \mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k})\end{aligned}$$

Plug this into $\mathbf{x} = \frac{1}{2}\mathbf{\Sigma}^{-1}(\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} + \lambda_{k+1}\mathbf{1})$ as

$$\begin{aligned}\mathbf{x} &= \frac{1}{2}\mathbf{\Sigma}^{-1}[\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} + (\frac{1}{f}(2 - \mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k}))\mathbf{1}]\\ &= \frac{1}{2}\mathbf{\Sigma}^{-1}[\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k})\mathbf{1}] + \frac{1}{f}(\mathbf{\Sigma}^{-1}\mathbf{1})\end{aligned}$$

Because $(\mathbf{1}^T_{1\times n}\mathbf{\Sigma}^{-1}_{n\times n}\mathbf{C}_{2,3,\ldots,k\;n\times(k-1)}\,\boldsymbol{\lambda}_{2,3,\ldots,k\;(k-1)\times 1})$ is a scalar,

$$= \frac{1}{2}[\mathbf{\Sigma}^{-1}\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k})\mathbf{\Sigma}^{-1}\mathbf{1}] + \frac{1}{f}(\mathbf{\Sigma}^{-1}\mathbf{1})$$

By (7.95) $\mathbf{C}_{2,3,\ldots,k}\,\boldsymbol{\lambda}_{2,3,\ldots,k} = \begin{bmatrix}\mathbf{c}^2 & \mathbf{c}^3 & \ldots\mathbf{c}^k\end{bmatrix}\begin{bmatrix}\lambda_2\\ \lambda_3\\ \vdots,\\ \lambda_k\end{bmatrix} = \lambda_2\mathbf{c}^2 + \lambda_3\mathbf{c}^3 + \ldots + \lambda_k\mathbf{c}^k$. Plug this into the equation above and obtain

$$\begin{aligned}\mathbf{x} &= \frac{1}{2}[\mathbf{\Sigma}^{-1}(\lambda_2\mathbf{c}^2 + \lambda_3\mathbf{c}^3 + \ldots + \lambda_k\mathbf{c}^k) - \frac{1}{f}\mathbf{1}^T\mathbf{\Sigma}^{-1}(\lambda_2\mathbf{c}^2 + \lambda_3\mathbf{c}^3 + \ldots + \lambda_k\mathbf{c}^k)\mathbf{\Sigma}^{-1}\mathbf{1}]\\ &\quad + \frac{1}{f}(\mathbf{\Sigma}^{-1}\mathbf{1})\\ &= \frac{1}{2}\{\lambda_2[\mathbf{\Sigma}^{-1}\mathbf{c}^2 - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{c}^2)\mathbf{\Sigma}^{-1}\mathbf{1}] + \lambda_3[\mathbf{\Sigma}^{-1}\mathbf{c}^3 - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{c}^3)\mathbf{\Sigma}^{-1}\mathbf{1}]\\ &\quad + \ldots + \lambda_k[\mathbf{\Sigma}^{-1}\mathbf{c}^k - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{c}^k)\mathbf{\Sigma}^{-1}\mathbf{1}]\} + \frac{1}{f}(\mathbf{\Sigma}^{-1}\mathbf{1})\\ &= \lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \lambda_4\mathbf{\Delta}^4 + \ldots + \lambda_k\mathbf{\Delta}^k + \mathbf{x}^1 \qquad (7.96)\end{aligned}$$

where

$$\begin{aligned}
\mathbf{\Delta}^2 &= \frac{1}{2}[\mathbf{\Sigma}^{-1}\mathbf{c}^2 - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{c}^2)\mathbf{\Sigma}^{-1}\mathbf{1}] = \mathbf{\Delta}^2 \text{ defined in (7.53), because } \mathbf{c}^2 = \boldsymbol{\mu} \\
\mathbf{\Delta}^3 &= \frac{1}{2}[\mathbf{\Sigma}^{-1}\mathbf{c}^3 - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{c}^3)\mathbf{\Sigma}^{-1}\mathbf{1}] = \mathbf{\Delta}^3 \text{ defined in (7.54), because } \mathbf{c}^3 = \boldsymbol{\delta} \\
\mathbf{\Delta}^4 &= \frac{1}{2}[\mathbf{\Sigma}^{-1}\mathbf{c}^4 - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{c}^4)\mathbf{\Sigma}^{-1}\mathbf{1}] \\
&\vdots \\
\mathbf{\Delta}^k &= \frac{1}{2}[\mathbf{\Sigma}^{-1}\mathbf{c}^k - \frac{1}{f}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{c}^k)\mathbf{\Sigma}^{-1}\mathbf{1}] \\
\mathbf{x}^1 &= \frac{1}{f}\mathbf{\Sigma}^{-1}\mathbf{1} = \mathbf{x}^1 \text{ defined in (7.55)}
\end{aligned}$$

Similar to (7.57), as $\lambda_2, \lambda_3, \ldots, \lambda_k$ vary, the efficient set of (7.88) is given by

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \ldots + \lambda_k\mathbf{\Delta}^k + \mathbf{x}^1, \lambda_2 \geq 0, \lambda_3 \geq 0, \ldots, \lambda_k \geq 0\} \quad (7.97)$$

This is a translated cone with $\mathbf{\Delta}^2, \mathbf{\Delta}^3, \ldots, \mathbf{\Delta}^k$ as the generators and $\mathbf{x}^1$ as the extreme point. It is a convex set. Then any convex combination of any points belonging to the cone is also in the cone. The implication is that if all investors hold efficient portfolios of (7.88), then the market portfolio, as a convex combination of these portfolios is also efficient of (7.88). This cone generalizes the efficient sets of (7.1) and (7.7) by extending Theorems 7.7 and 7.9 as follows.

Theorem 7.23. *The minimum-variance portfolio* $\mathbf{x}^1$ *of* (7.88) *is also the minimum-variance portfolio of* (7.1) *and* (7.7).

Proof. Three models have the same minimum-variance portfolio $\mathbf{x} = \frac{1}{f}\mathbf{\Sigma}^{-1}\mathbf{1}$. □

Theorem 7.24. *The efficient sets* (7.6) *of model* (7.1) *and* (7.57) *of model* (7.7) *are subsets of the efficient set* (7.97) *of model* (7.88).

Proof. Similar to the proof of Theorem 7.9 and (7.57) is a subset of (7.97) by setting $\lambda_4 = 0, \lambda_5 = 0, \ldots, \lambda_k = 0$ in (7.97). □

Theorem 7.25. *A multiple criteria portfolio selection model takes the first j objectives of model* (7.88) *as follows*

$$
\begin{aligned}
&min\{\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} = z_1\} \\
&max\{\ \mathbf{c}^{2^T}\mathbf{x} = z_2\} \\
&max\{\ \mathbf{c}^{3^T}\mathbf{x} = z_3\} \\
&\qquad\vdots \\
&max\{\ \mathbf{c}^{j^T}\mathbf{x} = z_j\} \\
&s.t.\quad \mathbf{1}^T\mathbf{x} = 1
\end{aligned}
\tag{7.98}
$$

where $j < k$. Then the efficient set of model (7.98) *is*

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \lambda_2\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\Delta}^3 + \ldots + \lambda_j\boldsymbol{\Delta}^j + \mathbf{x}^1, \lambda_2 \geq 0, \lambda_3 \geq 0, \ldots, \lambda_j \geq 0\} \tag{7.99}$$

which is a subset of the efficient set (7.97) *of model* (7.88)*.*

Proof. The derivation of E of model (7.98) is similar to that of model (7.88) except for the first j objectives of model (7.88). Then similar to the proof of Theorem 7.24, E of model (7.98) is a subset of (7.97) by setting $\lambda_{j+1} = 0, \lambda_{j+2} = 0, \ldots, \lambda_k = 0$. □

Theorem 7.24 and Theorem 7.25 demonstrate that the previously efficient portfolios will still stay efficient as a suitable-portfolio investor gradually adds objectives. Furthermore the suitable-portfolio investor can enjoy greater flexibility to choose from the higher dimensional nondominated surface.

Theorem 7.25 can also serve as an experiment to illustrate the possible loss of information that can happen in a general multiple criteria optimization context, similar to the experiment in Section 2.4. Suppose the true (complete) model is (7.88) deployed by **suitable-portfolio investor**^{k}. With $j < k$, **suitable-portfolio investor**^{j} takes the first j objectives of (7.88) and deploys (7.98), while **standard investor**^{2} takes the first 2 objectives of (7.88) and deploys (7.1). Then the efficient sets of these three investors are in (7.97), (7.99) and (7.6), respectively. **suitable-portfolio investor**^{k} have the complete information. Then the potential loss of information of **suitable-portfolio investor**^{j} is that she or he does not observe $\boldsymbol{\Delta}^{j+1}, \boldsymbol{\Delta}^{j+2}, \ldots, \boldsymbol{\Delta}^k$ and the solutions spanned from $\boldsymbol{\Delta}^{j+1}, \boldsymbol{\Delta}^{j+2}, \ldots, \boldsymbol{\Delta}^k$. Similarly, the potential loss of information of **standard investor**^{2} is that she or he does not observe $\boldsymbol{\Delta}^3, \boldsymbol{\Delta}^4, \ldots, \boldsymbol{\Delta}^k$ and the solutions spanned from $\boldsymbol{\Delta}^3, \boldsymbol{\Delta}^4, \ldots, \boldsymbol{\Delta}^k$. This experiment is more realistic in objective functions but possibly less realistic in constraints than the experiment in Section 2.4. It also confirms the potential loss of information of a standard investor. Future research will be taken to check whether $\boldsymbol{\Delta}^2, \boldsymbol{\Delta}^3, \ldots, \boldsymbol{\Delta}^k$ are linearly independent. The educated guess is yes, based on Theorem 7.8 and Theorem 7.19. Then the potential loss of information can be described strictly by dimensionality.

The nondominated surface of model (7.88) can be delineated by plugging the efficient solution (7.96) as follows

$$
\begin{aligned}
z_1 &= (\lambda_2\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\Delta}^3 + \ldots + \lambda_k\boldsymbol{\Delta}^k + \mathbf{x}^1)^T\,\boldsymbol{\Sigma}\,(\lambda_2\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\Delta}^3 + \ldots + \lambda_k\boldsymbol{\Delta}^k + \mathbf{x}^1)\\
z_2 &= \mathbf{c}^{2^T}(\lambda_2\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\Delta}^3 + \ldots + \lambda_k\boldsymbol{\Delta}^k + \mathbf{x}^1)\\
&\vdots\\
z_k &= \mathbf{c}^{k^T}(\lambda_2\boldsymbol{\Delta}^2 + \lambda_3\boldsymbol{\Delta}^3 + \ldots + \lambda_k\boldsymbol{\Delta}^k + \mathbf{x}^1)
\end{aligned}
$$

The e-constraint approach (7.89) and the weighted-sums approach (7.94) can be connected by

$$
\begin{aligned}
\mathbf{c}^{2^T}\mathbf{x} &= z_2 \quad \text{and let} \quad z_2 = e_2\\
\mathbf{c}^{3^T}\mathbf{x} &= z_3 \quad \text{and let} \quad z_3 = e_3\\
&\vdots\\
\mathbf{c}^{k^T}\mathbf{x} &= z_k \quad \text{and let} \quad z_k = e_k
\end{aligned}
$$

Pre-multiply $\mathbf{x}$ (7.96) by $\mathbf{c}^{2^T}, \mathbf{c}^{3^T}, \ldots, \mathbf{c}^{k^T}$, respectively and plug the products into the equations above.

$$\mathbf{c}^{2^T}(\lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \ldots + \lambda_k\mathbf{\Delta}^k + \mathbf{x}^1) = e_2$$
$$\mathbf{c}^{3^T}(\lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \ldots + \lambda_k\mathbf{\Delta}^k + \mathbf{x}^1) = e_3$$
$$\vdots$$
$$\mathbf{c}^{k^T}(\lambda_2\mathbf{\Delta}^2 + \lambda_3\mathbf{\Delta}^3 + \ldots + \lambda_k\mathbf{\Delta}^k + \mathbf{x}^1) = e_k$$

When a suitable-portfolio investor projects the feasible region Z of (7.88) from high-dimensional space to (variance, expected return) space, Theorem 7.14 still holds. Theorem 7.16 is extended as follows.

Theorem 7.26. *The nondominated frontier of model* (7.1) *is the "minimizing variance-maximizing expected return" boundary of the projection of the nondominated surface of model* (7.88).

Proof. Similar to the proof of Theorem 7.16 and by Theorem 7.24. □

Together with Theorem 7.16, this theorem can fortify a suitable-portfolio investor's explanation of buried market portfolio. The projection of the market portfolio can be buried deeper and deeper below the nondominated frontier as a suitable-portfolio investor adds more linear objectives.

Chapter 8

The Nondominated Surface of a Two quadratic and Two Linear Criteria Optimization

This chapter outlines a research being undertaken by Steuer, Hirschberger and Qi. This research is about computing the nondominated surface of the following model

$$\begin{aligned} \min \{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} &= z_1\} \\ \min \{\mathbf{x}^T \boldsymbol{\Pi} \mathbf{x} &= z_2\} \\ \max \{\boldsymbol{\mu}^T \mathbf{x} \quad &= z_3\} \\ \max \{\boldsymbol{\delta}^T \mathbf{x} \quad &= z_4\} \\ \text{s.t.} \quad \mathbf{x} \quad &\in S \end{aligned} \tag{8.1}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Pi}$ are $n \times n$ positive semidefinite matrices, $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$ are $n \times 1$ vectors, and S is created by linear constraints. Note that (8.1) is a two quadratic and two linear criteria optimization in operations research.

Model (8.1) can be used as a substitute problem (mean-variance approach) of a suitable-portfolio investor's portfolio selection model. For example, if a suitable-portfolio investor has a two (linear) objective stochastic programming model to maximize portfolio return and to maximize portfolio dividend, then (8.1) is a substitute problem (mean-variance approach) of the stochastic programming model by taking $\boldsymbol{\Sigma}$ and $\boldsymbol{\Pi}$ as the covariance matrices of the n securities' returns and dividend yield, respectively, and by taking $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$ as the expected values of the n securities' returns and dividend yield, respectively. Assumption 6.1 that dividend yields, social responsibility, and $R\&D$ are deterministic criteria can be relaxed by

(8.1). Chow [16] 's model to minimize (return) variance, minimize tracking error variance, and maximize expected return can also be an example of (8.1) in portfolio selection.

The nondominated surface of (8.1) is difficult to compute, because of the two quadratic objective functions. The author is not aware of major research breakthroughs in this area.

8.1 A Weighted-sums Approach for Obtaining a Discretized Representation of the Nondominated Surface

Because $\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}$ and $\mathbf{x}^T\boldsymbol{\Pi}\mathbf{x}$ are convex functions with respect to $\mathbf{x}$, a weighted-sums approach can be applied to (8.1) as follows

$$\max\left\{-\lambda_1\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} - \lambda_2\mathbf{x}^T\boldsymbol{\Pi}\mathbf{x} + \lambda_3\boldsymbol{\mu}^T\mathbf{x} + \lambda_4\boldsymbol{\delta}^T\mathbf{x} = z_w\right\} \quad \text{s.t.} \quad \mathbf{x} \in S \tag{8.2}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is a weighting vector and $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, and $\boldsymbol{\Lambda}$ is the weighting space as

$$\boldsymbol{\Lambda} = \{(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \in \mathbb{R}^4 \mid \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1,\ \lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0\} \tag{8.3}$$

Every maximizing solution of (8.2) is a properly efficient solution of (8.1) and therefore an efficient solution of (8.1). The criterion vector (image) of the maximizing solution is on the nondominated surface. The following lemma demonstrates a property of (8.2).

Lemma 8.1. *Let $a_1 > 0$ and $a_2 > 0$. If $\mathbf{A}^1$ and $\mathbf{A}^2$ are two $n \times n$ positive semidefinite matrices, then $a_1\mathbf{A}^1 + a_2\mathbf{A}^2$ is also positive semidefinite.*

Proof. Note that for any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^T(a_1\mathbf{A}^1 + a_2\mathbf{A}^2)\mathbf{x} = a_1\mathbf{x}^T\mathbf{A}^1\mathbf{x} + a_2\mathbf{x}^T\mathbf{A}^2\mathbf{x} \geq 0$, because $a_1 > 0$ and $a_2 > 0$, and $\mathbf{x}^T\mathbf{A}^1\mathbf{x} \geq 0$ and $\mathbf{x}^T\mathbf{A}^2\mathbf{x} \geq 0$. Therefore, $a_1\mathbf{A}^1 + a_2\mathbf{A}^2$ is positive semidefinite. □

The objective function of (8.2) can be re-expressed as

$$-\lambda_1\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} - \lambda_2\mathbf{x}^T\boldsymbol{\Pi}\mathbf{x} + \lambda_3\boldsymbol{\mu}^T\mathbf{x} + \lambda_4\boldsymbol{\delta}^T\mathbf{x} = -\mathbf{x}^T(\lambda_1\boldsymbol{\Sigma} + \lambda_2\boldsymbol{\Pi})\mathbf{x} + (\lambda_3\boldsymbol{\mu} + \lambda_4\boldsymbol{\delta})^T\mathbf{x}$$

By Lemma 8.1, for given $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, $-\mathbf{x}^T(\lambda_1\boldsymbol{\Sigma} + \lambda_2\boldsymbol{\Pi})\mathbf{x} + (\lambda_3\boldsymbol{\mu} + \lambda_4\boldsymbol{\delta})^T\mathbf{x}$ is a convex function with respect to $\mathbf{x}$. Consequently, (8.2) is a quadratic programming model (it is usually assumed that the objective function of a quadratic programming model in maximization sense is convex). It would be ideal to find a quadratic parametric programming algorithm for (8.2). Unfortunately, such an algorithm does not exist. However, (8.2) can provide a discretized representation of the nondominated surface of (8.1) by the following steps.

1. Select a set of $\boldsymbol{\lambda}$ vectors from $\boldsymbol{\Lambda}$. It is preferable to have these $\boldsymbol{\lambda}$ uniformly distributed in $\boldsymbol{\Lambda}$, so that they can be an even representation of $\boldsymbol{\Lambda}$. Simplex lattices can be a good candidate, which will be introduced shortly.

2. For each of these $\boldsymbol{\lambda}$ vectors, solve (8.2) by a quadratic programming solver (for example Hirschberger, Qi and Steuer [40]) and get a maximizing solution.

3. Such a maximizing solution is an efficient solution of (8.1). Therefore, its criterion vector (image) is on the nondominated surface of (8.1).

4. The criterion vectors (images) of these maximizing solutions are a discretized representation of the nondominated surface.

This approach can be practical. For example, for a 500 security portfolio selection problem of (8.1), selecting 220 $\boldsymbol{\lambda}$ vectors, it takes Hirschberger's procedure [40] $220 \times 72 = 15,840$ seconds =4.4 hours in expectation to finish this approach. The advantage of this approach is being straightforward and practical. However, this approach has the following disadvantages.

1. This style of repetitive solving (optimization) is time-consuming.

2. It loses its appeal in six objective (three quadratic and three linear) or more models, because the number of weighting vectors increases dramatically with the number of objectives. For example, for a six objective (three quadratic and three linear) problem, 2,002 weighting vectors are needed, and nine times of the 4.4 hours are needed.

3. It can not reveal the structure of the efficient set of (8.1).

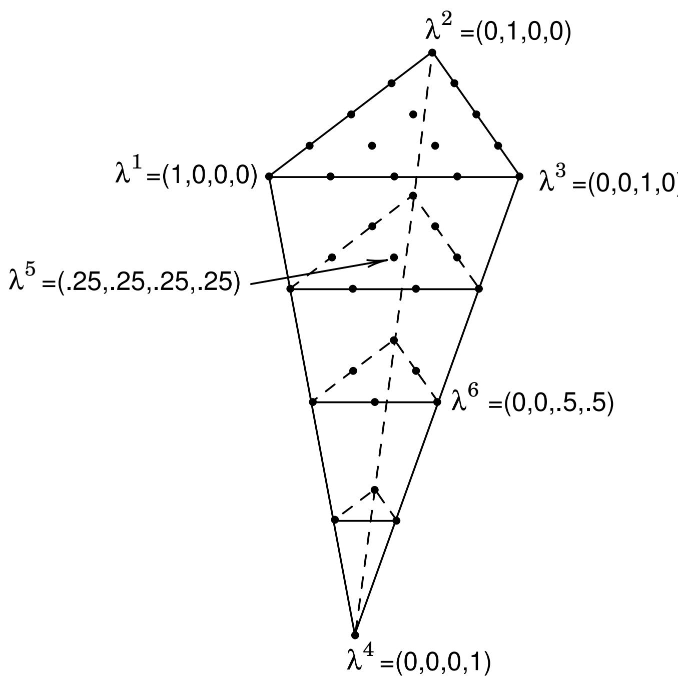

Figure 8.1: Simplex lattices in $\mathbb{R}^4$

4. It has difficulty in obtaining an even representation of the nondominated surface, although the weighting vectors are uniformly distributed.

Simplex lattices can be an appropriate way to select the weighting vectors. A *simplex* in $\mathbb{R}^k$ is $\{(\lambda_1, \lambda_2, \ldots, \lambda_k) \in \mathbb{R}^k \mid \lambda_1 + \lambda_2 + \ldots + \lambda_k = 1,\ \lambda_1, \lambda_2, \ldots, \lambda_k \geq 0\}$ *Simplex lattices* are a set of evenly positioned points belonging to a simplex. For example, a simplex in $\mathbb{R}^4$ is in Figure 8.1. The simplex is represented by a unit tetrahedron in the graph. The four extreme points of the tetrahedron are $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2, \boldsymbol{\lambda}^3$, and $\boldsymbol{\lambda}^4$. Each edge of the tetrahedron is partitioned into 4 equal parts with 5 partition points including the end points. Then the simplex lattices contain 35 points. Each point is represented by a dot. Each point corresponds to a weighting vector. The weighting vectors of $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2, \ldots, \boldsymbol{\lambda}^6$ are $(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1), (.25,.25,.25,.25)$ and $(0,0,.5,.5)$, respectively.

Scheffe [81] and Liou [57] discussed properties of simplex lattices. For example, the number of points that a simplex lattices contain and the positions of these points are fixed, after the number of partition points on an edge is determined. The previously used 220 weighting vectors come from the fact that if a decision maker partitions an edge of a simplex in $\mathbb{R}^4$ into 9 equal parts with 10 partition points, then the simplex lattices contain 220 points (weighting vectors). The previously used 2,002 weighting vectors come from the fact that if a decision maker partitions an edge of a simplex in $\mathbb{R}^6$ into 9 equal parts with 10 partition points, then the simplex lattices contain 2,002 points (weighting vectors).

As an alternative to the approach to obtain a discretized representation of the nondominated surface, interactive weighted-sums approach (see for example Steuer and Schuler [91]) can be applied to (8.1). This approach tries to locate a decision maker's optimal solution or near optimal solution in several iterations, by selecting j weighting vectors, presenting the j criterion vectors of the j maximizing solutions to a decision maker, asking the decision maker to choose one criterion vector as the best, selecting j more finely distributed weighting vectors based on the weighting vector corresponding to the chosen criterion vector by the decision maker, and following the same style for the next iteration. For example, take $j = 3$, and in order to visualize, plot the weighting vectors in two dimensional space (although the weighting vectors are in four dimensional space). It works as follows.

1. Select three weighting vectors $\boldsymbol{\lambda}^{1,1}, \boldsymbol{\lambda}^{1,2}$ and $\boldsymbol{\lambda}^{1,3}$ represented in Figure 8.2 *top*.

2. Solve (8.2) three times based on $\boldsymbol{\lambda}^{1,1}, \boldsymbol{\lambda}^{1,2}$ and $\boldsymbol{\lambda}^{1,3}$, and obtain three criterion vectors (images) of the three maximizing solutions.

3. Present the three criterion vectors to a decision maker. Suppose the decision maker chooses the maximizing solution corresponding to $\boldsymbol{\lambda}^{1,2}$.

4. Centering at $\boldsymbol{\lambda}^{1,2}$, draw three weighting vectors $\boldsymbol{\lambda}^{2,1}, \boldsymbol{\lambda}^{2,2}$ and $\boldsymbol{\lambda}^{2,3}$ in Figure 8.2 *middle*. Weighting vectors $\boldsymbol{\lambda}^{2,1}, \boldsymbol{\lambda}^{2,2}$ and $\boldsymbol{\lambda}^{2,3}$ are more finely distributed than $\boldsymbol{\lambda}^{1,1}, \boldsymbol{\lambda}^{1,2}$ and $\boldsymbol{\lambda}^{1,3}$.

5. Solve (8.2) three times based on $\boldsymbol{\lambda}^{2,1}, \boldsymbol{\lambda}^{2,2}$ and $\boldsymbol{\lambda}^{2,3}$, and obtain three criterion vectors (images) of the three maximizing solutions.

6. Present the three criterion vectors to the decision maker. Suppose the decision maker chooses the maximizing solution corresponding to $\boldsymbol{\lambda}^{2,3}$.

7. Then obtain weighing vector $\boldsymbol{\lambda}^{3,1}, \boldsymbol{\lambda}^{3,2}$ and $\boldsymbol{\lambda}^{3,3}$ in Figure 8.2 *bottom*. Weighting vectors $\boldsymbol{\lambda}^{3,1}, \boldsymbol{\lambda}^{3,2}$ and $\boldsymbol{\lambda}^{3,3}$ are more finely distributed than $\boldsymbol{\lambda}^{2,1}, \boldsymbol{\lambda}^{2,2}$ and $\boldsymbol{\lambda}^{2,3}$. Repeat Steps $\mathbf{4}, \mathbf{5}, \mathbf{6}$ until the decision maker gets a satisfactory solution, or start again from Step $\mathbf{1}$.

Simplex lattices can create finer and finer positioned points (weighting vectors) in the interactive weighted-sums approach. For example, partition an edge of a simplex in $\mathbb{R}^4$ into 2 equal parts, 2^2 equal parts, and 2^3 equal parts in the first three iterations, and so forth. The resultant groups of simplex lattices become finer and finer for the first three iterations and so forth.

In order to try to reveal the structure of the efficient set of (8.1), a simplified version of (8.1) is introduced as

$$\begin{aligned}
&\min \{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} = z_1\} \\
&\min \{\mathbf{x}^T \boldsymbol{\Pi} \mathbf{x} = z_2\} \\
&\max \{\boldsymbol{\mu}^T \mathbf{x} \quad = z_3\} \\
&\max \{\boldsymbol{\delta}^T \mathbf{x} \quad = z_4\} \\
&\quad \text{s.t.} \;\; \mathbf{1}^T \mathbf{x} = 1
\end{aligned}$$

Then (8.2) is simplified as

$$\begin{aligned}
\max \{-\lambda_1 \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} - \lambda_2 \mathbf{x}^T \boldsymbol{\Pi} \mathbf{x} + \lambda_3 \boldsymbol{\mu}^T \mathbf{x} + \lambda_4 \boldsymbol{\delta}^T \mathbf{x} = z_w\} \\
\text{s.t.} \qquad\qquad \mathbf{1}^T \mathbf{x} = 1
\end{aligned} \tag{8.4}$$

where $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \in \boldsymbol{\Lambda}$. Then Lagrangian method can be applied to try to obtain closed-form results and reveal the structure of the efficient set of (8.1).

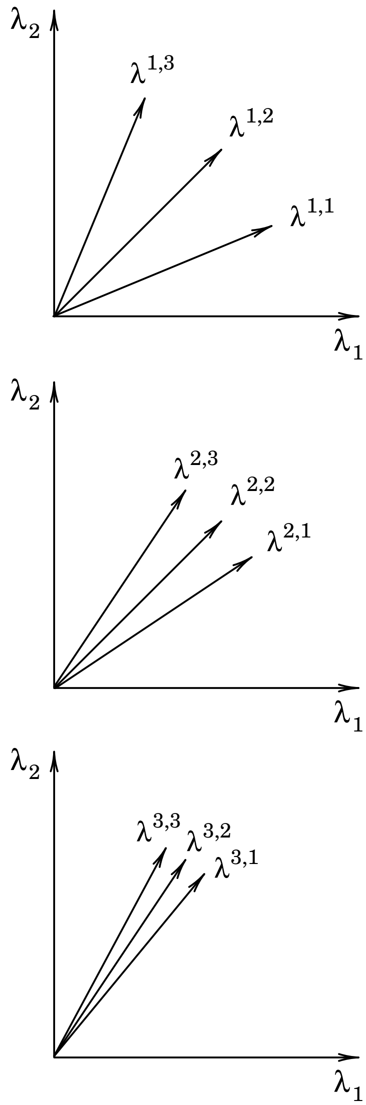

Figure 8.2: The weighting vectors of the first three iterations of interactive weighted-sum approach of (8.1). Suppose the decision maker selects the maximizing solutions corresponding to $\boldsymbol{\lambda}^{1,2}$ and $\boldsymbol{\lambda}^{2,3}$ in the first two iterations, respectively.

8.2 Approximating the Nondominated Surface by Linearizing one Quadratic Objective Function

This section describes an approach to linearize one quadratic objective function of (8.1) and then use the nondominated surface of the resultant model (one quadratic and three linear objectives) to approximate the nondominated surface of (8.1).

For a portfolio selection model

$$\begin{aligned} \min \{\mathbf{x}^T \mathbf{\Sigma} \mathbf{x} = z_1\} \\ \max \{\boldsymbol{\mu}^T \mathbf{x} = z_2\} \\ \text{s.t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{8.5}$$

suppose an investor has the realized or historical return matrix $(q_{ij})_{t \times n}$ for the n securities. That is, t is the number of time periods, $q_{1i}, q_{2i}, \ldots, q_{ti}$ are realized or historical returns of the ith security. Konno and Yamazaki [53] and Feinstein and Thapa [29] proposed to use mean absolute deviation (MAD) of portfolio return instead of standard deviation of portfolio return as follows

$$
\begin{array}{llr}
\min \{ & \frac{1}{t}v_1 + \frac{1}{t}v_2 + \ldots + \frac{1}{t}v_t + \frac{1}{t}w_1 + \frac{1}{t}w_2 + \ldots + \frac{1}{t}w_t & = z_1^{MAD}\} \\
\max \{ & \mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n & = z_2\} \\
\text{s.t.} & a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & \leq b_1 \\
& a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & \leq b_2 \\
& \vdots & \\
& a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n & \leq b_m \\
& d_{11}x_1 + d_{12}x_2 + \ldots + d_{1n}x_n \; - v_1 \; + w_1 & = 0 \\
& d_{21}x_1 + d_{22}x_2 + \ldots + d_{2n}x_n \; - v_2 \; + w_2 & = 0 \\
& \vdots & \\
& d_{t1}x_1 + d_{t2}x_2 + \ldots + d_{tn}x_n \; - v_t \; + w_t & = 0 \\
& x_1, x_2, \ldots, x_n & \geq 0 \\
& v_1, v_2, \ldots, v_t & \geq 0 \\
& w_1, w_2, \ldots, w_t & \geq 0
\end{array}
\tag{8.6}
$$

where $x_1, x_2, \ldots, x_n,\ v_1, v_2, \ldots, v_t,\ w_1, w_2, \ldots, w_t$ are variables, z_1^{MAD} is mean absolute deviation of portfolio return, z_2 is expected (portfolio) return, and $d_{ij} = q_{ij} - \bar{q_j}$ for $i = 1, 2, \ldots, t$ and $j = 1, 2, \ldots, n$ where $\bar{q_j} = \frac{1}{t}\sum_{i=1}^{t} q_{ij}$. Model (8.6) has 2 objectives, $(m+t)$ constraints, and $(n+t+t)$ variables. Moreover, objective z_1^{MAD} is linear. Therefore (8.6) is a linear model and ADBASE can be used to obtain all the efficient extreme points. The nondominated frontiers of (8.5) and (8.6) can be compared in (standard deviation, expected return) space in the following steps.

1. The standard deviations and expected returns of the efficient extreme points of (8.6) can be obtained by plugging the efficient extreme points into the objective functions of (8.5).

2. Such standard deviations and expected returns can be connected and form a discretized representation of the nondominated frontier in (standard deviation, expected return) space.

3. The nondominated frontier of (8.5) can be computed by Hirschberger's procedure [40] in (standard deviation, expected return) space.

The nondominated frontiers of (8.5) and (8.6) can be compared in (mean absolute deviation, expected return) space in the following steps.

1. The efficient set of (8.5) can be computed by Hirschberger's procedure.

2. The mean absolute deviations and expected returns of this efficient set can be computed.

3. Such mean absolute deviations and expected returns form the nondominated frontier of (8.5) in (mean absolute deviation, expected return) space.

4. The nondominated vertices of (8.6) can be obtained as the criterion vectors of the efficient extreme points of (8.6).

5. The nondominated frontier of (8.6) can be obtained by connecting the neighboring nondominated vertices in (mean absolute deviation, expected return) space.

Model (8.6) needs a realized return matrix $(q_{ij})_{t\times n}$. If an investor does not have such a matrix, then based on $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ she or he can use "PORTSIM" function of Matlab [2] to generate a $(q_{ij})_{t\times n}$ matrix as follows

> "PORTSIM Random simulation of correlated asset returns. Simulates returns of NASSETS assets over NUMOBS consecutive observation intervals. Returns are simulated as the increments of constant drift and volatility Brownian processes.
> RetSeries = portsim(ExpReturn, ExpCovariance, NumObs, RetIntervals, NumSim)"

Based on (8.6), randomly generate the covariance matrix $\boldsymbol{\Sigma}$ and expected return vector $\boldsymbol{\mu}$ for a 46 security portfolio selection problem. Bring the matrix $\boldsymbol{\Sigma}$ into Matlab to generate a realized return matrix $(q_{ij})_{120\times 46}$. Take the constraints $\mathbf{Ax} \leq \mathbf{b}$ as $x_i \leq 0.1$ for $i = 1, 2, \ldots, 46$.

Then follow the steps to compare the nondominated frontiers. Model (8.5) has 40 turning points. Model (8.6) has 117 efficient extreme points. The nondominated frontiers of (8.5) and (8.6) in (standard deviation-expected return) space are in Figure 8.3 *top*. The nondominated frontier of (8.5) is black, with the dots on it representing the criterion vectors of the 40 turning points. The nondominated frontier of (8.6) is gray (the true color is red in color prints), with the dots on it representing the criterion vectors of the 117 efficient extreme points. The nondominated frontiers of (8.5) and (8.6) in (mean absolute deviation-expected return) space are in Figure 8.3 *bottom*. The nondominated frontier of (8.5) is black, with the dots on it representing the criterion vectors of the 40 turning points. The nondominated frontier of (8.6) is gray, with the dots on it representing the criterion vectors of the 117 efficient extreme points.

The two frontiers are close to each other in both graphs, indicating that (8.6) can be a good approximating to (8.5).

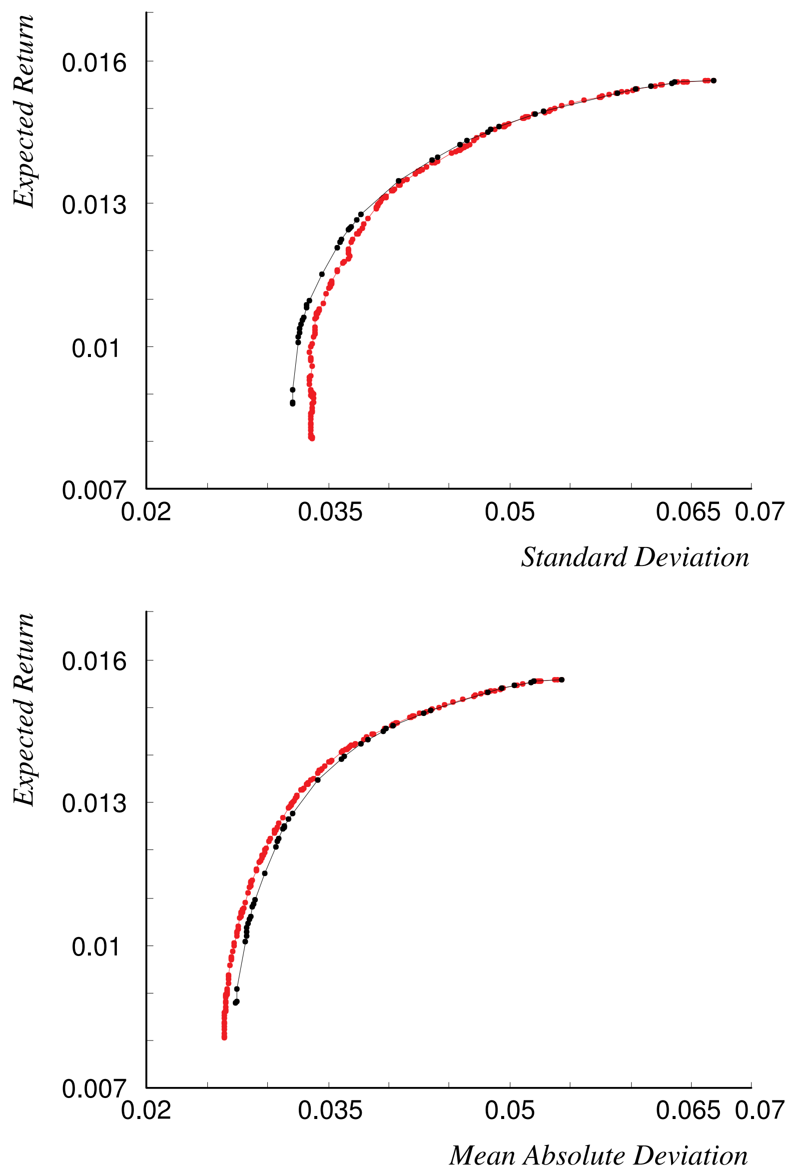

Figure 8.3: Nondominated frontiers of (8.5) and (8.6) in (standard deviation-expected return) space *top*, and nondominated frontiers of (8.5) and (8.6) in (mean absolute deviation-expected return) space *bottom* of a 46 security portfolio selection problem

Next the linearizing approach of (8.6) can be used in (8.1) as follows

$$
\begin{array}{lll}
\min\{ & \frac{1}{t}v_1 + \frac{1}{t}v_2 + \ldots + \frac{1}{t}v_t + \frac{1}{t}w_1 + \frac{1}{t}w_2 + \ldots + \frac{1}{t}w_t & = z_1^{MAD}\} \\
\min\{ & \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \pi_{ij} x_j & = z_2\} \\
\max\{ & \mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n & = z_3\} \\
\max\{ & \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_n x_n & = z_4\} \\
\text{s.t.} & a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & \leq b_1 \\
 & a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & \leq b_2 \\
 & \vdots & \\
 & a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n & \leq b_m \\
 & d_{11}x_1 + d_{12}x_2 + \ldots + d_{1n}x_n \; - v_1 \; + w_1 & = 0 \\
 & d_{21}x_1 + d_{22}x_2 + \ldots + d_{2n}x_n \; - v_2 \; + w_2 & = 0 \\
 & \vdots & \\
 & d_{t1}x_1 + d_{t2}x_2 + \ldots + d_{tn}x_n \; - v_t \; + w_t & = 0 \\
 & x_1, x_2, \ldots, x_n & \geq 0 \\
 & v_1, v_2, \ldots, v_t & \geq 0 \\
 & w_1, w_2, \ldots, w_t & \geq 0
\end{array}
\tag{8.7}
$$

where π_{ij} is the (i, j) element of matrix $\mathbf{\Pi}$. Model (8.7) has 4 objectives (one quadratic and three linear), $(m+t)$ constraints, and $(n+t+t)$ variables. Hirschberger, Qi, and Steuer [39] proposed an algorithm to compute the nondominated surface of one quadratic and several linear objective model. Therefore, this algorithm can be used to (8.7). The nondominated surface is piece-wise paraboloidal in $(z_1^{MAD}, z_2, z_3, z_4)$ space. The $\mathbf{x}^T\mathbf{\Sigma}\mathbf{x}$ values of this surface can be computed. Then, this surface can be plotted into the criterion space of (8.1), that is, (z_1, z_2, z_3, z_4) space. It must be pointed out that this surface is an approximation of the nondominated surface of (8.1). A discretized representation of the nondominated surface

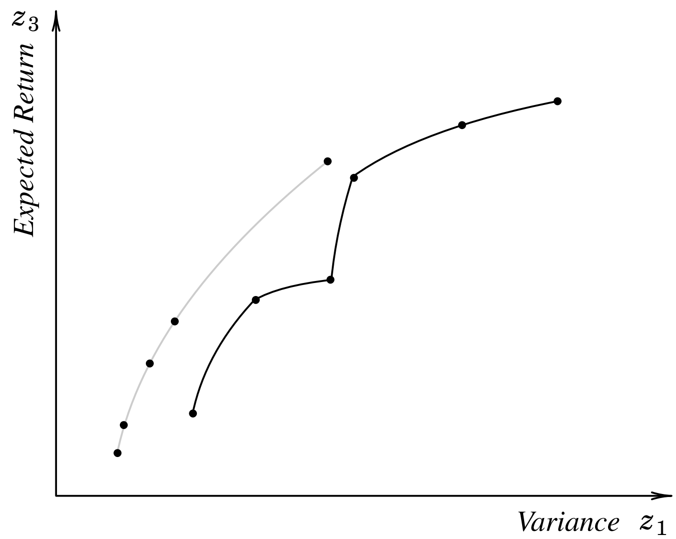

Figure 8.4: A discretized representation of the nondominated surface of (8.1) and an approximation of the nondominated surface of (8.1) by (8.7)

of (8.1) can be obtained in Section 8.1. Therefore, this approximation and the discretized representation can be compared in (z_1, z_2, z_3, z_4) space.

As a hypothesized example, in order to visualize, a discretized representation and an approximation are in (variance, expected return) space or (z_1, z_3) space of (8.1). The left frontier is the discretized representation. The five dots on it are the images of the maximizing solutions of (8.2) for some selected weighting vectors. The left frontier is gray, indicating an investor can not see it completely. What the investor sees is in fact the five dots on it. The right frontier is the approximation. The right frontier is black, indicating that an investor can see all of it. The six dots on it are the images of turning points of (8.7). Note that these two frontiers are purposely misaligned, because the discretized representation can be an uneven representation of the nondominated surface of (8.1).

Because the discretized representation of the nondominated surface of (8.1) and an approximation of the nondominated surface of (8.1) by (8.7) are in four dimensional space, and the representation and the approximation may not be aligned, a value function can be introduced to compare the representation and the approximation.

The advantages of this linearizing approach (8.7) are as follows.

1. The nondominated surface of (8.7) can be computed once and for all. Therefore, this approach can be faster than the repetitive optimization approach to obtain a discretized representation of the nondominated surface in Section 8.1.

2. This approach can still be effective in six objective (three quadratic and three linear) or more models.

Another usage of the approximation of the nondominated surface of (8.1) by (8.7) is that it can serve as the first generation (starting point) of multiple criteria evolutionary algorithms (for example, see Deb [19]). Taking advantage of the approximation, multiple criteria evolutionary algorithms can locate some points on the nondominated surface of (8.1) more easily.

Continue on the hypothesized example before. Randomly select several points on the right frontier as the first generation of multiple criteria evolutionary algorithms. After multiple criteria evolutionary algorithms stop, some points, for example $\mathbf{z}^1$, on the nondominated surface of (8.1) are found, while it is possible that some dominated points, for example $\mathbf{z}^2$, are also located.

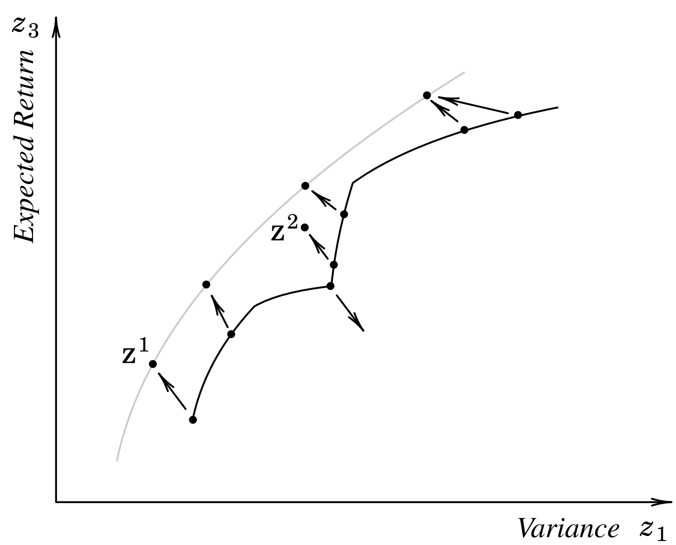

Figure 8.5: Taking an approximation of the nondominated surface of (8.1) by (8.7) as the first generation (starting point) of multiple criteria evolutionary algorithms

Chapter 9

Conclusion and Future Directions

In finance there is the classification between positive modelling and normative modelling defined by Sharpe [84] (2000, p. 2) and Hallerbach and Spronk [34] as

> "In a positive or descriptive model simplified assumptions are made in order to obtain a testable implication of the model. The validity of the model is evaluated according to the inability to reject the model's implication at some preset level of significance ... As long as the pricing relationship [1] is not falsified the model is accepted, irrespective whether the necessary assumptions are realistic or not ... In a conditional-normative model also simplified assumptions are made in order to obtain a tractable model. These assumptions relate to the preferences of the decision maker and to the representation of the set of choice alternatives. The object of the conditional-normative modelling is not to infer a testable implication but to obtain a decision rule" (2002, pp. 114-115)

In investment process in finance, positive models (for example Capital Asset Pricing Model) are needed to offer investors the exact prediction between risk and expected return. At the same time, normative models (for example portfolio selection) are also needed to help investors construct portfolios. Positive models typically prescribe for "average" investors and therefore may not be suitable for an individual investor. A closer look at the "average" investors reveals that this concept may not work well in practice, because investors are in different tax bracket, with distinct requirements towards dividend and capital gains, and in

[1] of CAPM

various stages of the life cycle. Confining all investors into an "average" investor may not be feasible in practice as also pointed out by Markowitz [63].

This argument is by no means intended to discount the insight and contribution of CAPM and other positive models. In fact, the insight and contribution are recognized. The point is that investment process in finance needs both positive models and normative models, so the fashion to over-emphasize positive models and try to apply positive models in the place of normative models can be inappropriate.

This dissertation tries to fit the need of an individual investor by modelling for a suitable-portfolio investor and proposing the solution methods. Meanwhile, the author acknowledges the difficulty of this approach as pointed out in Chapter 1. This dissertation belongs to normative modelling. The author hopes this dissertation could enrich the body of normative models, so that normative models can not only serve as a better starting point of positive models, but also proceed for their own sake.

The research projects of Steuer, Hirschberger, and Qi are in Figure 9.1. Each block represents one research project. Research projects "Suitable-portfolio investor", "Generating covariance matrix", "Quadratic parametric programming for portfolio selection", and "Compute nondominated surface of linear-constraint model" are approximately done. Research projects "Derive nondominated surface of short-sales allowed model" and "Compute nondominated surface of 2-quadratic and 2-linear model" are being undertaken. The other research projects are for future directions.

Start from the top of Figure 9.1, in "Suitable-portfolio investor" (Chapters 1 and 6), the idea of suitable-portfolio investor and the formulation are proposed. Follow the direction of "smooth cases". Smooth cases typically mean that all the variables are continuous and all the constraints are linear. In "Generating covariance matrix" (Chapter 5), a procedure to generate covariance with specified distributional characteristics is proposed. In "Quadratic parametric programming for portfolio selection" (Chapter 4), Hirschberger's procedure to compute nondominated frontiers is proposed. In "Computation comparison", Markowitz crit-

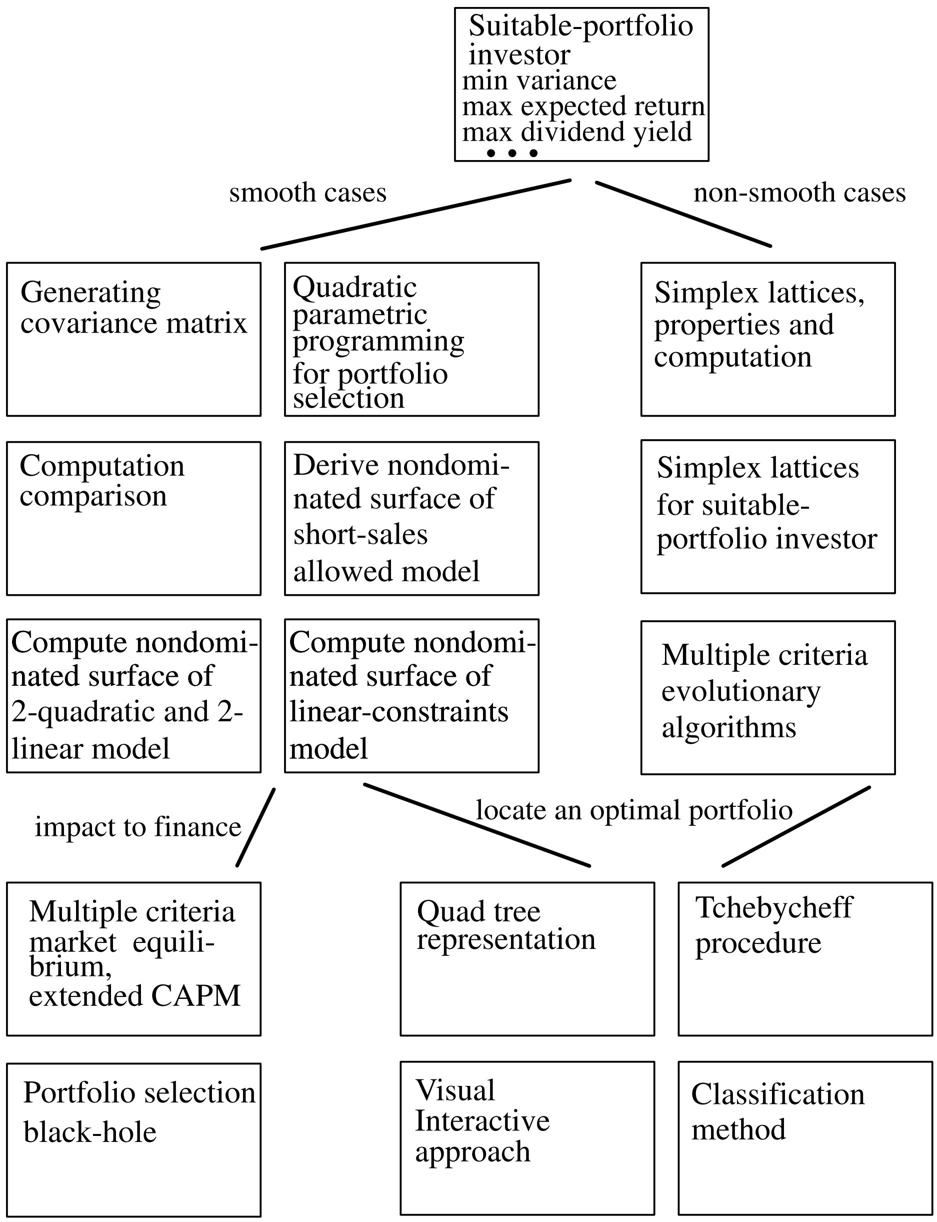

Figure 9.1: Research projects of Steuer, Hirschberger, and Qi

ical line algorithm, Hirschberger's procedure, and e-constraint approach will be compared. In "Derive nondominated surface of short-sales allowed model" (Chapter 7), closed-form results of (minimizing variance, maximizing expected return, and maximizing other linear objectives) models with the only constraint $\mathbf{1}^T\mathbf{x} = 1$ are obtained. In "Compute nondominated surface of 2-quadratic and 2-linear model" (Chapter 8), methods to approximate the nondominated surface of a (2-quadratic 2-linear) model with linear constraints are proposed. In "Compute nondominated surface of linear-constraint model" of Hirschberger, Qi and Steuer [39], a procedure to compute nondominated surfaces of (minimizing variance, maximizing expected return, and maximizing other linear objectives) models with linear constraints is proposed.

Follow the direction of "impact to finance". In "Multiple criteria market equilibrium, extended CAPM" (suggested in Chapter 7), Capital Asset Pricing Model for suitable-portfolio investor will be tried to obtain. In "Portfolio selection black-hole", a frequent observation that in large scale portfolio selection, most portfolios have a high probability to appear in a very small area of the feasible region Z will be studied. That is, there seems to exist a black-hole of portfolio selection. Although the black-hole is small, it contains most of the portfolios in Z.

Follow the direction of "non-smooth cases". Non-smooth cases mean that some of the variables are integer variables or semi-continuous variables. For example, the number of securities in portfolio involves integer variable. Semi-continuous variables can be used by fund managers. The managers will typically either hold a large amount of a stock or none of the stock. As a result, one would expect that nondominated surfaces in non-smooth cases are non-smooth. In "Simplex lattices, properties and computation", the properties and computation of simplex lattices will be studied. In "Simplex lattices for suitable-portfolio investor", simplex lattices can work as a discretized representation of a feasible region Z in small-scale portfolio selection. Simplex lattices can also be used in interactive weighted-sums approach. In “Multiple criteria evolutionary algorithms”, multiple criteria evolutionary algorithms will be used and tested to locate nondominated surface in non-smooth cases.

Follow the direction of “locate an optimal portfolio”. A suitable-portfolio investor may encounter millions of nondominated portfolios, either as a discretized representation of a nondominated surface or obtained from multiple criteria evolutionary algorithms. Searching through the millions of portfolios can be time-consuming. Employing “Quad tree representation” can significantly facilitate the searching process. “Tchebycheff procedure”, “Visual Interactive approach”, and “Classification method” are three interactive methods to locate a decision maker’s optimal or near-optimal solution. The three methods will be used and tested to locate a suitable-portfolio investor’s optimal or near-optimal portfolio.

Chapter 10

Appendix

10.1 Appendix for Chapter 1

10.1.1 Performance of Major Indices of Industrialized Countries

Example 1. Collected are monthly returns for the constituents of Standard & Poor's 500 Index from January 1997 to December 2002, represented by the round dots in Figure 10.1. [1] However, the sample covariance matrix is singular as pointed out by Markowitz [66] (2000, p. 43). The minimum-variance frontier with 50 randomly selected constituents serving as underlying assets is computed and portrayed as the solid line through which we can imagine the minimum-variance frontier with the complete constituents serving as underlying assets as the broken line to the left. Also shown are the performance of Dow Jones Industrial Average, Standard & Poor's 500 Index, and the value weighted index (more inclusive than Standard & Poor's 500 Index) offered by Wharton Research Data Services (WRDS) of the same time period, as squares denoted by $\mathbf{D}, \mathbf{S}$ and $\mathbf{V}$ respectively. These indices as market portfolio surrogates are much overwhelmed by the nondominated frontier, even lower than the minimum-variance portfolio. ◀

[1] Data source: the Center for Research in Security Prices (CRSP) through Wharton Research Data Service (WRDS) at the University of Pennsylvania and the website of Standard & Poor's 500 Index http://www.spglobal.com on July 20, 2003. Because of corporation merges, acquisitions, dropouts from the index, or possess missing values, 471 constituents with full records are utilized instead of 500 constituents.

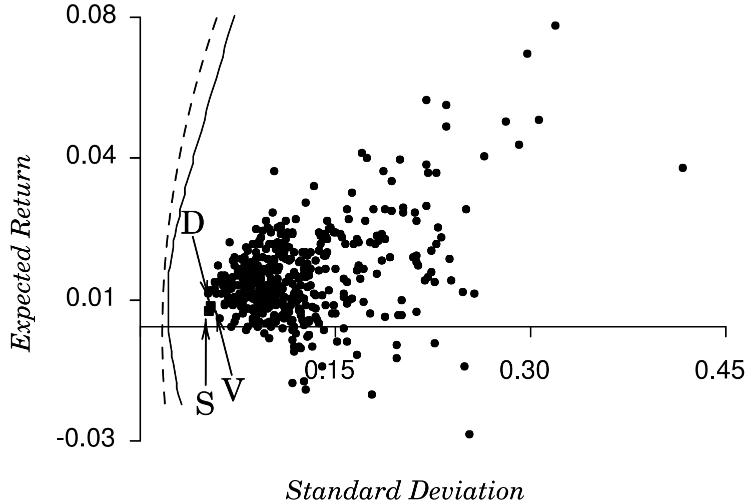

Figure 10.1: Graph of Example 1, minimum-variance frontier generated from Standard & Poor's 500 Index constituents and performance of market portfolio surrogates from 1997 to 2002

Example 2. Use the same method as Example 1 for Nikkei 225 Index and its constituents [2] from January 1997 to December 2002 (monthly returns) in Figure 10.2. Fairly apparent is the almost same result as in Example 1. ◀

Example 3. Use the same method as Example 1 for FTSE 100 Index and its constituents [3] from January 1997 to December 2002 (monthly returns) in Figure 10.3. Fairly apparent is the almost same result as in Example 1. ◀

Example 4. Use the same method as Example 1 for CAC 40 Index and its constituents [4] from January 1997 to December 2002 (monthly returns). Fairly apparent is the almost same result as in Example 1. ◀

[2]Data source: DataStream of Thomson Financial, September 6, 2003. Due to the same reason as in Example 1, 214 constituents with full records are utilized instead of 225 constituents.

[3]Data source: DataStream of Thomson Financial, September 6, 2003. Due to the same reason as in Example 1, 86 constituents with full records are utilized instead of 100 constituents.

[4]Data source: DataStream of Thomson Financial, September 6, 2003. Due to the same reason as in Example 1, 34 constituents with full records are utilized instead of 40 constituents.

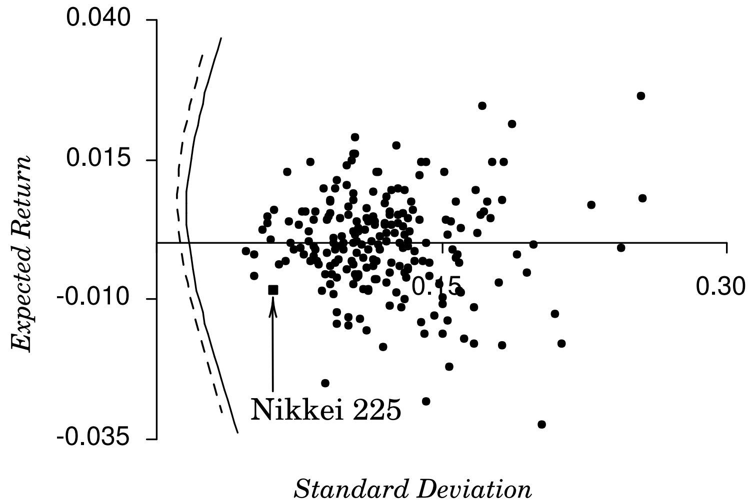

Figure 10.2: Graph of Example 2, minimum-variance frontier generated from Nikkei 225 Index constituents and performance of market portfolio surrogates from 1997 to 2002

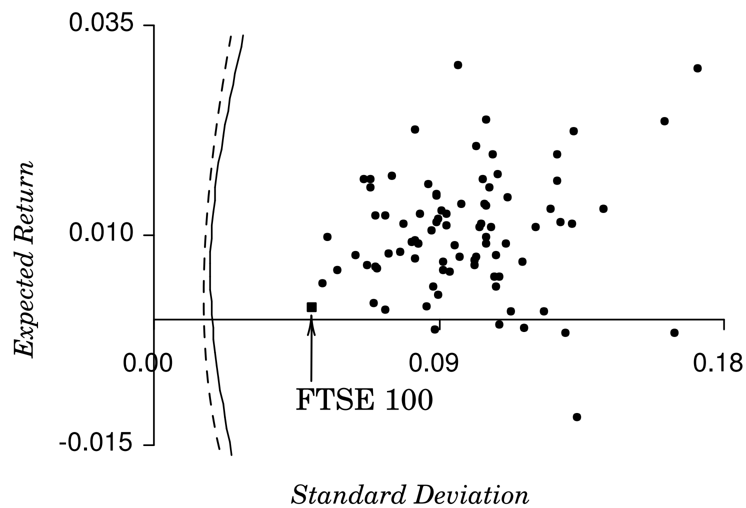

Figure 10.3: Graph of Example 3, minimum-variance frontier generated from FTSE 100 Index constituents and performance of market portfolio surrogates from 1997 to 2002

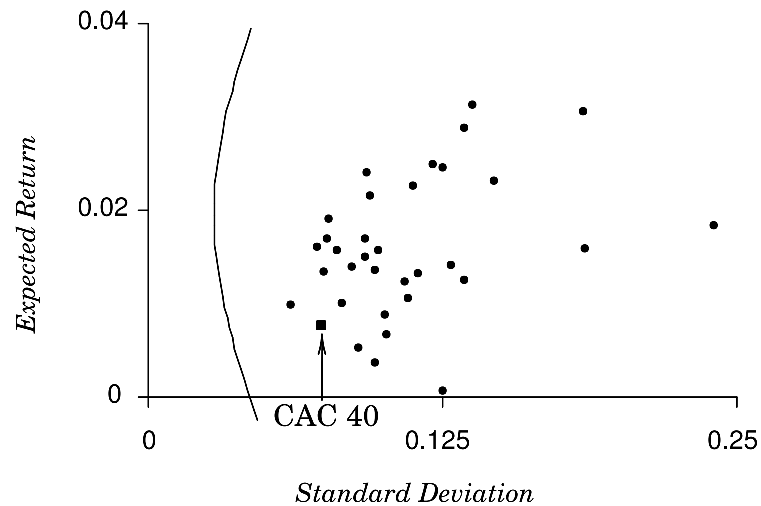

Figure 10.4: Graph of Example 4, minimum-variance frontier generated from CAC 40 Index constituents and performance of market portfolio surrogates from 1997 to 2002

Example 5. Use the same method as Example 1 for DAX 30 Index and its constituents [5] from January 1997 to December 2002 (monthly returns). Fairly apparent is the almost same result as in Example 1. ◀

10.1.2 Feasible Region Z with Ellipsoidal Boundary and Projection

Lemma 10.1. *For a multiple criteria optimization model below*

$$\begin{aligned} \min \{f_1(\mathbf{x}) = z_1\} \\ \max \{f_2(\mathbf{x}) = z_2\} \\ \vdots \\ \max \{f_k(\mathbf{x}) = z_k\} \\ s.t. \quad \mathbf{x} \in S \end{aligned}$$

[5]Data source: DataStream of Thomson Financial, September 6, 2003. Due to the same reason as in Example 1, 25 constituents with full records are utilized instead of 30 constituents.

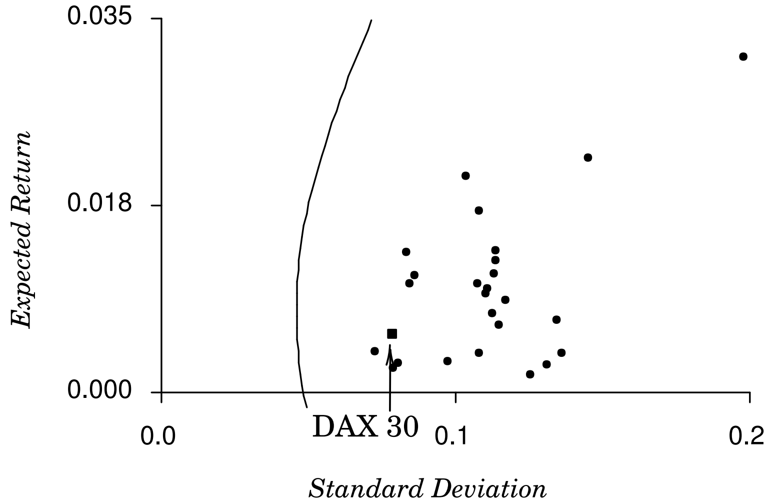

Figure 10.5: Graph of Example 5, minimum-variance frontier generated from DAX 30 Index constituents and performance of market portfolio surrogates from 1997 to 2002

the boundary of the feasible region Z is an ellipsoid. That is,

$$Z = \{(z_1, z_2, \ldots, z_k) \in \mathbb{R}^k \mid \frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_k^2}{a_k^2} \leq 1\}$$

for fixed $a_1 \geq 0, a_2 \geq 0, \ldots, a_k \geq 0$, then

(i) *the nondominated set is*

$$N = \{(z_1, z_2, \ldots, z_k) \in \mathbb{R}^k \mid \frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_k^2}{a_k^2} = 1, z_1 \leq 0, z_2 \geq 0, \ldots, z_k \geq 0\}$$

item the projection of N along the axes of $z_3, z_4, \ldots, z_k$ onto $z_1 - z_2$ space is

$$\{(z_1, z_2) \in \mathbb{R}^k \mid \frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} \leq 1, z_1 \leq 0, z_2 \geq 0\}$$

Proof. : Denote a subset of Z as

$$Q = \{(z_1, z_2, \ldots, z_k) \in \mathbb{R}^k \mid \frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_k^2}{a_k^2} \leq 1, z_1 \leq 0, z_2 \geq 0, \ldots, z_k \geq 0\}$$

For any $(z_1, z_2, \ldots, z_k) \in Z$ but $(z_1, z_2, \ldots, z_k) \notin Q$, then $(-|z_1|, |z_2|, \ldots, |z_k|) \in Q$ and $(-|z_1|, |z_2|, \ldots, |z_k|)$ dominates $(z_1, z_2, \ldots, z_k)$. That is, any point in Z but not in Q is dominated by some point in Q. Then the nondominated set must be a subset of Q.

For any $(z_1, z_2, \ldots, z_k) \in Q$ but $(z_1, z_2, \ldots, z_k) \notin N$ i.e. $\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_k^2}{a_k^2} < 1$ then rearrange terms as $1 - (\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_{k-1}^2}{a_{k-1}^2}) > \frac{z_k^2}{a_k^2}$ then $a_k\sqrt{1 - (\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_{k-1}^2}{a_{k-1}^2})} > z_k$. Point $(\, z_1, z_2, \ldots, z_{k-1}, a_k\sqrt{1 - (\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_{k-1}^2}{a_{k-1}^2})}\ )$ dominates point $(z_1, z_2, \ldots, z_k)$ and $(\, z_1, z_2, \ldots, z_{k-1}, a_k\sqrt{1 - (\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_{k-1}^2}{a_{k-1}^2})}\ ) \in N$. That is any point in Q but not in N is dominated by some point in N.

By the discussion above, **(i)** holds. For any $(z_1, z_2, \ldots, z_k) \in N$, $(z_3, z_4, \ldots, z_k)$ drop out by the projection. What left is $z_1 \leq 0, z_2 \geq 0$ and $\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} \leq 1$ (otherwise $\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} > 1$ violates $\frac{z_1^2}{a_1^2} + \frac{z_2^2}{a_2^2} + \ldots + \frac{z_k^2}{a_k^2} \leq 1$.) Therefore **(ii)** holds. $\square$

Bibliography

[1] *SAS/GRAPH User's Guide*. SAS Institute, Inc., Cary, North Carolina, 1985.

[2] *MATLAB The Language of Technical Computing, Using MATLAB Version 6.* The MathWorks, Inc., Natick, Massachusetts, 2000.

[3] *SAS Release 8.02 Online Manual*. SAS Institute, Inc., Cary, North Carolina, 2003.

[4] E. Ballestero and D. Pla-Santamaria. Portfolio selection on the Madrid exchange: A compromise programming model. *International Transactions in Operational Research*, 10(1):33–51, 2003.

[5] E. Ballestero and C. Romero. Portfolio selection: A compromise programming solution. *Journal of the Operational Research Society*, 47(11):1377–1386, 1996.

[6] C. A. Bana e Costa and J. O. Soares. Multicriteria approaches for portfolio selection: An overview. *Review of Financial Markets*, 4(1):19–26, 2001.

[7] V. Belton and T. J. Stewart. *Multiple Criteria Decision Analysis-an Integrated Approach*. Kluwer Academic Publishers, Boston, 2002.

[8] F. Black. Capital market equilibrium with restricted borrowing. *The Journal of Business*, 45(3):444–455, 1972.

[9] Z. Bodie, A. Kane, and A.J. Marcus. *Essentials of Investment*. McGraw-Hill/Irwin, Boston, 5th edition, 2002.

[10] Z. Bodie, A. Kane, and A.J. Marcus. *Investment*. McGraw-Hill/Irwin, Boston, 5th edition, 2002.

[11] A. Bouri, J. M. Martel, and H. Chabchoub. A multi-criterion approach for selecting attractive portfolio. *Journal of Multi-Criteria Decision Analysis*, 11(4-5):269–277, 2002.

[12] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, New York, New York, 2nd edition, 1991.

[13] R. Caballero, E. Cerdá, M. M. Muñoz, L. Rey, and I. M. Stancu-Minasian. Efficient solution concepts and their relations in stochastic multiobjective programming. *Journal of Optimization Theory and Applications*, 110(1):53–74, 2001.

[14] T.-J. Chang, N. Meade, J. E. Beasley, and Y. M. Sharaiha. Heuristics for cardinally constrained portfolio optimisation. *Computers & Operations Research*, 27:1271–1302, 2000.

[15] V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11, 1993.

[16] G. Chow. Portfolio selection based on return, risk, and relative performance. *Financial Analysts Journal*, pages 54–60, March-April, 1995.

[17] G. Colson and Ch. DeBruyn. An integrated multiobjective portfolio management system. *Mathematical and Computer Modelling*, 12(10-11):1359–1381, 1989.

[18] P. I. Davies and N. J. Higham. Numerically stable generation of correlation matrices and their factors. *BIT*, 40(4):640–651, 2000.

[19] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons, New York, 2001.

[20] C. Dominiak. An application of interactive multiple objective goal programming on the Warsaw stock exchange. In R. Caballero, F. Ruiz, and R. E. Steuer, editors, *Lecture*

Notes in Economics and Mathematical Systems, vol. 455, pages 66–74. Springer-Verlag, Berlin, 1997.

[21] M. Doumpos, M. Spanos, and C. Zopounidis. On the use of goal programming techniques in the assessment of financial risks. *Journal of Euro-Asian Management*, 5(1):83–100, 1999.

[22] M. Ehrgott. *Multicriteria Optimization, Volume 491 of Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 2000.

[23] M. Ehrgott. *Multicriteria Optimization, Volume 491 of Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 2000.

[24] M. Ehrgott, K. Klamroth, and C. Schwehm. An MCDM approach to portfolio optimization. *European Journal of Operational Research*, 155(3):752–770, 2004.

[25] E. J. Elton, M. J. Gruber, S. J. Brown, and W. Goetzmann. *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, New York, 6th edition, 2002.

[26] E. J. Elton, M. J. Gruber, and M. W. Padberg. Simple criteria for optimal portfolio selection. *The Journal of Finance*, 31(5):1341–1357, 1976.

[27] E. J. Elton, M. J. Gruber, and M. W. Padberg. Simple criteria for optimal portfolio selection with upper bounds. *Operations Research*, 25(6):952–967, 1977.

[28] E. J. Elton, M. J. Gruber, and M. W. Padberg. Simple criteria for optimal portfolio selection: tracing out the efficient frontier. *The Journal of Finance*, 33(1):296–302, 1978.

[29] C. D. Feinstein and M. N. Thapa. Notes: A reformulation of a mean-absolute deviation portfolio optimization model. *Management Science*, 39(4):1552–1553, 1993.

[30] K. L. Fisher and M. Statman. Investment advice from mutual fund companies. *The Journal of Portfolio Management*, 24(1):9–25, 1997.

[31] A. M. Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of Matheatical Analysis and Applications*, 22(3):618–630, 1968.

[32] J. Guddat. Stability in convex quadratic programming. *Operationsforschung und Statistik*, 7:223–245, 1976.

[33] J. B. Guerard and A. Mark. The optimization of efficient portfolios: The case for an R&D quadratic term. *Research in Finance*, 20:217–247, 2003.

[34] W. G. Hallerbach and J. Spronk. A multidimensional framework for financial-economic decisions. *Journal of Multi-Criteria Decision Analysis*, 11(3):111–124, 2002.

[35] W. G. Hallerbach and J. Spronk. The relevance of MCDM for financial decisions. *Journal of Multi-Criteria Decision Analysis*, 11(4-5):187–195, 2002.

[36] N. J. Higham. Matrix nearness problems and applications. In M. J. C. Grover and S. Barnett, editors, *Applications of Matrix Theory*, pages 1–27. Oxford University Press, Oxford, 1989.

[37] N. J. Higham. Computing the nearest correlation matrix – a problem for finance. *IMA Journal of Numerical Analysis*, 22:329–343, 2002.

[38] I. D. Hill, R. Hill, and R.L. Holder. Fitting Johnson curves by moments. *Applied Statistics*, 25(2):180–189, 1976.

[39] M. Hirschberger, Y. Qi, and R. E. Steuer. "Computation of efficient surfaces in tri-criterion portfolio optimization," Department of Banking and Finance, University of Georgia, Athens, March, 2004.

[40] M. Hirschberger, Y. Qi, and R. E. Steuer. "Quadratic parametric programming for portfolio selection with random problem generation and computational experience," Department of Banking and Finance, University of Georgia, Athens, March, 2004.

[41] M. Hirschberger, Y. Qi, and R. E. Steuer. "Randomly generating portfolio-selection covariance matrices with specified distributional characteristics," Department of Banking and Finance, University of Georgia, Athens, February, 2004.

[42] R. B. Holmes. On random correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 12(2):239–272, 1991.

[43] C. F. Huang and Litzenberger R. H. *Foundations for Financial Economics.* Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[44] Ch. Hurson and N. Ricci. Multicriteria decision making and portfolio management with arbitrage pricing theory. In C. Zopounidis, editor, *Operational Tools in The Management of Financial Risks*, pages 31–55. Kluwer Academic Publishers, Dordrecht, 1998.

[45] Ch. Hurson and C. Zopounidis. On the use of multi-criteria decision aid methods to portfolio selection. *Journal of Euro-Asian Management*, 1(2):69–94, 1995.

[46] J. E. Ingersoll. *Theory of Financial Decision Making.* Rowman & Littlefield, 1987.

[47] V. Jog, I. Kaliszewski, and W. Michalowski. Using attribute trade-off information in investment. 1999.

[48] N. L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36:149–176, 1949.

[49] C. P. Jones. *Investments: Analysis and Management.* John Wiley & Sons, New York, 2000.

[50] B. Kolman. *Introductory Linear Algebra With Applications.* Prentice-Hall, Upper Saddle River, New Jersey, 1997.

[51] H. Konno, H. Shirakawa, and H. Yamazaki. A mean-absolute deviation-skewness portfolio optimization model. *Annals of Operations Research*, 45(1-4):205–220, 1993.

[52] H. Konno and K.-I. Suzuki. A mean-variance-skewness portfolio optimization model. *Journal of the Operations Research Society of Japan*, 38(2):173–187, 1995.

[53] H. Konno and H. Yamazaki. Mean-absolute deviation portfolio optimization model and its application to the tokyo stock market. *Management Science*, 37(5):519–531, 1991.

[54] Y. Kroll, H. Levy, and H. M. Markowitz. Mean variance versus direct utility maximization. *The Journal of Finance*, 39(1):47–61, 1984.

[55] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492. University of California Press, Berkeley, CA, 1951.

[56] S. P. Lin and R. B. Bendel. Algorithm AS213: Generation of population correlation matrices with specified eignvalues. *Applied Statistics*, 34:193–198, 1985.

[57] F. H. Liou. "A Routine for Generating Grid Point Defined Weighting Vectors," Master's Degree thesis, University of Georgia, USA, 1985.

[58] A. W. Lo, C. Petrovz, and M. Wierzbicki. "It's 11pm-do you know where your liquidity is? the mean-variance-liquidity frontier," Sloan School of Management, Mmassachusetts Institute of Technology, Cambridge, Ma, January, 2003.

[59] Y.-D. Lyuu. *Financial Engineering and Computation: Principles, Mathematics, Algorithms.* Cambridge University Press, Cambridge, 2002.

[60] R. Mansini, W. Ogryczak, and M. G. Speranza. On LP solvable models for portfolio selection. *Informatica*, 14(1):37–62, 2003.

[61] H. M. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[62] H. M. Markowitz. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly*, 3:111–133, 1956.

[63] H. M. Markowitz. *Portfolio Selection: Efficient Diversification in Investments.* John Wiley & Sons, New York, 1st edition, 1959.

[64] H. M. Markowitz. Normative portfolio analysis: past, present, and future. *Journal of Economics and Business*, 42:99–103, 1990.

[65] H. M. Markowitz. Foundations of portfolio selection. *The Journal of Finance*, 46(2):469–477, 1991.

[66] H. M. Markowitz and G. P. Todd. *Mean-Variance Analysis in Portfolio Choice and Capital Markets.* Frank J. Fabozzi Associates, New Hope, Pennsylvania, 2000.

[67] H. M. Markowitz, P. Todd, G. Xu, and Y. Yamane. Computation of mean-semivariance efficient sets by the critical line algorithm. *Annals of Operations Research*, 45(1-4):307–318, 1993.

[68] H. M. Markowitz, P. Todd, G. Xu, and Y. Yamane. Fast computation of mean-variance efficient sets using historical covariances. *Journal of Financial Engineering*, 1(2):117–132, 1993.

[69] H. P. Mayo. *Investments: An Introduction.* Harcourt, Fort Worth, Texas, 2000.

[70] R. C. Merton. An analytical derivation of the efficient portfolio frontier. *The Journal of Financial and Quantitative Analysis*, 7(4):1851–1872, 1972.

[71] W. Michalowski and W. Ogryczak. Extending the MAD portfolio optimization model to incorporate downside risk aversion. *Naval Research Logistics*, 48:185–200, 2001.

[72] K. Miettinen. *Nonlinear Multiobjective Optimization.* Kluwer Academic Publishers, Boston, 1999.

[73] W. Ogryczak. Multiple criteria linear programming model for portfolio selection. *Annals of Operations Research*, 97:143–162, 2000.

[74] S. Pafka and I. Kondor. Noisy covariance matrices and portfolio optimization II. *Physica*, A(319):487, 2003. e-print cond-mat/0205119.

[75] S. Pafka, M. Potters, and I. Kondor. "Exponential Weighting and Random Matrix Theory Based Filtering of Financial Covariance Matrices for Portfolio Optimization," Department of Physics of Complex Systems, Eötvös University, Budapest, e-print cond-mat/0402573, 2004.

[76] J. Pang. A new and efficient algorithm for a class of portfolio selection problems. *Operations Research*, 28(3):754–767, 1980.

[77] M. Arenas Parra, A. Bilbao Terol, and M. V. Rodríguez Uría. A fuzzy goal programming approach to portfolio selection. *European Journal of Operational Research*, 133(2):287–297, 2001.

[78] A.F. Perold. 1984. *Management Science*, 30(10):1143–1160, 1984.

[79] A. Prékopa. *Stochastic Programming.* Kluwer Academic Publishers, Dordrecht, 1995.

[80] S. Sayin. An algorithm based on facial decomposition for finding the efficient set in multiple objective linear programming. *Operations Research Letters*, 19(2):87–94, 1996.

[81] H. Scheffe. Experiments with mixtures. *Journal of the Royal Statistical Society. Series B(Methodology)*, 20(2):344–360, 1958.

[82] W. F. Sharpe. "Optimal portfolios without bounds on holdings," online lecture, http://www.stanford.edu/∼wfsharpe/mia/opt, Graduate School of Business, Stanford University, Stanford, CA, 2004.

[83] W. F. Sharpe. Capital asset prices: a theory of market equilibrium. *The Journal of Finance*, 19:425–442, 1964.

[84] W. F. Sharpe. *Portfolio Theory and Capital Markets.* McGraw-Hill, New York, 2000.

[85] T. Shifrin. *Abstract Algebra - A Geometric Approach.* Prentice-Hall, Englewood Cliffs, New Jersey, 1996.

[86] J. Spronk and W. G. Hallerbach. Financial modelling: Where to go? with an illustration for portfolio management. *European Journal of Operational Research*, 99(1):113–127, 1997.

[87] I. M. Stancu-Minasian. *Stochastic Programming with Multiple Objective Functions.* D. Reidel Publishing Company, Dordrecht, 1984.

[88] R. E. Steuer. "ADBASE: A multiple objective linear programming solver for all efficient extreme points and all unbounded efficient edges," Terry College of Business, University of Georgia, Athens, Georgia, 2003.

[89] R. E. Steuer. *Multiple Criteria Optimization: Theory, Computation and Application.* John Wiley & Sons, New York, 1986.

[90] R. E. Steuer, Y. Qi, and M. Hirschberger. "Suitable-portfolio investors, nondominated frontier sensitivity, and the effect of multiple objectives on standard portfolio selection," Department of Banking and Finance, University of Georgia, Athens, January, 2004.

[91] R. E. Steuer and A. T. Schuler. An interactive multiple objective linear programming approach to a problem in forest management. *Operations Research*, 26(2):254–269, 1978.

[92] B. K. Stone. A linear programming formulation of the general portfolio selection problem. *The Journal of Financial and Quantitative Analysis*, 8(4):621–636, 1973.

[93] M. Tamiz, R. Hasham, D. F. Jones, B. Hesni, and E. K. Fargher. A two staged goal programming model for portfolio selection. In M. Tamiz, editor, *Lecture Notes in Economics and Mathematical Systems,* vol. 432, pages 386–399, 1996.

[94] J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, New Jersey, 1953.

[95] P. Wolfe. The simplex method for quadratic programming. *Econometrica*, 27(3):382–398, 1959.

[96] G. Y. Yu. A multiple criteria approach to choosing an efficient stock portfolio at the Helsinki stock exchange. *Journal of Euro-Asian Management*, 3(2):53–85, 1997.

[97] M. Zeleny. Multidimensional measure of risk: The prospect ranking vector. In S. Zionts, editor, *Lecture Notes in Economics and Mathematical Systems,* vol. 155, pages 529–548. Springer-Verlag, Heidelberg, 1977.

[98] W. T. Ziemba. *The Stochastic Programming Approach to Asset, Liability, and Wealth Management.* Research Foundation of the AIMR, Charlottesville, Virginia, 2003.

[99] C. Zopounidis and M. Doumpos. INVESTOR: A decision support system based on multiple criteria for portfolio selection and composition. In A. Colorni, M. Paruccini, and B. Roy, editors, *A-MCD-A Aide Multi Critere a la Decision - Multiple Criteria Decision Aiding*, pages 371–381. European Commission Joint Research Centre, Brussels, 2000.