

Journal of Business Economics manuscript No.
(will be inserted by the editor)

Overviewing the transition of Markowitz bi-criterion portfolio selection to tri-criterion portfolio selection

Ralph E. Steuer · Maximilian Wimmer ·
Markus Hirschberger

Preprint as of 20Oct2012

The final publication is available at www.springerlink.com

Abstract Over sixty years ago, Markowitz introduced the mean-variance efficient frontier to finance. While mean-variance is still the predominant model in portfolio selection, it has endured many criticisms. One serious one is that it does not allow for additional criteria. The difficulty is that the efficient frontier becomes a surface. With it now possible to compute such a surface, we provide an overview on how Markowitz's risk-return (bi-criterion) portfolio selection can be extended to tri-criterion portfolio selection. With a focus on the geometry of the extension, many graphs are used to illustrate.

Keywords Efficient frontiers · Nondominated surfaces · Parametric quadratic programming · Portfolio selection · Stability sets · Hyperbolic platelets

1 Introduction

In finance there has long been the problem of portfolio selection—how to invest in a collection of securities for maximum benefit. Typically, this has been taken to mean to maximize the stochastic objective (i.e., random variable) of portfolio return. The task is difficult because how to maximize a stochastic quantity requires an interpretation. One possibility is to replace the problem having a stochastic objective with an equivalent problem having one or more deterministic objectives (Caballero, 2003). In his seminal work of (1952, 1956, 1959), and as reiterated in his books of (1987,

R. E. Steuer
Department of Finance, University of Georgia, Athens, Georgia 30602-6253 USA
E-mail: rsteuer@uga.edu

M. Wimmer
Department of Finance, University of Regensburg, 93040 Regensburg, Germany
E-mail: maximilian.wimmer@ur.de

M. Hirschberger
Munich Re, 80802 Munich, Germany

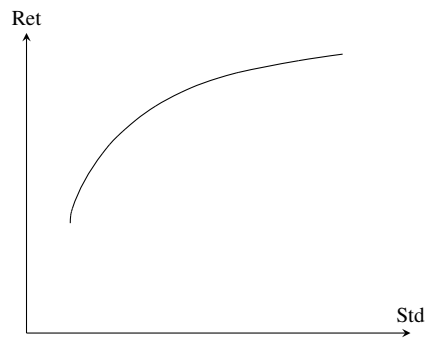


Fig. 1 Typical nondominated frontier with, in this paper, standard deviation on the horizontal axis.

2000), Markowitz did just that. In his mean-variance approach, the stochastic objective is replaced by the two deterministic objectives of to maximize expected return and to minimize variance. The appeal of his equivalent problem is that with its two objectives it focuses on the conflicting criteria of risk and return, where *return* is measured in terms of the expected value of the portfolio return random variable and *risk* is measured in terms of the square root of the variance of the portfolio return random variable. Variance's role in the approach is that produces the same solutions as standard deviation but it is by means of variance that they are easier to obtain.

After data are obtained and constraints are constructed to equip the bi-criterion nature of his equivalent problem, Markowitz's approach can be thought of as existing in two stages. In the first stage, with variance (a monotonic stand-in for standard deviation) as one of its objective functions and expected return as the other, the equivalent problem is solved for its efficient frontier (which we henceforth call the *nondominated frontier*). With each portfolio having a criterion vector describing its risk-return attributes, the nondominated frontier is the plot of all portfolio criterion vectors that are nondominated in *standard deviation, expected return* space. As a nonnegatively sloped, concave curve, a typical nondominated frontier is as in Figure 1. With standard deviation for risk, the nondominated frontier shows how, if one wants low risk, one must be prepared to accept a low rate of return, and how, if one wishes a high rate of return, one must be prepared to endure high risk. A property of Markowitz's nondominated frontier is that, under the assumption that a decision maker's utility function is increasing and quadratic or that security returns follow a multivariate normal distribution, it is precisely the set of all *contenders for optimality*. That is, if a portfolio's criterion vector can possibly optimize a decision maker's utility function, it will be on the frontier, and if it can not, it will not be on the frontier. Thus, a person's most preferred point on the nondominated frontier is that person's optimal solution to the problem.

In the second stage, the graph of the nondominated frontier is displayed before the decision maker for the decision maker to select his or her most preferred criterion vector on it. From a decision-making perspective, displaying the nondominated frontier so that all nondominated points can be viewed at once is an important part of the approach. This is because in portfolio selection it is usually not possible for a de-

cision maker to recognize an optimal solution in the absolute. Normally, the solution that gets selected is not selected because of its greatness. It is only selected because everything else is worse.

With this the general background situation, it will be our preference in this paper to call a decision maker's final nondominated choice his or her *most preferred* solution. Our intention is to shy away from an overuse of the term "optimal" in this context because it seems to us a bit presumptuous to apply the term to a solution that a decision maker may not like, but is forced to accept given the current investment reality, like for instance 1% on US Treasuries at the moment¹. It just seems that "most preferred" is a softer term to use. Consequently, to allow for a final solution to be arrived at by means of a process of elimination, being able to present the nondominated frontier all at once is ideal. In this way, with the nondominated frontier containing only all contenders for optimality, a decision maker need not ever be haunted by the worry that there might be other candidates that should be shown, because there are none.

While mean-variance is still the predominant model in portfolio selection (He and Zhou, 2011), it has endured many criticisms. As seen from the analysis in Steuer and Na (2003), one serious criticism is that it does not allow for additional criteria. The stumbling block is that the nondominated frontier becomes a *surface*, and, until very recently, it has not been possible to compute such a surface. But by utilizing the CIOS procedure from Hirschberger et al (2013) for the computation of nondominated surfaces, in this paper we are able to show how the two stages of Markowitz portfolio selection

- (a) computing the nondominated set
- (b) displaying it before the decision maker all at once

can be extended in full robustness to tri-criterion portfolio selection. In this way, Markowitz-style portfolio selection can be adapted to handle, beyond risk and return, an additional linear criterion (such as liquidity, dividends, social responsibility, and so forth).

The paper is organized as follows. In Section 2, Markowitz's model for portfolio selection is introduced, and after some preliminaries in Section 3, different methods for computing the nondominated frontier are discussed in Sections 4 and 5. To set the stage for tri-criterion portfolio selection, parametric quadratic programming is reviewed in Section 6. After discussing additional criteria in portfolio selection in Section 7, the model for use in tri-criterion portfolio selection is presented in Section 8. Section 9 covers multi-parametric quadratic programming by which tri-criterion nondominated surfaces can be computed. After some tri-criterion nondominated surface graphical illustrations in Section 10, the paper concludes in Section 11.

¹ Or, in more colorful terminology, the solution is accepted simply because it is the "cleanest dirty shirt" which is a quote from an internet interview with William H. Gross, co-chief investment officer, Pacific Investment Management, on June 28, 2012.

2 Bi-criterion formulation

We begin a review of Markowitz portfolio selection. However our treatment is somewhat nonstandard. This is because our end purpose is not bi-criterion portfolio selection. It is tri-criterion portfolio selection—and certain features that might not otherwise attract attention must be brought out to accommodate the extension. In bi-criterion format, the problem of Markowitz portfolio selection is as in (M)

$$\begin{aligned} \min \{ \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} \} & \quad \text{variance} & \text{(M)} \\ \max \{ \boldsymbol{\mu}^T \mathbf{x} \} & \quad \text{expected return} \\ \text{s.t. } & \mathbf{x} \in S \end{aligned}$$

where

- (a) $S \subset \mathbb{R}^n$ is the feasible region in *decision space* defined by

$$\begin{aligned} \mathbf{H}\mathbf{x} &= \mathbf{d} \\ \mathbf{G}\mathbf{x} &\leq \mathbf{b} \\ \mathbf{x} &\in [\mathbf{0}, \boldsymbol{\beta}] \end{aligned}$$

- (b) n is the number of securities in the pool of securities eligible for investment
 (c) $\mathbf{x} = (x_1, \dots, x_n)$. Provided that the x_i sum to one, \mathbf{x} is said to be a *portfolio*.
 When this is the case, x_i is the proportion of a portfolio invested in security i .
 (d) $\boldsymbol{\mu}$ is an n -vector of individual security expected returns
 (e) $\boldsymbol{\Sigma}$ is an $n \times n$ covariance matrix of the security returns

To ensure that \mathbf{x} is a portfolio in our analyses, it is assumed that the equality constraints $\mathbf{H}\mathbf{x} = \mathbf{d}$ always contain the condition $\mathbf{1}^T \mathbf{x} = 1$. If there are industry or sector constraints like at least 5% of a portfolio is to be invested in the pharmaceutical industry or no more than 20% of a portfolio is to be invested in oil, they would be housed in the inequality constraints of $\mathbf{G}\mathbf{x} \leq \mathbf{b}$. Although $\mathbf{x} \in [\mathbf{0}, \boldsymbol{\beta}]$ might give the impression that short positions or lower bounds greater than zero are not to be treated, this is not the case. Should there be any such security lower bounds, they are handled by a translation of the coordinate system. Let $\boldsymbol{\ell} \in \mathbb{R}^n$ be a problem's starting vector of lower bounds (long and short) on holdings. Substituting $\mathbf{x} - \boldsymbol{\ell}$ for \mathbf{x} throughout the formulation translates the coordinate system's origin to $\boldsymbol{\ell}$, with the new \mathbf{x} satisfying $\mathbf{x} \in [\mathbf{0}, \boldsymbol{\beta}]$. After solution, all \mathbf{x} from (M) are then translated back to original form.

Normally there are not many constraints in $\mathbf{H}\mathbf{x} = \mathbf{d}$ and $\mathbf{G}\mathbf{x} \leq \mathbf{b}$. Even a problem in which \mathbf{G} is vacuous would not raise eyebrows. Thus, apart from the $[\mathbf{0}, \boldsymbol{\beta}]$ box constraints on \mathbf{x} , a characteristic of portfolio selection problems is that in most cases they can be anticipated to possess many times as many variables (i.e., securities) as constraints in $\mathbf{H}\mathbf{x} = \mathbf{d}$ and $\mathbf{G}\mathbf{x} \leq \mathbf{b}$.

Being a multiple objective programming problem, mathematically, the solution of (M) is the set of all \mathbf{x} -vectors in S that produce nondominated vectors in *criterion space*. For books on multiple objective programming, see Steuer (1986), Meittinen (1999) or Ehrgott (2005). While some authors choose *variance, expected return* space for criterion space in portfolio selection, since we have stated that we will be displaying nondominated frontiers in *standard deviation, expected return* space, we will use

that space for criterion space in this paper. *Standard deviation, expected return* space is definitely our choice as the units of standard deviation are much more interpretable than those of variance for conveying risk (more on this in next section).

Let \mathbf{x} be a portfolio. Then when speaking about the criterion vector of an \mathbf{x} , it will be assumed, unless something is said to the contrary, that it is given by

$$(\sqrt{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}}, \boldsymbol{\mu}^T \mathbf{x}). \quad (1)$$

Continuing under the adoption of *standard deviation, expected return* space, where S is the feasible region in decision space, the feasible region in criterion space, designated Z , is given by

$$Z = \{\mathbf{z} \in \mathbb{R}^2 \mid \mathbf{z} = (\sqrt{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}}, \boldsymbol{\mu}^T \mathbf{x}), \mathbf{x} \in S\}.$$

It is to be noted that all of the nondominated frontiers that we see in books and so forth are graphs in criterion space, not decision space which is the space of the \mathbf{x} -variables, and when the graphs have standard deviation as opposed to variance on the horizontal axis, the graphs are in *standard deviation, expected return* space. Then what is actually shown is the “northwest” boundary of the feasible region Z in criterion space. This is exactly what is shown in Figure 1. Often we do not see much of the rest of Z graphed under the rationale that none of the points not shown can be optimal for anyone.

3 Some preliminaries

Because of the role of nondominated criterion vectors in this paper, we have a few more words to say about them in this section. Then, as opposed to nondominance, we will define the term efficiency as used in this paper. Finally, we comment about covariance matrices and how a covariance matrix is instrumental in the calculation of a portfolio’s standard deviation measure.

Consider a set of criterion vectors. Then a point in the set is *nondominated* if and only if there exists no other point in the set that is better in one component without being worse in another. To illustrate, consider the criterion vectors of length 2 whose components come from the first two columns of Table 1. Note that in the risk column, lower scores are better. Then, among the seven criterion vectors of length 2, vectors 1, 4, 6 and 7 are nondominated, and vectors 2, 3 and 5 are dominated. Vectors 2, 3 and 5 are dominated by 1, 1 and 4, and 4, respectively.

Using the numbers in the liquidity column, let a third component be appended to the criterion vectors so that they are of length 3, exactly as shown in Table 1. Then, among the seven criterion vectors of length 3, vectors 1, 2, 4, 5 and 6 are nondominated, and vectors 3 and 7 dominated. Whereas 3 was formerly dominated by 1 and 4, it is now only dominated by 4. But vector 7, which was formerly nondominated, is now dominated by 6. What is going on here?

What we are observing are two properties of dominated and nondominated criterion vectors. In the property we will call the first, let all criterion vectors in a collection be unique, like criterion vectors 1 to 6 of length 2 in Table 1. Then, when

Table 1 Nondominated status of criterion vectors when appending components

portfolio	risk	return	liquidity
1	1	2	1
2	2	1	2
3	3	2	4
4	3	4	5
5	4	3	7
6	5	5	2
7	5	5	1

Table 2 Return data and resulting covariance matrix

period	stock 1	stock 2	stock 3
1	0.026	0.019	0.058
2	-0.027	0.037	0.030
\vdots	\vdots	\vdots	\vdots
60	0.001	-0.006	0.005
Σ	0.0048	0.0008	0.0023
	0.0008	0.0034	-0.0003
	0.0023	-0.0003	0.0039

appending components, no nondominated criterion can lose its nondominated status, but a dominated criterion vector can become nondominated. Thus the nondominated set can only grow to supersets of itself. This is the general situation. The other property is more specialized in that it only pertains to nondominated criterion vectors that have replicates, like criterion vectors 6 and 7 of length 2 in Table 1. This could happen in portfolio selection, for example, should there be different \mathbf{x} -vectors that produce the same criterion vector via (2). Then, when appending a component to a given set of replicates, those of the replicates that don't receive the most superior value of the additional component will lose their nondominated status.

As for the term efficiency, we observe the difference in definition between it and nondominance that is common practice in multiple criteria optimization (Steuer, 1986, Chap. 6). The distinction is this. Let $\mathbf{x} \in S$. Then \mathbf{x} is *efficient* if and only if its criterion vector (in Z) is nondominated. Thus, efficiency is a decision space concept as points in S are either efficient or inefficient, and nondominance is a criterion space concept as points in Z are either nondominated or dominated. In other words, if the criterion vector of an $\mathbf{x} \in S$ is on the nondominated frontier, then \mathbf{x} is an efficient portfolio, and vice versa. We need this more crisp distinction between the two terms to help communications as we go from two to three criteria in portfolio selection.

Moving on to covariance matrices, let the first part of Table 2 be period return data on three stocks. For instance, in period 2, the value of an investment in stock 1 declined by 2.7%. From the return data, assume that covariance matrix Σ results. In portfolio selection, covariance matrices can often be highly dense as here. Covariance matrices are important because with them the variance or standard deviation of any portfolio can be calculated by means of $\mathbf{x}^T \Sigma \mathbf{x}$ or $\sqrt{\mathbf{x}^T \Sigma \mathbf{x}}$, respectively. For example, in relation to the stocks of Table 2, let \mathbf{x} be a portfolio invested 50% in stock 1, 30% in stock 2, and 20% in stock 3. As computed in (2), \mathbf{x} 's variance is 0.002326%/month². Beyond smaller values of variance being better than larger values, the actual numerical value of variance is uninterpretable (look at the units). But after calculating \mathbf{x} 's standard deviation in (3), we have 4.82%/month which is highly interpretable given that expected return is also measured in percent per month.

$$\begin{aligned}
 \text{Var} &= \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} & (2) \\
 &= \begin{bmatrix} 0.5 & 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} 0.0048 & 0.0008 & 0.0023 \\ 0.0008 & 0.0034 & -0.0003 \\ 0.0023 & -0.0003 & 0.0039 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix} \\
 &= 0.002326\%/ \text{month}^2
 \end{aligned}$$

$$\begin{aligned}
 \text{Std} &= \sqrt{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}} & (3) \\
 &= 4.82\% / \text{month}
 \end{aligned}$$

4 Methods for constructing nondominated frontiers discretely

We now consider the computation of the nondominated frontier. Even though the nondominated frontier of (M) is a continuous curve, in practice, the world computes its nondominated frontiers discretely so as to produce dotted representations as in Figures 2 and 3.

The overall favorite for the discrete construction of nondominated frontiers is the e -constraint method. A traditional method, it has been around for as long as anyone can remember. Often utilized in multiple criteria optimization, the method has been known by the name of “ e -constraint” since Haimes (1971). Before that it apparently didn’t have a name.

In the e -constraint method, all objectives but one are converted to constraints with e right-hand sides. It is from this notation that the method derives its name. Then by employing a thoughtful series of different e values and solving, the idea is that a useful series of different solutions will result. In the case of (M), we have the simplest e -constraint case in that only one objective is to be converted to a constraint. With expected return the most practical choice, we have, for generating the portfolios that minimize risk for q different levels of expected return, the “ e -constraint” formulation of (E) where

$$\begin{aligned}
 &\min \{ \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} \} & (E) \\
 &s.t. \quad \boldsymbol{\mu}^T \mathbf{x} = e \quad \text{for each } e \in \{e_{\min}, e_2, \dots, e_{q-1}, e_{\max}\} \\
 &\quad \mathbf{x} \in S
 \end{aligned}$$

The resulting \mathbf{x} -portfolios are efficient because their criterion vectors $(\sqrt{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}}, \boldsymbol{\mu}^T \mathbf{x})$ are points on the nondominated frontier. Before deploying, however, some homework is to be carried out. This involves ascertaining the range $[e_{\min}, e_{\max}]$ of expected return values over the nondominated set, where e_{\min} is the minimum expected return value over the nondominated set and e_{\max} is the maximum expected return value over the nondominated set. Then from $[e_{\min}, e_{\max}]$, a string of values (typically equally spaced over the range) is selected so as to construct a desired dotted representation of the nondominated frontier. Suppose we have q equally spaced expected return values for e obtained in this way. Then the procedure is as follows. Solve (E) q times, once for each of the q pre-chosen values for e , note the resulting \mathbf{x} -vector, and then compute its criterion vector.

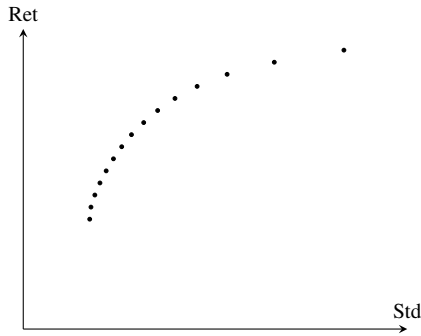


Fig. 2 Dotted representation of nondominated frontier with 15 dots.

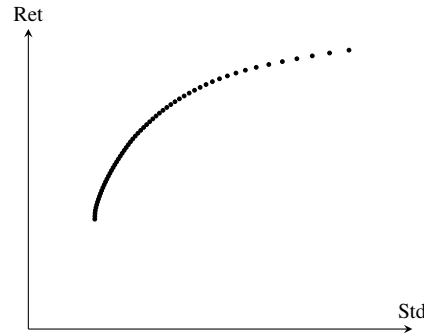


Fig. 3 Dotted representation of nondominated frontier with 60 dots.

Of course, because of the repetitiveness of the approach, the task can be time consuming. Requiring a quadratic programming (QP) run for each different value of e , depending upon q , n and the density of Σ , CPU-times can add up. If q were 15, then a dotted representation of the nondominated frontier would look as in Figure 2. If q were 60, then it would look as in Figure 3. While 15 may be enough for pedagogical purposes, many more dots is likely to be what is required in practice. Notice how the distance between dots increases as we proceed up the nondominated frontier. While this could be minimized by a more deliberate choice of e -values, it could take a lot of extra runs to get rid of this completely.

The approach is for sure intuitive. One need not be a math major to understand the concept of the nondominated frontier by its e -constraint construction—that for each level of expected return, the e -constraint method computes the portfolio that minimizes risk. A further advantage of the approach is that (E) is only a single-objective program, albeit a quadratic one. Consequently, it only requires a standard QP solver (in plentiful supply) to solve. However, because of the repetitiveness of the process, total CPU-time must always be kept in mind on larger problems.

Another method for computing a discretized representation of the nondominated frontier of (M) would be to use (L) where, under the intention of producing q points on the nondominated frontier, we have

$$\begin{aligned} \max\{-\mathbf{x}^T \Sigma \mathbf{x} + \lambda_2 \boldsymbol{\mu}^T \mathbf{x}\} \quad & \text{for each } \lambda_2 \in \{0, \lambda_{2,1}, \dots, \lambda_{2,q-1}, \lambda_{2,\max}\} \quad (\text{L}) \\ \text{s.t. } \quad & \mathbf{x} \in S \end{aligned}$$

We use the notation λ_2 because that is the weight on the second objective; 1 is the weight on the first. For a dotted representation of the nondominated frontier, instead of repetitively solving (E) for different pre-chosen values of e , in this approach we repetitively solve (L) for different pre-chosen values of λ_2 . As (L) is a weighted-sums formulation, each λ_2 value must be nonnegative, or in other words, be a member of the interval $[0, +\infty)$.

Problem (L) looks simple enough. It even has one fewer constraint than (E). However, with (L) a nettlesome issue enters the picture. It is that there is no simple way of knowing $\lambda_{2,\max}$ in advance where $\lambda_{2,\max}$ is the smallest value of λ_2 that causes (L) to

solve for the maximum expected return point on the nondominated frontier. In some cases, $\lambda_{2,\max}$ could be very large. In other cases it could be very small. It can't be foretold from the outside. The frustrations set in when solving (L) for a pre-chosen group of λ_2 's. If the largest λ_2 in the group is less than $\lambda_{2,\max}$, we will fail to learn the topmost point of the nondominated frontier. If the largest λ_2 in the group is greater than $\lambda_{2,\max}$, the last few λ_2 's in the group may well generate the same point. In the first case, we fall short of capturing all of the upper part of the frontier. In the second case, we run the risk of generating fewer points than desired.

(L) can also be disconcerting with regard to the spacing of the points generated along the nondominated frontier. At least with the ϵ -constraint method we have some control. We specify in advance the expected return component of each generated point, but with (L), we don't know either component until after the optimization is over.

5 Methods for constructing nondominated frontiers continuously

While computing nondominated frontiers discretely is about the only way seen in practice, it is possible, using specialized algorithms, to compute the full continuous curve of the nondominated frontier of (M). The specialized algorithms are all parametric quadratic programming procedures of one form or another. As parametric procedures, their strategies are to model (M) with a parameter and then vary the parameter continuously until the nondominated frontier is fully traced out. We use the term "specialized" because parametric quadratic programming procedures exist, it is just that there are no implementations in any commercial packages of which we are aware. One needs to consult the open literature or research community.

Best known among these algorithms is the critical line method of Markowitz (1956). Other contributions to parametric quadratic programming procedures for portfolio selection have been by Best (1996), Korhonen and Yu (1998), Stein et al (2008), Niedermayer and Niedermayer (2010), and Hirschberger et al (2010). The last three underscore the speed at which research implementations of parametric quadratic programming can now compute the nondominated frontier. Perhaps the most tested (see Steuer et al, 2011), we use the CIOS procedure from Hirschberger et al (2010) to help us detail the nature of the true nondominated frontiers of (M).

For solving for the nondominated frontier of (M), the formulation of interest now is (P)

$$\begin{aligned} \max \{-\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} + \lambda_2 \boldsymbol{\mu}^T \mathbf{x}\} \quad & \text{for all } \lambda_2 \geq 0 \\ \text{s.t. } \quad & \mathbf{x} \in S \end{aligned} \quad (\text{P})$$

This is the same formulation as (L), but instead of solving repetitively using different pre-chosen values of λ_2 , we vary λ_2 continuously over the interval $[0, +\infty)$ to obtain the nondominated frontier in one run. Actually, in the specialized algorithms, it is generally more convenient to start at $+\infty$ and continuously vary λ_2 down to 0, rather than the other way around.

Regardless, what we find when varying λ_2 continuously is that the nondominated frontier, with standard deviation on the horizontal axis, is piecewise hyperbolic. By

this we mean that the nondominated frontier is composed of a connected collection of curved segments, each coming from a different hyperbola. Representative of the group of parametric quadratic programming algorithms, we now review the operation of the CIOS procedure from Hirschberger et al (2010), not only to see how this is known, but also to set the stage for the algorithm to be discussed for tri-criterion portfolio selection in Section 9.

6 Parametric quadratic programming

CIOS from Hirschberger et al (2010) computes the hyperbolic segments of the non-dominated frontier by processing the Kuhn-Tucker Conditions for (P). To see the Kuhn-Tucker Conditions for (P) formed, consider from Winston (2003) the generic formulation

$$\begin{aligned}
 & \max \{f(\mathbf{x})\} && \text{(W)} \\
 \text{s.t. } & h_i(\mathbf{x}) = d_i \quad i = 1, \dots, l \\
 & g_j(\mathbf{x}) \leq b_j \quad j = 1, \dots, m \\
 & \mathbf{x} \leq \boldsymbol{\beta} \\
 & \mathbf{x} \geq \mathbf{0}
 \end{aligned}$$

where l is the number of equality constraints and m is the number of inequality constraints. If in (W) $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is concave and all constraints are convex, then it is known that $\mathbf{x} \in \mathbb{R}^n$ solves (W) if and only if there exist vectors $\mathbf{v} \in \mathbb{R}^l$, $\mathbf{u}^y \in \mathbb{R}^m$, $\mathbf{u}^\beta \in \mathbb{R}^n$ and $\mathbf{u}^x \in \mathbb{R}^n$ such that \mathbf{x} satisfies the following Kuhn-Tucker Conditions

$$\begin{aligned}
 \frac{\partial f(\mathbf{x})}{\partial x_j} - \sum_{i=1}^l v_i \frac{\partial h_i(\mathbf{x})}{\partial x_j} - \sum_{i=1}^m u_i^y \frac{\partial g_i(\mathbf{x})}{\partial x_j} - u_j^\beta + u_j^x &= 0 \quad j = 1, \dots, n \\
 h_i(\mathbf{x}) &= d_i \quad i = 1, \dots, l \\
 g_j(\mathbf{x}) &\leq b_j \quad j = 1, \dots, m \\
 \mathbf{x} &\leq \boldsymbol{\beta} \\
 \mathbf{x} &\geq \mathbf{0} \\
 \mathbf{u}^y &\geq \mathbf{0} \\
 \mathbf{u}^\beta &\geq \mathbf{0} \\
 \mathbf{u}^x &\geq \mathbf{0} \\
 (b_j - g_j(\mathbf{x}))u_j^y &= 0 \quad j = 1, \dots, m \\
 (\beta_j - x_j)u_j^\beta &= 0 \quad j = 1, \dots, n \\
 x_j u_j^x &= 0 \quad j = 1, \dots, n \\
 \mathbf{v} &\text{ unrestricted}
 \end{aligned}$$

Meeting the objective function and constraint requirements, the Kuhn-Tucker Conditions (KTC) for (P), in matrix format, are

$$2\Sigma\mathbf{x} - \lambda_2\boldsymbol{\mu} + \mathbf{H}^T\mathbf{v} + \mathbf{G}^T\mathbf{u}^y + \mathbf{I}_n\mathbf{u}^\beta - \mathbf{I}_n\mathbf{u}^x = \mathbf{0} \quad (4.1)$$

$$\mathbf{H}\mathbf{x} = \mathbf{d} \quad (4.2)$$

$$\mathbf{G}\mathbf{x} \leq \mathbf{b} \quad (4.3)$$

$$\mathbf{I}_n\mathbf{x} \leq \boldsymbol{\beta} \quad (4.4)$$

$$\mathbf{x} \geq \mathbf{0}, \mathbf{u}^y \geq \mathbf{0}, \mathbf{u}^\beta \geq \mathbf{0}, \mathbf{u}^x \geq \mathbf{0}$$

\mathbf{v} unrestricted

$$(b_j - g_j(\mathbf{x}))u_j^y = 0, \quad j = 1, \dots, m \quad (\beta_j - x_j)u_j^\beta = 0, \quad j = 1, \dots, n$$

$$x_j u_j^x = 0, \quad j = 1, \dots, n$$

With just (4.1–4.4) being $(2n + l + m) \times (3n + l + 2m)$, a difficulty with the above is its size. Fortunately, it is possible to reduce the system by modeling the upper bound constraints $\mathbf{I}_n\mathbf{x} \leq \boldsymbol{\beta}$ of (4.4) implicitly. Also, it is possible, as shown in Hirschberger et al (2010), to consolidate \mathbf{u}^β and \mathbf{u}^x into a single vector $\bar{\mathbf{u}}^x \in \mathbb{R}^n$ because when a u_j^β is needed, its counterpart u_j^x is not, and vice versa. With this reducing (4.1–4.4) to $(n + l + m) \times (2n + l + 2m)$, we have the more manageable system that we call the *reduced KTC*

$$2\Sigma\mathbf{x} + \mathbf{H}^T\mathbf{v} + \mathbf{G}^T\mathbf{u}^y - \bar{\mathbf{I}}_n\bar{\mathbf{u}}^x = \mathbf{0} + \lambda_2\boldsymbol{\mu} \quad (5.1)$$

$$\mathbf{H}\mathbf{x} = \mathbf{d} \quad (5.2)$$

$$\mathbf{G}\mathbf{x} + \mathbf{I}_m\mathbf{y} = \mathbf{b} \quad (5.3)$$

$$\boldsymbol{\beta} - \mathbf{x} \geq \mathbf{0}, \mathbf{x} \geq \mathbf{0}, \mathbf{u}^y \geq \mathbf{0}, \bar{\mathbf{u}}^x \geq \mathbf{0}, \mathbf{y} \geq \mathbf{0} \quad (5.4)$$

\mathbf{v} unrestricted (5.5)

$$\mathbf{x}^T\bar{\mathbf{u}}^x = 0, \mathbf{y}^T\mathbf{u}^y = 0 \quad (5.6)$$

where $\bar{\mathbf{I}}_n$ in (5.1) is a diagonal matrix whose j -th diagonal element is -1 or 1 depending upon whether u_j^β is in the problem at the moment or not, respectively², and (5.6) displays the new complementary slackness constraints which have been reduced as well.

Since the reduced KTC is equivalent to the KTC, then by virtue of Eaves (1971), the reduced KTC is solvable if and only if there is a basis of (5.1–5.3) that solves the system (5.1–5.6). Assume that S is nonempty. We know this, not only because of the $[\mathbf{0}, \boldsymbol{\beta}]$ box constraints on \mathbf{x} , but also because $\mathbf{1}^T\mathbf{x} = 1$ is a constraint in $\mathbf{H}\mathbf{x} = \mathbf{d}$. Thus, there is a solution to (P) for all $\lambda_2 \in \mathbb{R}$. Hence, there is a basis of (5.1–5.3) that solves the system for what we need, all $\lambda_2 \in [0, +\infty)$. A task is to find a basis of (5.1–5.3) that solves the system from which to commence pivoting to all other bases of (5.1–5.3) that solve (5.1–5.6) that are necessary for the generation of the nondominated frontier.

² As a technical matter, slight adjustments are made to \mathbf{H} , \mathbf{d} , \mathbf{G} and \mathbf{b} in (5.2–5.3) depending upon which elements are -1 in $\bar{\mathbf{I}}_n$

A way to proceed on the task is to solve the linear program

$$\begin{aligned} \max \{ & \boldsymbol{\mu}^T \mathbf{x} \} && \text{(LP)} \\ \text{s.t. } & \mathbf{x} \in S \end{aligned}$$

for a solution that gives us the expected return of maximum value over the nondominated frontier. We do this because it is not too difficult to obtain from the solution of (LP) \mathbf{x} , \mathbf{v} , \mathbf{u}^y , $\bar{\mathbf{u}}^x$ and \mathbf{y} that enable us to form a basis of (5.1–5.3) with $\lambda_2 = \lambda_{2,\max}$ that solves the system to start the process of tracing out the nondominated frontier. It is only in cases of alternative optima that the \mathbf{x} -vector from (LP) might be different from the \mathbf{x} -vector of this initial basis. Recall that $\lambda_{2,\max}$ is the smallest value of λ_2 for which the topmost point on the nondominated frontier solves (P). This is a good basis from which to start because there is nothing above $\lambda_{2,\max}$. It is only necessary to range λ_2 downward from there to 0 to trace out the nondominated frontier.

As we begin, in this basis, ranging λ_2 downward from $\lambda_{2,\max}$, the variables in the basis, including any components of \mathbf{x} that might be in the basis, adjust to accommodate the changes in λ_2 . Due to the linearity of the equations that constitute (5.1–5.3), all of the variables that adjust do so linearly with λ_2 . At some point in the ranging downward of λ_2 , provided we don't reach 0 first, a point will be encountered at which the basis would lose feasibility if we were to go further. The interval over which λ_2 can range in a given basis without forcing a pivot is called a *stability set*. At this point we pivot to a new basis of (5.1–5.3) that solves the system. Because there is a solution to (P) for all values of λ_2 , a new basis will always be available.

In continuing to further decrease λ_2 , there will be another range over which the new basis holds, but then will come a value of λ_2 , provided we don't reach 0 first, at which one or more of the basic variables, which have been adjusting linearly all along, would lose feasibility if we were to go anymore with λ_2 . Recording the stability set, we pivot to achieve another basis over which to continue ranging λ_2 downward. Also, recording how \mathbf{x} linearly adjusts over each basis, we continue with λ_2 , pivoting when necessary, until λ_2 reaches 0, at which point we are done.

With reference to Figures 4 to 7, we illustrate the procedure. Consider the 22 dots in Figure 4. They are from the nondominated frontier of Figure 1. But they are not just dots. They have meaning. The first dot is where, with $\lambda_2 = \lambda_{2,\max}$, we start. Apart from the bottommost dot, where $\lambda_2 = 0$, all of the others are where we pivot. For the part of the nondominated frontier between any two adjacent dots, it is taken from a hyperbola. With the \mathbf{x} -vector tracking a straight line as λ_2 is ranged from one end of a basis to the other (i.e., over the stability set of the basis), the criterion vectors of the \mathbf{x} 's along the linear line segment trace a hyperbolic arc, because of the $\sqrt{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}}$ term.

In Figure 5 we see the hyperbola that contributes the first hyperbolic arc of the nondominated frontier coming down the curve. In Figure 6 we see the hyperbolas that contribute the first two hyperbolic arcs of the nondominated frontier coming down the curve. In Figure 7, we see the hyperbolas that contribute the first seven hyperbolic arcs, and so forth. Note that when we pivot coming down the curve, we switch from the end of one hyperbolic arc to the beginning of the next.

Perhaps some readers are familiar with the expression that the nondominated frontier is “piecewise linear in decision space and piecewise hyperbolic in criterion

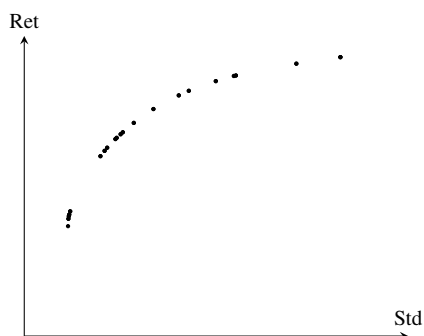


Fig. 4 22 dots of meaning on the nondominated frontier of Figure 1.

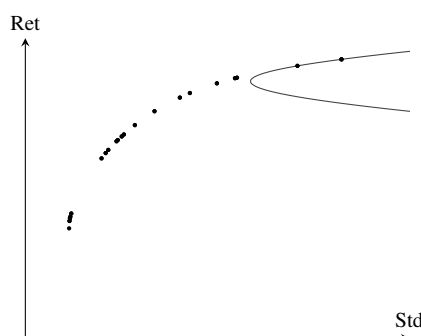


Fig. 5 The hyperbola from which the first hyperbolic arc of the nondominated frontier is taken.

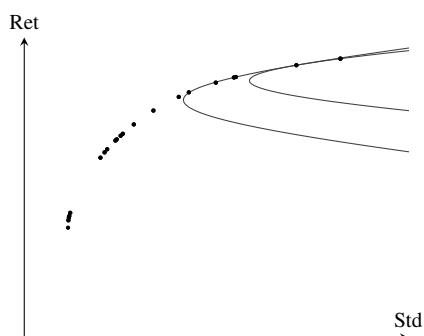


Fig. 6 The hyperbolas from which the first two hyperbolic arcs of the nondominated frontier are taken.

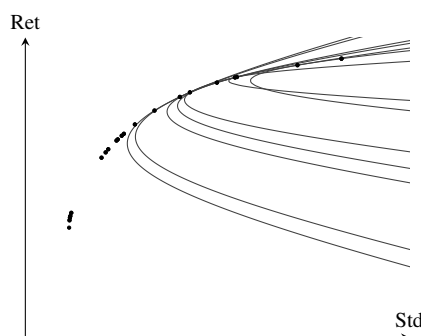


Fig. 7 The hyperbolas from which the first seven hyperbolic arcs of the nondominated frontier are taken.

space.” Now we can see why. For the first part of the expression, consider the bases of (5.1–5.3) encountered coming down the nondominated frontier. As we know, within each, \mathbf{x} tracks a linear line segment as λ_2 is varied over the stability set of the basis. With each pair of successive linear line segments connected at their point of pivot, all are connected, and hence we can see how the linear line segments, which are the efficient portfolios of (M), form a piecewise linear path in S (i.e., decision space). And as each linear line segment gives rise to a hyperbolic arc, the nondominated frontier is piecewise hyperbolic in criterion space.

Occasionally, there will occur a nondominated frontier with a *kink* as in Figure 8. In the basis of (5.1–5.3) that solves the system at the kink (it suffices for there to be one basis for purposes of explanation), what happens is that none of the components of \mathbf{x} are among those that undergo linear adjustments as λ_2 is ranged over the stability set of the basis.

Now we can see what parametric quadratic programming does. It decomposes $\lambda_2 \geq 0$ (i.e., it decomposes the non-negative portion of the real line) into intervals, with $[\lambda_{2,max}, +\infty)$ corresponding to the topmost point of the nondominated frontier. In this way, the topmost point on the nondominated frontier is a kink in that there is

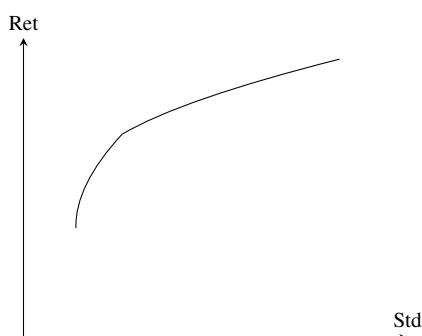


Fig. 8 A nondominated frontier with a kink other than at the top of the frontier.

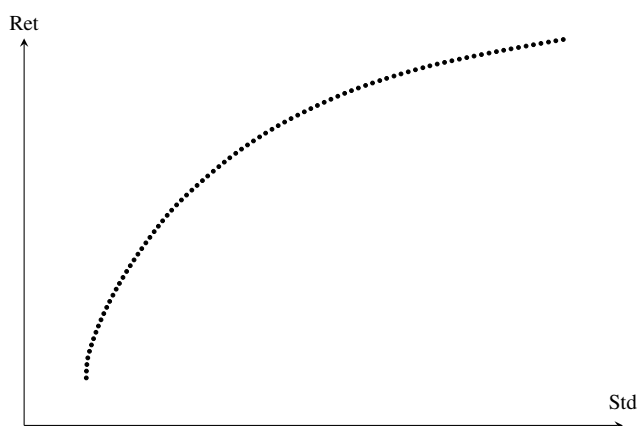


Fig. 9 A nondominated frontier represented by 100 equally spaced points indicating that the larger the graph the more points that are required for a quality representation.

no movement in \mathbf{x} for any movement of λ_2 in $[\lambda_{2,\max}, +\infty)$. While the terminology may seem a little peculiar at this point in the paper, the nondominated frontier can be viewed as consisting of 0-dimensional kinks (rare other than at the point at the top of the frontier) and 1-dimensional hyperbolic arcs. The dimensionalities of these aspects of the nondominated frontier are taken from the dimensionalities of the sets of \mathbf{x} -vectors corresponding to the different stability sets.

Because nondominated frontiers are so often constructed discretely and then presented in the form of dotted representations, which are not by the way aesthetically displeasing, there may be users who do not wish to see nondominated frontiers presented to them in the form of a continuous curve. They just wish to continue seeing nondominated frontiers in the way to which they have become accustomed.

Whereas 60 dots in Figure 3 might seem plenty (apart from the tendency to be spread out at the top), this is because the figure is so small. In practice, nondominated frontiers would be shown in a larger form, such as on a regularly-sized piece of paper or on a computer screen. Then in all likelihood 60 would not be enough. We can get a sense of this from Figure 9, which is still not a big graph.

If one wants more than a few dots to represent a nondominated frontier, actually, the fastest way to do this is to employ a parametric quadratic programming code to compute the whole curve of the nondominated frontier. This is because codes such as Todd's implementation of Markowitz's critical line algorithm, called Optimizer and found in the appendix of Markowitz and Todd (2000), or CIOS from Hirschberger et al (2010) can solve for a full mathematical specification of a nondominated frontier in not much more time than it takes a typical QP solver to solve for a single ϵ -constraint point on the nondominated frontier. Then, since we would know from output of Optimizer or CIOS the equation of each hyperbolic segment, dots could be placed on the nondominated frontier in any pre-determined fashion, including equally-spaced. As pointed out in Qi et al (2009), it would not take more than a second or two for a user-written routine running under something like Matlab to do this. Figure 9 is a dotted representation of a nondominated frontier produced in this way. With not only 100 points, but equally spaced as well, this is a feat not so easily accomplished by a discrete method.

7 Additional criteria in portfolio selection

Over the many years of Markowitz portfolio theory, perhaps the most persistent criticism that it has had to endure has been about its lack of a formal mechanism for accommodating additional criteria (beyond risk and return). As surveyed in Steuer and Na (2003), many papers have been written on the topic and numerous approaches have been suggested. To mention a few, there are Stone (1973), Colson and DeBryn (1989), Chow (1995), Ogryczak (2000), Bana e Costa and Soares (2004), Ben Abdelaziz et al (2007), and so forth. However, despite such efforts, little long-term traction has been achieved. We think we have at least a partial explanation.

As set out by Hwang and Masud (1979), procedures for solving a multiple criteria decision problem, which is what portfolio selection clearly is with additional criteria, can be grouped into three categories depending upon when preference information is elicited from the decision maker. The categories are *a priori*, *progressive articulation*, and *a posteriori*. In an *a priori* method, information about the decision maker's preferences is sought before an optimization problem based upon it is solved for the decision maker's supposedly optimal solution. Difficulties here include that it may be hard to obtain the information that is required in advance and that the decision maker is not given any kind of picture of the nondominated set other than for the point generated.

In the progressive articulation category are interaction procedures. Phases of computation are interleaved with phases of decision making. In these procedures, preference information is elicited iteratively to guide the solution process. This is better, but at best, the decision maker only gets a partial look at the nondominated set.

In an *a posteriori* method, the nondominated set in its entirety is computed first. Then it is shown to the decision maker for the selection of his or her most preferred solution. *A posteriori* methods, of course, are not applicable to problems for which the nondominated set can not be computed, and even in problems where sufficient representations of the nondominated set are possible, one must be mindful about in-

formation overload when conveying the volume of information to the decision maker. But if these challenges can be dealt with, it is hard to see why this wouldn't be the way to approach a multiple criteria problem.

In addition to its modeling, a further reason for the appeal of Markowitz portfolio selection is that, in its decision making second stage, it is an a posteriori approach. After computing the nondominated frontier, a graph of the nondominated frontier is shown to the decision maker so all can be seen at once. After being brought into contact with the a posteriori totality of a problem in this way, the decision maker is in a very natural position to compare and contrast solution candidates, and ultimately decide upon a final one.

8 Tri-criterion formulation

The difficulty with previously written papers on multiple criteria portfolio selection is that they have almost all revolved around a priori procedures and consequently don't provide much information about what other types of solutions are candidates. Also, in a priori procedures, the computer produces the solution. The decision maker doesn't get to do much picking. In the few papers that have attempted a posteriori approaches, the main difficulties have been with the approximations of the nondominated set they have been stuck having to work with.

With all of this background, we now proceed with the main part of the paper which is to extend Markowitz's portfolio selection to tri-criterion portfolio selection without changing any of the principles and practices that have made Markowitz's bi-criterion portfolio selection so famous. To accommodate a linear criterion in addition to expected return, instead of (M), the formulation now is (T)

$$\begin{aligned}
 & \min \{ \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} \} & (T) \\
 & \max \{ \boldsymbol{\mu}^T \mathbf{x} \} \\
 & \max \{ \mathbf{v}^T \mathbf{x} \} \\
 & s.t. \quad \mathbf{x} \in S
 \end{aligned}$$

where the third criterion could be liquidity as in Lo et al (2003), dividend yield as in Ehrgott et al (2004), social responsibility as in Ballesteros et al (2012), and so forth. One of the first things realized about (T) is that the nondominated frontier becomes the nondominated surface. Thus, the challenges, to stay within the successes of the Markowitz's framework, are to

- (a) obtain a highly comprehensive if not exact rendition of the nondominated surface
- (b) be able to display the nondominated surface graphically so that the decision maker will not have undue difficulties in working with it.

Because of the popularity of discrete methods in regular portfolio selection, one might think of them first, but they will not work here. In a problem where 100 points might be necessary to portray the nondominated frontier, with a third criterion added, thousands would be needed to portray the problem's nondominated surface to achieve the

same type of dotted representation density, and even then, it would be very hard to get the dots to graph attractively. With discrete methods effectively ruled out, and none of the parametric quadratic programming approaches alluded to earlier that are based on Markowitz's (1956) critical line algorithm able to address the situation because the critical line method has no provisions for additional criteria, we have a problem. But as it turns out, in a break with the past, it is now possible to compute the non-dominated surface of (T) exactly. Being able to compute the nondominated surface of (T) is possible by means of the multi-parametric extension of the CIOS procedure of Hirschberger et al (2010), that we will refer to as CIOSm, that is developed and described for this purpose in Hirschberger et al (2013).

As a multi-parametric quadratic programming procedure, CIOSm operates in a fashion that parallels that of the CIOS parametric quadratic programming procedure outlined in Section 6, albeit with additional details. Kuhn-Tucker Conditions are formed, bases are identified, pivoting takes place, and stability sets are recorded until the entire nondominated surface is computed. How CIOSm functions in this fashion is described in Section 9.

In this way, in tri-criterion portfolio selection, we are able to meet the full requirements of (a) above. As for (b), in this paper, PGFPlots by Feuersänger (2010) is employed. As of this writing, PGFPlots is our graphics choice because for us we are in control of the software, rather than the other way around. That is, with PGFPlots we can pretty much create any kind of graph that we can imagine, rather than having restrictions imposed upon us as to what we can and can't do. Thus, with PGFPlots we feel that we can make with precision the kinds of graphs that are necessary to give to the decision-making stage of tri-criterion portfolio selection the same type of strength that the decision-making stage of bi-criterion portfolio selection has always enjoyed.

9 Multi-parametric quadratic programming

Whereas (P) was the starting formulation for computing exactly the nondominated frontier of (M), (MP) is now the starting formulation for computing exactly the non-dominated surface of (T)

$$\begin{aligned} \max \{ & -\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} + \lambda_2 \boldsymbol{\mu}^T \mathbf{x} + \lambda_3 \mathbf{v}^T \mathbf{x} \} & \text{for all } \lambda_2, \lambda_3 \geq 0 & \text{(MP)} \\ \text{s.t. } & \mathbf{x} \in S \end{aligned}$$

The third term in (MP) is because of the additional linear criterion and λ_3 is the weight on that criterion. However, here we have to solve for all $\lambda_2, \lambda_3 \geq 0$, whereas before it was only to solve for all $\lambda_2 \geq 0$. Thus, instead of partitioning the nonnegative portion of the real line (or in other words, partitioning the nonnegative orthant of \mathbb{R}^1) into interval stability sets, here the task is to partition the nonnegative orthant of \mathbb{R}^2 into polyhedron stability sets. We can see why by looking at the below reduced KTC for (T)

$$2\Sigma\mathbf{x} + \mathbf{H}^T\mathbf{v} + \mathbf{G}^T\mathbf{u}^y - \bar{\mathbf{I}}_n\bar{\mathbf{u}}^x = \mathbf{0} + \lambda_2\boldsymbol{\mu} + \lambda_3\boldsymbol{\nu} \quad (6.1)$$

$$\mathbf{H}\mathbf{x} = \mathbf{d} \quad (6.2)$$

$$\mathbf{G}\mathbf{x} + \mathbf{I}_m\mathbf{y} = \mathbf{b} \quad (6.3)$$

$$\boldsymbol{\beta} - \mathbf{x} \geq \mathbf{0}, \mathbf{x} \geq \mathbf{0}, \mathbf{u}^y \geq \mathbf{0}, \bar{\mathbf{u}}^x \geq \mathbf{0}, \mathbf{y} \geq \mathbf{0} \quad (6.4)$$

$$\mathbf{v} \text{ unrestricted} \quad (6.5)$$

$$\mathbf{x}^T\bar{\mathbf{u}}^x = 0, \mathbf{y}^T\mathbf{u}^y = 0 \quad (6.6)$$

Note that instead of just one parametric term on the right in (5.1), in (6.1) we have two. So it is no longer seeing how far λ_2 can be range over $[0, +\infty)$ without forcing a pivot. It is over what region of the nonnegative orthant of \mathbb{R}^2 can the tuple (λ_2, λ_3) range without forcing a pivot. This then is the stability set of the selected basis of (6.1–6.3) undergoing study. To illustrate, consider the $n = 4$ security tri-criterion problem in which \mathbf{G} is vacuous, \mathbf{H} contains only the full-investment constraint $\mathbf{1}^T\mathbf{x} = 1$, $\boldsymbol{\beta} = \mathbf{1}$, and

$$\Sigma = \begin{bmatrix} 6.09\text{E-}3 & 1.20\text{E-}3 & 9.68\text{E-}3 & 2.13\text{E-}3 \\ & 7.02\text{E-}3 & 6.63\text{E-}3 & -1.05\text{E-}3 \\ & & 2.98\text{E-}3 & 4.43\text{E-}3 \\ & & & 5.94\text{E-}3 \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} 1.23\text{E-}3 \\ 7.85\text{E-}3 \\ 5.61\text{E-}3 \\ 5.91\text{E-}3 \end{bmatrix} \quad \boldsymbol{\nu} = \begin{bmatrix} -9.17\text{E-}3 \\ 9.33\text{E-}3 \\ 1.96\text{E-}3 \\ 8.25\text{E-}3 \end{bmatrix}.$$

After solving (LP) and obtaining from the solution a basis of (6.1–6.3) that solves (6.1–6.6), CIOSm identifies the stability set of this basis to be the dotted-white area in Figure 10left. The criterion space image of this stability set is the 0-dimensional point given by the dot in Figure 10right. Not counting the λ_2 axis, the stability set has two bounding constraints. To cross either requires a pivot. Since it can't do both at the same time, it pivots across its leftmost boundary, and puts the other possible pivot in a List to be gotten to later.

The stability set of the new basis is the light-gray region in Figure 10left. The criterion space image of this stability set is the 1-dimensional hyperbolic arc shown emanating from the 0-dimensional point in Figure 10right. Not counting the λ_2 axis, the light-gray stability set has three bounding constraints. Since there is no sense in going back to the dotted-white area, the leftmost boundary is chosen for pivoting with the other boundary put in the List. Now there are two in the List. The stability set of the new basis is the gray region in Figure 10left. It is smaller, but size does not always make that much difference. The criterion space image of the gray stability set is the 2-dimensional hyperboloidic *platelet* shown in Figure 10right on whose relative boundary is the 1-dimensional hyperbolic arc. Note the color scheme that we are following for stability sets: dotted-white for 0-dimensional points, light-gray for 1-dimensional hyperbolic arcs, and gray for 2-dimensional hyperboloidic platelets.

In Figure 11, the most recent addition to the List is chosen for pivoting. The stability set of the new basis obtained in this way is the new gray polyhedron in Figure 11left. The criterion space image of this gray stability set is the new 2-dimensional hyperboloidic platelet shown in Figure 11right.

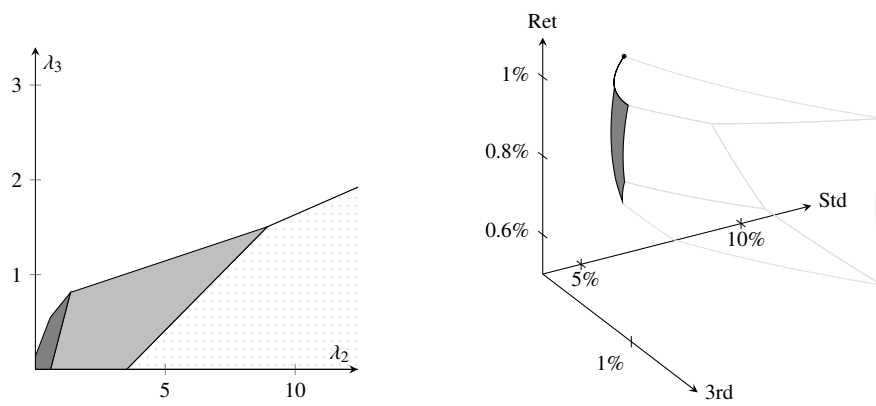


Fig. 10 First three stability sets in parameter space on the left, image of the first three stability sets in criterion space on the right.

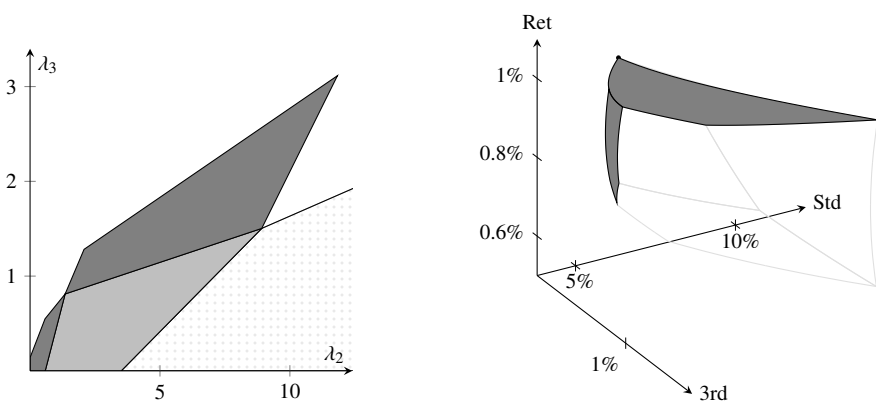


Fig. 11 First four stability sets in parameter space on the left, image of the first four stability sets in criterion space on the right.

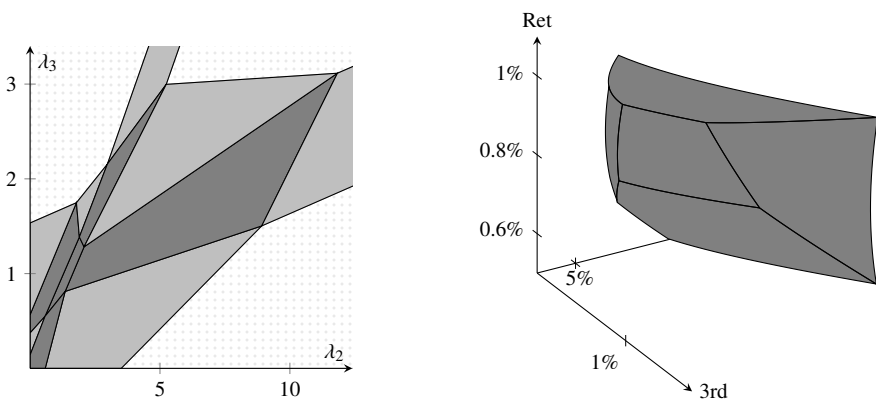


Fig. 12 All stability sets in parameter space on the left, entire nondominated surface in criterion space on the right.

Figure 12*left* shows the full decomposition of the nonnegative orthant of λ_2, λ_3 parameter space. When this is achieved we are done. As seen, this problem has 15 stability sets: three dotted-white, seven light-gray, and five gray. Thus, the surface has five 2-dimensional hyperboloidic platelets, or what are simply called “platelets.” When gray stability sets have a boundary in common, that means that along the corresponding common boundary in criterion space the two platelets blend into each other in a continuously differentiable fashion. Otherwise, if there is a light-gray stability set separating two gray stability set boundaries, the two platelets connect with each other along the corresponding boundary in the form of a ridge.

As seen from examining the stability sets in Figure 12*left*, there is some blending together of the platelets in Figure 12*right*, but in several instances they connect with one another abruptly. From the light-gray areas in Figure 12*left*, we see that this problem has a large proportion of ridges, but that is because it is a small problem. As problem size grows, the proportion of 1-dimensional ridges relative to 2-dimensional platelets drops rapidly.

10 Additional tri-criterion illustrations

Having illustrated the CIOSm procedure on a 4-security example, we now show the nondominated surfaces of three randomly generated problems that have more securities to get a better idea of how things would work on more realistically sized problems. For the problems, the elements of μ and ν were randomly sampled from a normal distribution with a mean of 10% and a standard deviation of 6%. For the covariance matrices of the problems, we employed the procedure for simulating portfolio-selection covariance matrices described in Hirschberger et al (2007). In accordance with the procedure, the diagonal entries of the covariance matrices were drawn from a distribution with an expected value of 0.012 and a standard deviation of 0.012, and the off-diagonal entries of the covariance matrices were drawn from a distribution with an expected value of 0.0025 and a standard deviation of 0.0025.

Figure 13 shows the nondominated surface of a problem created in this way with $n = 10$ securities. In this problem, all upper bounds on the securities are 0.20. The *left* and *right* figures show the problem’s nondominated surface from two different angles. On the left, the nondominated surface is shown from a 60/30 azimuth/elevation viewpoint, and on the right, the nondominated surface is shown from a 105/20 azimuth/elevation viewpoint. It seems instructive to also plot a “shadow,” i.e., the projection of the nondominated surface onto the *standard deviation, 3rd objective* plane, to clarify the suspension of the nondominated surface in the three-dimensional space. On this problem, CIOSm reveals that the nondominated surface consists of 61 two-dimensional platelets. However, the platelets vary in size with about half of the platelets in the example too small to be recognizable on the graph. It is not uncommon for a nondominated surface to have many small platelets like this.

In Figures 14 and 15, we continue with two larger-size examples. In Figure 14 we see, from two angles, the nondominated surface of a 50-security tri-criterion problem generated as above, but with an upper bound of 0.10 on each security. On this problem, CIOSm reveals that the nondominated surface consists of 548 two-dimensional

platelets. In this problem, too, there are many platelets too small to be recognizable on the graph. In larger-sized problems, we have noticed that many of the small platelets tend to cluster around the minimum standard deviation point. In Figure 15 we see, from two angles, the nondominated surface of a 75-security tri-criterion problem generated as above. The upper bound here on each security is again 0.10. On this problem, CIOsm reveals that the nondominated surface consists of 1,120 two-dimensional platelets.

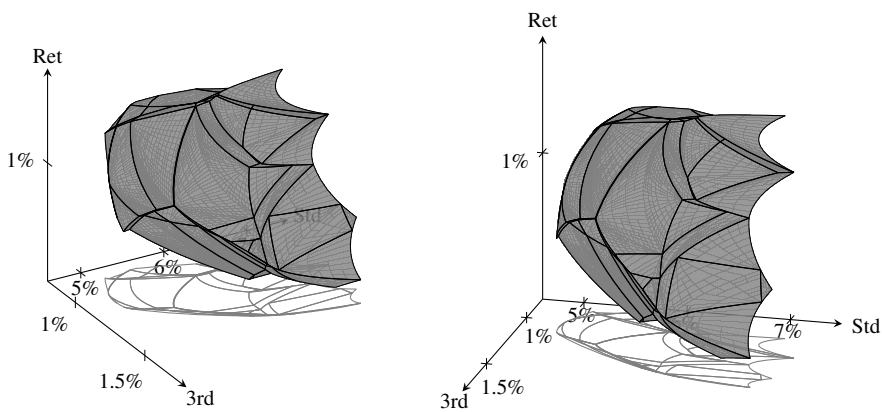


Fig. 13 Nondominated surface of a 10-security example having 61 two-dimensional platelets shown twice.

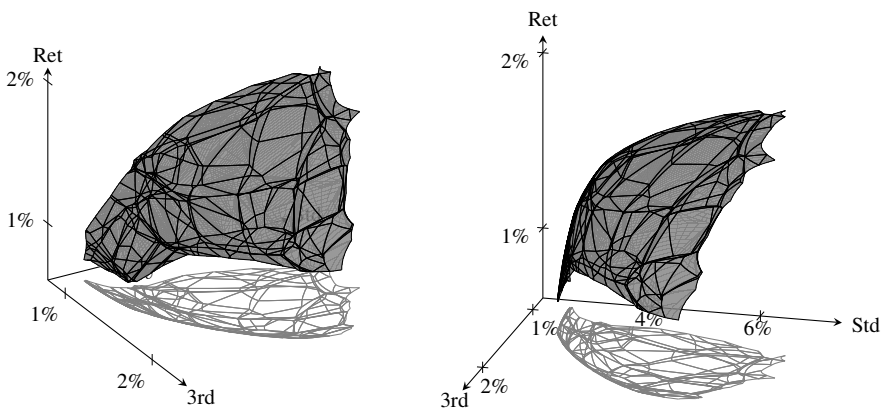


Fig. 14 Nondominated surface of a 50-security example having 548 two-dimensional platelets shown twice.

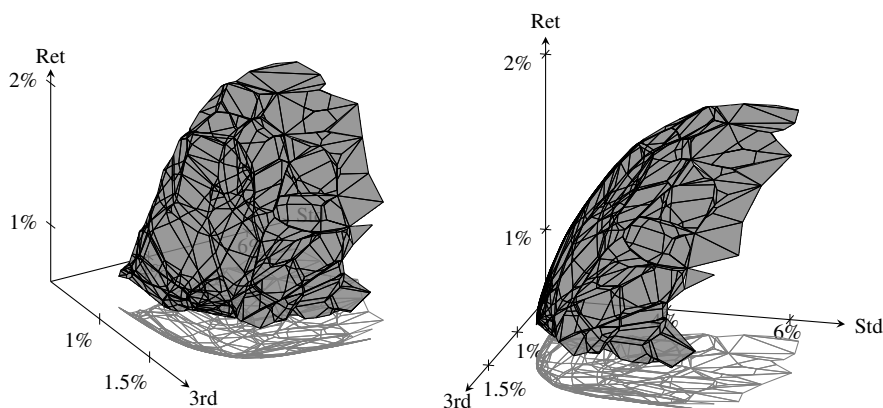


Fig. 15 Nondominated surface of a 75-security example having 1,120 two-dimensional platelets shown twice.

11 Conclusions

We started this paper by reviewing methods to compute the standard Markowitz bi-criterion nondominated frontier. The most common approach is the e -constraint method, which discretizes values along the vertical axis and then minimizes variance for each discretized level of expected return. Another discrete method is the weighted-sums approach. In this approach, we use a scalarized version of (M) in the form of (L). The scalarized version also provides a vantage point for analytically computing the entire nondominated frontier. It turns out that the nondominated frontier is composed of a sequence of adjacent segments, each stemming from a different hyperbola, and there are algorithms available to compute the segments analytically.

Following this, we discussed an extension to the problem to include a third, linear criterion. Such a criterion could be liquidity, momentum, social responsibility, and so forth. In this tri-criterion setting, the nondominated frontier in two-dimensional criterion space becomes a nondominated surface in three-dimensional criterion space. While the nondominated frontier consists of segments taken from two-dimensional hyperbolas, the nondominated surface consists of patches taken from three-dimensional hyperboloids.

Though it might be desirable to include even further linear and possibly also quadratic criteria in portfolio selection in the future, the advantage of the tri-criterion approach is that the dimensionality of criterion space is still such that it is possible to graph the nondominated set. In this way, the entire surface can be shown to the decision maker so that he or she can select from it, while seeing all, his or her most preferred solution. Thus, the approach falls into the *a posteriori* category of Hwang and Masud, which presents a transparent and intuitive way for the decision maker. For portfolio problems with more than three criteria, one would presumably have to fall back upon interactive procedures from the *progressive articulation* category, which might be less accepted, but in such situations there might be no other choice.

As for CPU-times, all of the nondominated surfaces of this paper were computed in less than 10 seconds cumulative on a Pentium 3.80GHz computer, not a particularly fast machine. The time taken by PGFPlots to graph the surfaces, however, took several times longer, but at this point in our research, we have made no attempt to optimize the graphing process.

References

- Abdelaziz FB, Aouni B, El-Fayedh R (2007) Multi-objective stochastic programming for portfolio selection. *European Journal of Operational Research* 177(3):1811–1823
- Ballester E, Bravo M, Perez-Gladish B, Arenas-Parra M, Plà-Santamaría D (2012) Socially responsible investment: A multicriteria approach to portfolio selection combining ethical and financial objectives. *European Journal of Operational Research* 216(2):487–494
- Best MJ (1996) An algorithm for the solution of the parametric quadratic programming problem. In: Fischer H, Riedmüller B, Schäffler S (eds) *Applied Mathematics and Parallel Computing: Festschrift for Klaus Ritter*, Physica-Verlag, pp 57–76
- Caballero R, Cerdá E, Muñoz MM, Rey L, Stancu-Minasian IM (2001) Efficient solution concepts and their relations in stochastic multiobjective programming. *Journal of Optimization Theory and Applications* 110(1):53–74
- Chow G (1995) Portfolio selection based on return, risk, and relative performance. *Financial Analysts Journal* 52(2):54–60
- Colson G, DeBruyn C (1989) An integrated multiobjective portfolio management system. *Mathematical and Computer Modelling* 12(10-11):1359–1381
- Bana e Costa CA, Soares JO (2004) A multicriteria model for portfolio management. *European Journal of Finance* 10(3):198–211
- Eaves BC (1971) On quadratic programming. *Management Science* 17(11):698–711
- Ehrgott M (2005) *Multicriteria Optimization*, 2nd edn. Springer
- Ehrgott M, Klamroth K, Schwehm C (2004) An MCDM approach to portfolio optimization. *European Journal of Operational Research* 155(3):752–770
- Feuersänger C (2010) Manual for package PGFPLOTS, version 1.4.1, Universität Bonn.
- Haimes YY, Lasdon LS, Wismer DA (1971) On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man and Cybernetics* SMC-1(3):296–297
- He XD, Zhou XY (2011) Portfolio choice under cumulative prospect theory: An analytical treatment. *Management Science* 57(2):315–331
- Hirschberger M, Qi Y, Steuer RE (2007) Randomly generating portfolio-selection covariance matrices with specified distributional characteristics. *European Journal of Operational Research* 177(3):1610–1625
- Hirschberger M, Qi Y, Steuer RE (2010) Large-scale MV efficient frontier computation via a procedure of parametric quadratic programming. *European Journal of Operational Research* 204(3):581–588

- Hirschberger M, Qi Y, Steuer RE, Utz S, Wimmer M (2013) Computing the nondominated surface in tri-criterion portfolio selection. *Operations Research* (in press)
- Hwang CL, Masud ASM (1979) *Multiple Objective Decision Making – Methods and Applications: A State-of-the-Art Survey*. Lecture Notes in Economics and Mathematical Systems, Vol. 164, Springer Verlag, Berlin
- Korhonen P, Yu GY (1998) On computing objective function values in multiple objective quadratic-linear programming. *European Journal of Operational Research* 106(1):184–190
- Lo AW, Petrov C, Wierzbicki M (2003) It's 11pm – Do you know where your liquidity is? The mean-variance-liquidity frontier. *Journal of Investment Management* 1(1):55–93
- Markowitz HM (1952) Portfolio selection. *Journal of Finance* 7(1):77–91
- Markowitz HM (1956) The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly* 3(1-2):111–133
- Markowitz HM (1959) *Portfolio Selection: Efficient Diversification of Investments*. John Wiley and Sons
- Markowitz HM (1987) *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Basil Blackwell, Oxford
- Markowitz HM, Todd P (2000) *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Frank J. Fabozzi Associates, New Hope, Pennsylvania
- Miettinen KM (1999) *Nonlinear Multiobjective Optimization*. Kluwer, Boston
- Niedermayer A, Niedermayer D (2010) Applying Markowitz's critical line algorithm. In: Guerard JB (ed) *Handbook of Portfolio Construction*, Springer-Verlag, Berlin, pp 383–400
- Ogryczak W (2000) Multiple criteria linear programming model for portfolio selection. *Annals of Operations Research* 97:143–162
- Qi Y, Hirschberger M, Steuer RE (2009) Dotted representations of mean-variance efficient frontiers and their computation. *INFOR* 47(3):15–21
- Stein M, Branke J, Schmeck H (2008) Efficient implementation of an active set algorithm for large-scale portfolio selection. *Computers & Operations Research* 35(12):3945–3961
- Steuer RE (1986) *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley, New York
- Steuer RE, Na P (2003) Multiple criteria decision making combined with finance: A categorized bibliography. *European Journal of Operational Research* 150(3):496–515
- Steuer RE, Qi Y, Hirschberger M (2011) Comparative issues in large-scale mean-variance efficient frontier computation. *Decision Support Systems* 51(2):250–255
- Stone BK (1973) A linear programming formulation of the general portfolio selection problem. *Journal of Financial and Quantitative Analysis* 8(4):621–636
- Winston WL (2004) *Operations Research: Applications and Algorithms*, 4th edn. Brooks Cole