Annals of Operations Research
https://doi.org/10.1007/s10479-018-3101-y

S.I.: MCDM 2017

On the analytical derivation of efficient sets in quad-and-higher criterion portfolio selection

Yue Qi[1] · **Ralph E. Steuer**[2]

Abstract
This paper provides results in the area of the analytical derivation of the efficient set of a mean-variance portfolio selection problem that has more than three criteria. By "analytical" we mean derived by formula as opposed to being computed by algorithm. By "more than three criteria", we mean that beyond the mean and variance of regular portfolio selection, the problems addressed have two or more additional linear objectives. The additional objectives might include sustainability, dividend yield, liquidity, and R&D as extra objectives like these are being seen with greater frequency. While not all multiple criteria portfolio selection problems lend themselves to an analytical derivation, a certain class does and the problems in this class are covered by the mathematics of this paper.

Keywords Multiple criteria portfolio selection · Analytical derivation · Minimum-variance surface · Nondominated set · Efficient set · Paraboloid

1 Introduction

From Markowitz (1952) we have the mean-variance model of portfolio selection which can be expressed, in bi-criterion format, as

$$\begin{aligned} \min\{z_1 &= \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}\} \\ \max\{z_2 &= \boldsymbol{\mu}^T \mathbf{x}\} \\ \text{s.t.} \ \ \mathbf{1}^T \mathbf{x} &= 1 \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \tag{1}$$

where $\mathbf{x}$ is a vector specifying the proportion of capital invested in each stock. Because this vector specifies a portfolio, $\mathbf{x}$ is often called a *portfolio*. In this model, because of the nonnegativity restriction on $\mathbf{x}$, short selling is not allowed. With $\boldsymbol{\Sigma}$ an $n \times n$ covariance matrix of stock returns, the first objective (z_1) is portfolio variance. With $\boldsymbol{\mu}$ a vector of expected stock returns, the second objective (z_2) is portfolio expected return.

✉ Ralph E. Steuer
rsteuer@uga.edu

1 Department of Financial Management, Business School Nankai University, Tianjin 300071, China

2 Terry College of Business, University of Georgia, Athens, GA 30602, USA

Published online: 28 November 2018

Let $S \subset \mathbb{R}^n$ be the feasible region in *decision space*. Here $S = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{1}^T\mathbf{x} = 1,\ \mathbf{x} \geq \mathbf{0}\}$. Because the problem has more than one objective, the problem has a second representation of the feasible region, designated $Z \subset \mathbb{R}^k$, called the feasible region in *criterion space*, where k is the number of objectives. Criterion space is the space of the objectives, and any $\mathbf{z} \in Z$ is a *criterion vector*. With a criterion vector resulting from the values of the objectives at a given point in S, here Z is given by

$$Z = \{\mathbf{z} \in \mathbb{R}^2 \mid z_1 = \mathbf{x}^T \mathbf{\Sigma} \mathbf{x},\ z_2 = \boldsymbol{\mu}^T \mathbf{x},\ \mathbf{x} \in S\}.$$

In this way, Z is the set of all images of the points in S, and S is the set of all inverse images of the criterion vectors in Z.

With respect to Z, a criterion vector $\bar{\mathbf{z}} \in Z$ is *nondominated* iff there exists no $\mathbf{z} \in Z$ such that $z_1 \leq \bar{z}_1,\ z_2 \geq \bar{z}_2$ with $\mathbf{z} \neq \bar{\mathbf{z}}$. The set of all nondominated criterion vectors is called the nondominated set and is designated N. Similarly, $\bar{\mathbf{z}} \in Z$ is *weakly nondominated* iff there exists no $\mathbf{z} \in Z$ such that $z_1 < \bar{z}_1,\ z_2 > \bar{z}_2$. The difference in the directions of the inequalities is due to the first objective being in minimization form and the second objective being in maximization form. Note that the set of all weakly nondominated criterion vectors subsumes N. With respect to S, a portfolio $\bar{\mathbf{x}} \in S$ is *efficient* iff $\bar{\mathbf{x}}$ is an inverse image of some nondominated vector in Z. The set of all efficient portfolios is called the efficient set and is designated E. Similarly, $\bar{\mathbf{x}} \in S$ is *weakly efficient* iff $\bar{\mathbf{x}}$ is an inverse image of some weakly nondominated vector in Z.

To define optimality in a portfolio problem, let $U : Z \rightarrow \mathbb{R}$ be the decision maker's utility function. Hence, a $\mathbf{z}^* \in Z$ that maximizes U is an *optimal* criterion vector, and any inverse image $\mathbf{x}^* \in S$ of $\mathbf{z}^*$ is an *optimal* portfolio. We are interested in nondominated criterion vectors and efficient portfolios because under a U in which *less-is-better-than-more* for all minimization objectives and *more-is-better-than-less* for all maximization objectives, $\mathbf{z}^* \in N$ and $\mathbf{x}^* \in E$. Thus, it suffices to find one's most preferred nondominated criterion vector to find an optimal portfolio.

With this, the idea of Markowitz is to, first, compute all nondominated criterion vectors. Second, portray them graphically which, when doing so, takes on the form of what is called the *nondominated frontier*. Third, have the decision maker select from the nondominated frontier his or her's most preferred point on it. Fourth, obtain an optimal portfolio of the decision maker by taking an inverse image of the selected criterion vector.

In steps three and four we see two reasons why Markowitz's model has maintained its position as the most influential model in portfolio selection since its inception over sixty-five years ago. One is that it allows different decision makers to have different optimal solutions. A second is that, when attempting to identify one's optimal portfolio, one does so within the presence and knowledge of all other candidates for optimality. This is useful when a decision maker doesn't naturally like his optimal solution, but only comes to accept it because it can be seen that everything else is worse.

But perhaps the most frequently given reason for the success of Markowitz's mean-variance model is its mathematical tractability. This is demonstrated in the analytical derivation paper by Merton (1972) which is a classic in the area. In this paper, with $S = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{1}^T\mathbf{x} = 1\}$ so short sales are allowed, Merton (analytically) derives formulas for the efficient set and other salient quantities. In the paper, Merton also proves the mean-variance nondominated frontier to be the vertex and upper portion of a rightward opening parabola.

Despite the success of Markowitz's model, it has not been without criticisms. One of the most persistent is that it does not accommodate additional criteria such as sustainability, dividend yield, liquidity, R&D, and others that are now being mentioned in the literature (see

Lo et al. 2003; Guerard and Mark 2003; Steuer et al. 2007; Ballestero et al. 2012; Utz et al. 2015; and so forth) with greater frequency. With the additional objectives in linear form, this gives rise to the following *extended* mean-variance formulation

$$\begin{aligned} \min \{z_1 &= \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}\} \\ \max \{z_2 &= \boldsymbol{\mu}^{2^T} \mathbf{x}\} \\ &\vdots \\ \max \{z_k &= \boldsymbol{\mu}^{k^T} \mathbf{x}\} \\ \text{s.t.} \quad & \mathbf{x} \in S \end{aligned} \tag{2}$$

where $\boldsymbol{\mu}^2$ is the $\boldsymbol{\mu}$ of (1), $\boldsymbol{\mu}^3 \ldots\ \boldsymbol{\mu}^k$ are the coefficient vectors of the additional criteria, and $z_3\ \ldots\ z_k$ are criterion values. All of the other notation and concepts are the same after care is exercised to take the extra criteria into account. For instance, when $Z \subset \mathbb{R}^k$ where $k > 2$, a $\bar{\mathbf{z}} \in Z$ is nondominated iff there exists no $\mathbf{z} \in Z$ such that $z_1 \leq \bar{z}_1,\ z_2 \geq \bar{z}_2, \ldots, z_k \geq \bar{z}_k$ with $\mathbf{z} \neq \bar{\mathbf{z}}$, and so forth. In this way, we have the extension of (1) to (2).

2 Analytical versus algorithmic

Portfolio problems such as (1) and (2), but where the feasible region is simply given by S, where S can vary in how it is defined by linear constraints, can be divided into two classes. One class consists of problems amenable to analytical methods. By this we mean problems for which we can solve for quantities such as the efficient set, the nondominated set, special points within these sets, and so forth, using formulas that have already been derived by analytical methods or use analytical methods to derive formulas for quantities for which formulas do not yet exist. The analytical methods of which we speak mostly involve calculus and matrix algebra. Formulas that have been derived by analytical methods have been found to be very insightful and are heavily utilized on the theoretical side of portfolio selection.

The other class of portfolio problems consists of those that are not amenable to analytical methods. Their efficient sets, nondominated sets, special points within these sets, and so forth, can only be solved by algorithmic methods. By this we mean mostly techniques from mathematical programming or evolutionary procedures (Deb 2001). Unfortunately, such techniques typically require considerable computation that can easily run into the minutes and longer. Obviously, one would prefer to obtain needed results by formula, but not all problems qualify.

To be amenable to analytical methods, a problem must meet two requirements. One is that the problem must possess a positive definite[1] covariance matrix. The other is that the problem's feasible region S must be defined by equality constraints. In Merton (1972), the assumed covariance matrix and S in the form of $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{1}^T \mathbf{x} = 1\}$ meet these requirements. While such an S might be overly simplistic in practice, this is not a drawback in theoretical portfolio selection. Allowing short selling, the simplicity of the feasible region better enables one to express the mathematical relationships that exist among the various sets, points of special interest, and properties of a mean-variance model.

As for a review of the literature on the use of analytical methods in portfolio selection, for problems with more than two assets, there was little other than graphs for the display of

[1] A matrix $\mathbf{P}$ is (i) *positive definite* if the scalar $\mathbf{x}^T \mathbf{P} \mathbf{x}$ is positive for all nonzero vectors $\mathbf{x}$, and (ii) *positive semidefinite* if the scalar $\mathbf{x}^T \mathbf{P} \mathbf{x}$ is nonnegative for all nonzero vectors $\mathbf{x}$.

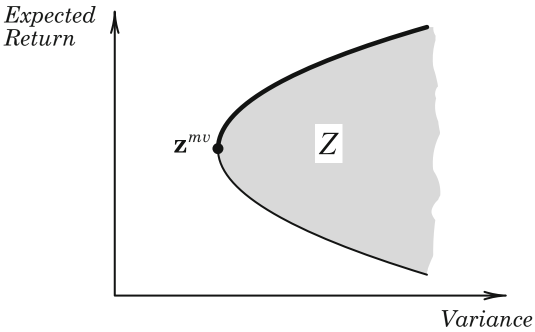

Fig. 1 For $k = 2$, with *Variance* = $\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}$ on the horizontal axis, and *Expected Return* = $\boldsymbol{\mu}^T \mathbf{x}$ (also the "mean" in mean-variance) on the vertical axis, the graph shows feasible region Z in criterion space. In the graph we also see the curve of the minimum-variance boundary, the minimum-variance point $\mathbf{z}^{mv}$, and the nondominated frontier as the dark line top of the minimum-variance boundary

quantitative information prior to 1972, but a complete job on the mathematics of (2) with $k = 2$ was not done until Merton (1972). The problem with a complete job is that it leaves little for others in its wake, and that is true about Merton (1972).

One of the benefits of Merton (1972) is that the material has been especially useful in instruction. This is reflected in the treatments of the material in Roll (1977), Ingersoll (1987), Huang and Litzenberger (1988), Campbell et al. (1997) and Luenberger (1997) where each represents advances in the presentation of the material. Consequently, these references are popular sources on the theory of portfolio selection in Ph.D. programs in finance. However, it must be pointed out that the treatments in these references only cover *standard* mean-variance portfolio selection. By "standard", we mean when $k = 2$. Resulting from the influences these sources have had over the last twenty years, the $k = 2$ analytical side of portfolio selection is often seen in the construction of models in empirical finance.

To illustrate some of the sets, special points, and properties of standard $k = 2$ portfolio selection that can be analytically determined by the treatments in the above references, consider Fig. 1. As noticed from the axes, the figure is in criterion space. The curve in the figure is a rightward opening parabola. This curve is the boundary of the feasible region Z, where Z is the curve and everything to the right of it. The curve is called the *minimum-variance boundary*, and the point on it labeled $\mathbf{z}^{mv}$ is the minimum-variance point. The inverse image of this point is the minimum-variance portfolio. The upper portion of the minimum-variance boundary plus the vertex of the parabola (i.e., the minimum-variance point) is the nondominated set (frontier). The inverse image set of the nondominated frontier is the mean-variance efficient set.

Different from the analytical side, problems on the algorithmic side of portfolio selection are allowed to have covariance matrices that are merely positive semidefinite (not necessarily positive definite) and feasible regions S that have inequality constraints in their specification. These problems cannot be solved analytically. By the way, algorithmic methods normally can not solve problems that analytical methods can solve because all of the algorithmic methods for solving portfolio problems of which we are aware require S to be bounded. At this point, we only mention a few $k = 2$ references on this side of portfolio selection such as Stein et al. (2008), Niedermayer and Niedermayer (2010), Hirschberger et al. (2010) and Woodside-Oriakhi et al. (2011), as not $k = 2$, but $k > 3$ is the focus in this paper.

For when $k = 3$, that is, when (2) contains one quadratic and two linear objectives, the only analytical derivation paper of which we are aware is Qi et al. (2017). In it, among other things, it is shown how the feasible region Z in criterion space has as its boundary a rightward opening *paraboloid*, how the minimum-variance boundary is now more appropriately called the *minimum-variance surface*, and how the nondominated set, being a portion of the paraboloid, is called the *nondominated surface*. Like Fig. 1, all of this can be shown as it is in Qi et al. (2017), but it requires 3D. For when $k > 3$, we are confident in saying that there has been nothing in the literature up until now on the analytical side. This paper contributes to the literature in this area on the analytical side.

In contrast to the analytical side, there has been research activity, mostly since about 2000, on the algorithmic side when $k \geq 3$. Apart from attempting to compute the whole nondominated surface when $k = 3$ as in the case of Hirschberger et al. (2013), strategies when $k \geq 3$ have generally been to convert the problem to a single-objective one and then apply (a) a weighted-sums approach, (b) an epsilon constraint approach, or (c) a goal programming approach. Notable references in this regard include Ehrgott et al. (2004), Ben Abdelaziz et al. (2007), Calvo et al. (2011), Ballestero et al. (2012) and Aouni et al. (2014).

With $k > 3$, covariance matrix $\boldsymbol{\Sigma}$ positive definite, and an equality-constraint feasible region in decision space generalized to $S = \{\mathbf{x} \mid \mathbf{A}^T\mathbf{x} = \mathbf{b}\}$, we have the rest of the paper. In Sect. 3 we analytically derive the minimum-variance surface. In Sect. 4 we analytically derive the efficient and nondominated sets. In Sect. 5, we discuss the dimensionality of the efficient set. In Sect. 6 we study relationships among efficient sets as additional linear objectives are added to a model. In Sect. 7 we provide an illustration of the material of the paper, and in Sect. 8 we discuss future directions.

3 Deriving the minimum-variance surface

To begin the process of providing the analytical derivations of this paper that cover the four-objective and higher cases of the extended mean-variance formulation of (2) with S generalized to $S = \{\mathbf{x} \mid \mathbf{A}^T\mathbf{x} = \mathbf{b}\}$, $\mathbf{b} \in \mathbb{R}^m$, we first focus on developing a formula for the minimum-variance boundary, which of course is a surface when $k > 2$. Highlighting the key assumptions used in this paper, we have

Assumption 1 Covariance matrix $\boldsymbol{\Sigma}$ is positive definite (and thus invertible).

Assumption 2 Vectors $\boldsymbol{\mu}^2 \ldots \boldsymbol{\mu}^k$ and all rows of $\mathbf{A}^T$ (altogether $k-1+m$ vectors) are linearly independent.

Applying an epsilon-constraint method to (2), we have

$$\begin{aligned} &\min \{z_1 = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}\} \\ &\text{s.t.} {\boldsymbol{\mu}^2}^T\mathbf{x} = e_2 \\ &\quad\vdots \\ &{\boldsymbol{\mu}^k}^T\mathbf{x} = e_k \\ &\mathbf{A}^T\mathbf{x} = \mathbf{b} \end{aligned} \tag{3}$$

where the e_i, $i = 2 \ldots k$, are parameters. The union of all optimal criterion vectors resulting from varying the e_i, $i = 2 \ldots k$, over all possible real values is called the minimum-variance surface.

We now employ the following notation

$$\mathbf{M} = \left[\boldsymbol{\mu}^2 \ \ldots \ \boldsymbol{\mu}^k \ \mathbf{A}\right]_{n\times(k-1+m)} \qquad \mathbf{f} = \begin{bmatrix} e_2 \\ \vdots \\ e_k \\ \mathbf{b} \end{bmatrix}_{(k-1+m)\times 1}$$

We rewrite the constraints of (3) as $\mathbf{M}^T\mathbf{x} = \mathbf{f}$. The system of linear equations $\mathbf{M}^T\mathbf{x} = \mathbf{f}$ is always solvable by Assumption 2.

Taking the Lagrangian to solve (3), we obtain

$$L(\mathbf{x}, \boldsymbol{\ell}) = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} + \boldsymbol{\ell}^T(\mathbf{M}^T\mathbf{x} - \mathbf{f})$$

where $\boldsymbol{\ell}$ is the vector of Lagrange multipliers. Because $L(\mathbf{x}, \boldsymbol{\ell})$ is convex, $\mathbf{x}$ is the optimal solution of (3) if and only if

$$\frac{\partial L}{\partial \mathbf{x}} = 2\boldsymbol{\Sigma}\mathbf{x} + \mathbf{M}\boldsymbol{\ell} = \mathbf{0}$$
$$\frac{\partial L}{\partial \boldsymbol{\ell}} = \mathbf{M}^T\mathbf{x} - \mathbf{f} = \mathbf{0}$$

Pre-multiplying the first equation by $\boldsymbol{\Sigma}^{-1}$ yields $\mathbf{x} = -\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\ell}$. Substituting $\mathbf{x}$ into the second equation yields $(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})\boldsymbol{\ell} = -2\mathbf{f}$. We now have Lemma 1 to help us with the invertibility of $\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M}$. The invertibility is used to compute the minimum-variance surface.

Lemma 1 *Let* $\mathbf{T}$ *be an* $n \times h$ *matrix of linearly independent columns. Then* $\mathbf{T}^T\boldsymbol{\Sigma}^{-1}\mathbf{T}$ *is positive definite (and hence invertible).*

Proof Since $\boldsymbol{\Sigma}$ is positive definite, $\boldsymbol{\Sigma}^{-1}$ is positive definite (Lax 2007). This means that there exists a random vector $\mathbf{s} \in \mathbb{R}^n$ such that $\boldsymbol{\Sigma}^{-1}$ is its covariance matrix. Consider $\mathbf{T}^T\mathbf{s}$. Then, $\mathbf{T}^T\boldsymbol{\Sigma}^{-1}\mathbf{T}$ is the covariance matrix of $\mathbf{T}^T\mathbf{s}$ (Brockwell and Davis 1987, pp. 33-35). Let $\mathbf{y} \neq \mathbf{0} \in \mathbb{R}^h$. Of course $\mathbf{y}^T(\mathbf{T}^T\boldsymbol{\Sigma}^{-1}\mathbf{T})\mathbf{y} = (\mathbf{T}\mathbf{y})^T\boldsymbol{\Sigma}^{-1}(\mathbf{T}\mathbf{y})$. Letting $\mathbf{u} = \mathbf{T}\mathbf{y}$, $\mathbf{u} \neq \mathbf{0}$ because $\mathbf{T}$ has full column rank by assumption. Then, $\mathbf{y}^T(\mathbf{T}^T\boldsymbol{\Sigma}^{-1}\mathbf{T})\mathbf{y} = \mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u} > 0$, because $\boldsymbol{\Sigma}^{-1}$ is positive definite. Therefore $\mathbf{T}^T\boldsymbol{\Sigma}^{-1}\mathbf{T}$ is positive definite and thus invertible. □

Substituting $\mathbf{M}$ for $\mathbf{T}$, on the basis of Lemma 1, we pre-multiply $(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})\boldsymbol{\ell} = -2\mathbf{f}$ by $(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}$ to obtain $\boldsymbol{\ell} = -2(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{f}$. We substitute $\boldsymbol{\ell}$ into $\mathbf{x} = -\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\ell}$ to obtain the optimal solution of (3) as

$$\mathbf{x} = \boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{f} \tag{4}$$

We now state Lemma 2 concerning the rank of $\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}$. The rank is used to demonstrate the dimensionality of the inverse image set of the minimum-variance surface in Theorem 1.

Lemma 2 *The* $n \times (k - 1 + m)$ *matrix* $\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}$ *has full column rank.*

Proof We know that for matrices $\mathbf{B}$ and $\mathbf{C}$ such that $\mathbf{BC}$ exists, rank $(\mathbf{BC}) \leq \min\{\text{rank } (\mathbf{B}), \text{rank } (\mathbf{C})\}$. Pre-multiplying (4) by $\mathbf{M}^T$, observe that

$$\mathbf{M}^T(\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}) = (\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1} = \mathbf{I}_{k-1+m}$$

where $\mathbf{I}_{k-1+m}$ is an identity matrix of order $k - 1 + m$. Letting $\mathbf{B} = \mathbf{M}^T$ and $\mathbf{C} = \boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}$, then rank $(\mathbf{C}) \geq k - 1 + m$. But because of the dimensions of $\mathbf{C}$, we know that the rank of $\mathbf{C}$ is at most $k - 1 + m$. Thus, the rank of $\mathbf{C} = \boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}$ is $k - 1 + m$. □

To express in the form of columns, let

$$\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1} = \left[\mathbf{p}^2 \ \ldots \ \mathbf{p}^k \ \mathbf{q}^1 \ \ldots \ \mathbf{q}^m\right]_{n\times(k-1+m)} \tag{5}$$

From (4), we obtain the inverse image set of the minimum-variance surface as

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = e_2\mathbf{p}^2 + \ldots + e_k\mathbf{p}^k + (b_1\mathbf{q}^1 + \ldots + b_m\mathbf{q}^m), \text{ for all } e_i \in \mathbb{R}\} \tag{6}$$

We now state Theorem 1.

Theorem 1 *The inverse image set of the minimum-variance surface is a* $(k-1)$*-dimensional affine subspace in* $\mathbb{R}^n$*. The subspace is spanned by* $\mathbf{p}^2 \ldots \mathbf{p}^k$ *and then translated to* $b_1\mathbf{q}^1 + \ldots + b_m\mathbf{q}^m$*. Any portfolio* $\mathbf{x}$ *whose criterion vector is on the minimum-variance surface can be expressed as a linear combination of any k affinely independent portfolios whose criterion vectors are also on the minimum-variance surface.*

The theorem follows from Lemma 2, (6), and the affine subspace property that a $(k-1)$-dimensional affine subspace is determined by any k affinely independent vectors.

From (6), we see that $b_1\mathbf{q}^1 + \ldots + b_m\mathbf{q}^m$ is the optimal solution of (3) when $e_2 = 0, \ldots, e_k = 0$. We substitute (4) into $z_1 = \mathbf{x}^T \boldsymbol{\Sigma}\mathbf{x}$ to compute the minimum-variance surface as

$$\begin{aligned} z_1 &= (\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{f})^T \, \boldsymbol{\Sigma} \, (\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{f}) \\ &= (\mathbf{f}^T (\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{\Sigma}^{-1}) \, \boldsymbol{\Sigma} \, (\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{f}) \\ &= \mathbf{f}^T (\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{f} \\ &= \mathbf{f}^T (\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{f} \end{aligned}$$

for all $z_i \in \mathbb{R}$, $i = 2 \ldots k$, in $\mathbf{f}$ where

$$\mathbf{f} = \begin{bmatrix} z_2 \\ \vdots \\ z_k \\ \mathbf{b} \end{bmatrix}$$

Substituting $z_2 \ldots z_k$ for the $e_2 \ldots e_k$ in $\mathbf{f}$, we have the following as the equation for the minimum-variance surface in the criterion space

$$z_1 = \begin{bmatrix} z^2 \ \ldots \ z^k \ \mathbf{b}^T \end{bmatrix} (\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1} \begin{bmatrix} z_2 \\ \vdots \\ z_k \\ \mathbf{b} \end{bmatrix} \tag{7}$$

Before analyzing the surface's property, we comment about paraboloids as follows. A paraboloid in $(y, x_1, \ldots, x_j)$ space can be written as $y = a_1x_1^2 + \ldots + a_jx_j^2$ with $a_1 \geq 0, \ldots, a_j \geq 0$ as coefficients. The x_i here have nothing to do with the $\mathbf{x} \in S$ mentioned elsewhere in the paper. If some $a_i = 0$, the paraboloid is *degenerate*. Otherwise, the paraboloid is *nondegenerate*. A paraboloid is degenerate if it has at least one *ruling* along which we may slide in either direction without ever leaving the paraboloid. For example, for $y = 0x_1^2 + x_2^2$ in (y, x_1, x_2) space, choose any point (e.g., $\begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}$) on the parabola $y = x_2^2$ in (y, x_2) subspace and then slide from that point in either x_1 direction. Then the complete ruling (i.e., $\{\begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} + t\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^3 \mid t \in \mathbb{R}\}$) still belongs to the paraboloid. A nondegenerate paraboloid does not have any such rulings.

Theorem 2 *The minimum-variance surface (7) is a nondegenerate paraboloid in criterion space.*

Proof Since $\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M}$ is positive definite, so is its inverse. Because $(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}$ is symmetric, by Lax (2007), there exists a $(k-1+m)\times(k-1+m)$ normal matrix $\mathbf{N}$ such that

$$(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1} = \mathbf{N}^T \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & v_{k-1+m} \end{bmatrix} \mathbf{N}$$

where $v_1 > 0, \dots, v_{k-1+m} > 0$ are the eigenvalues of $(\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}$. Let

$$\begin{bmatrix} w_1 \\ \vdots \\ w_{k-1+m} \end{bmatrix} = \mathbf{N} \begin{bmatrix} z_2 \\ \vdots \\ z_k \\ \mathbf{b} \end{bmatrix}$$

Then the minimum-variance surface is

$$\begin{aligned} z_1 &= \left[z^2 \; \dots \; z^k \; \mathbf{b}^T\right](\mathbf{M}^T\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1} \begin{bmatrix} z_2 \\ \vdots \\ z_k \\ \mathbf{b} \end{bmatrix} \\ &= \left[z^2 \; \dots \; z^k \; \mathbf{b}^T\right]\mathbf{N}^T \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & v_{k-1+m} \end{bmatrix} \mathbf{N} \begin{bmatrix} z_2 \\ \vdots \\ z_k \\ \mathbf{b} \end{bmatrix} \\ &= \left[w_1 \; \dots \; w_{k-1+m}\right] \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & v_{k-1+m} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_{k-1+m} \end{bmatrix} \\ &= v_1 w_1^2 + \dots + v_{k-1+m} w_{k-1+m}^2 \end{aligned}$$

With the last m w_i being constants, note that the first $k-1$ w_i effect a change of coordinate system. With $v_1 > 0, \dots, v_{k-1+m} > 0$, this then demonstrates the minimum-variance surface to be a nondegenerate paraboloid after the change of coordinate system. □

With it established that the feasible region Z in criterion space is bounded by a nondegenerate paraboloid, we now explore the mathematics of the efficient set and how it produces the portion of the bounding paraboloid (where the bounding paraboloid is the minimum-variance surface) that is the nondominated set.

4 Deriving the efficient set and nondominated set

To identify the nondominated portion of the minimum-variance surface (where the minimum-variance surface is the bounding paraboloid), we form the weighted-sums model

$$\begin{aligned} &\min \{\alpha_1 \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} - \alpha_2 {\boldsymbol{\mu}^2}^T \mathbf{x} - \ldots - \alpha_k {\boldsymbol{\mu}^k}^T \mathbf{x}\} \\ &\text{s.t. } \mathbf{A}^T \mathbf{x} = \mathbf{b} \end{aligned} \tag{8}$$

where $\boldsymbol{\alpha} \geq \mathbf{0}$. We know from Geoffrion (1968) that the weakly efficient set is the set of all optimal solutions of (8) as a result of the fact that all objective functions and S are convex. We now show Theorem 3.

Theorem 3 *The efficient set equals the weakly efficient set.*

This is shown by demonstrating that all weakly nondominated criterion vectors are nondominated criterion vectors. The existence of a weakly nondominated criterion vector which is not nondominated would imply that the minimum-variance surface has at least one feasible ruling. However, this is not possible because of the absence of rulings possessed by nondegenerate paraboloids. Thus, each weakly nondominated criterion vector is nondominated, and consequently, each weakly efficient point is efficient.

Let

$$\tilde{\mathbf{M}} = \left[\boldsymbol{\mu}^2 \ \ldots \ \boldsymbol{\mu}^k\right]_{n\times(k-1)} \qquad \tilde{\boldsymbol{\alpha}} = \begin{bmatrix} \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}_{(k-1)\times 1}$$

Then, (8) can be rewritten as $\min \{\alpha_1 \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} - \tilde{\boldsymbol{\alpha}}^T \tilde{\mathbf{M}}^T \mathbf{x} \mid \mathbf{A}^T \mathbf{x} = \mathbf{b}\}$ in which there are two cases.

Case 1 is when $\alpha_1 = 0$. This causes the quadratic term to drop out. Then (8) reduces to the following linear program:

$$\begin{aligned} &\max \{\tilde{\boldsymbol{\alpha}}^T \tilde{\mathbf{M}}^T \mathbf{x}\} \\ &\text{s.t. } \mathbf{A}^T \mathbf{x} = \mathbf{b} \end{aligned} \tag{9}$$

Lemma 3 states the status of (9)'s optimality.

Lemma 3 *Linear program* (9) *does not have an upper bound (and thus has no optimal solution).*

This results from the fact that feasible region S is unbounded and that the objective function coefficient vector $\tilde{\boldsymbol{\alpha}}^T \tilde{\mathbf{M}}^T$ is linearly independent of the rows of $\mathbf{A}^T$. In this way, (9) has no optimal solution. Thus nothing is lost when attempting to derive the efficient and nondominated sets by discarding the $\alpha_1 = 0$ case.

Case 2 is when $\alpha_1 > 0$. We divide the $\boldsymbol{\alpha}$ of (8) by α_1 to obtain $\left[1 \ \frac{\alpha_2}{\alpha_1} \ \ldots \ \frac{\alpha_k}{\alpha_1}\right]^T$. In this way, $\min \{\alpha_1 \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} - \alpha_2 {\boldsymbol{\mu}^2}^T \mathbf{x} - \ldots - \alpha_k {\boldsymbol{\mu}^k}^T \mathbf{x}\}$ and $\min \{\mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} - \frac{\alpha_2}{\alpha_1} {\boldsymbol{\mu}^2}^T \mathbf{x} - \ldots - \frac{\alpha_k}{\alpha_1} {\boldsymbol{\mu}^k}^T \mathbf{x}\}$ are equivalent when $\alpha_1 > 0$. Making the substitution $\frac{\alpha_2}{\alpha_1} \mapsto \lambda_2, \ldots, \frac{\alpha_k}{\alpha_1} \mapsto \lambda_k$, we then form

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix}_{(k-1)\times 1}$$

A financial interpretation of $\boldsymbol{\lambda}$ is that of a *risk tolerance* vector. For instance, the larger λ_i, the larger the decision maker's tolerance of taking on more risk in pursuit of more of objective i.

Taking the Lagrangian, we have

$$\tilde{L}(\mathbf{x}, \tilde{\boldsymbol{\ell}}) = \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} - \boldsymbol{\lambda}^T \tilde{\mathbf{M}}^T \mathbf{x} + \tilde{\boldsymbol{\ell}}^T (\mathbf{A}^T \mathbf{x} - \mathbf{b})$$

where $\tilde{\boldsymbol{\ell}}$ is the m-Lagrangian-multiplier vector. By this, $\mathbf{x}$ is the optimal solution of (8) if and only if

$$\frac{\partial \tilde{L}}{\partial \mathbf{x}} = 2\boldsymbol{\Sigma}\mathbf{x} - \tilde{\mathbf{M}}\boldsymbol{\lambda} + \mathbf{A}\tilde{\boldsymbol{\ell}} = \mathbf{0}$$
$$\frac{\partial \tilde{L}}{\partial \tilde{\boldsymbol{\ell}}} = \mathbf{A}^T\mathbf{x} - \mathbf{b} = \mathbf{0}$$

Pre-multiplying the first equation by $\boldsymbol{\Sigma}^{-1}$ yields $\mathbf{x} = \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{M}}\boldsymbol{\lambda} - \mathbf{A}\tilde{\boldsymbol{\ell}})$. Substituting $\mathbf{x}$ into the second equation yields $\mathbf{A}^T \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{M}}\boldsymbol{\lambda} - \mathbf{A}\tilde{\boldsymbol{\ell}}) - \mathbf{b} = \mathbf{0}$. Rearranging we have

$$(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})\tilde{\boldsymbol{\ell}} = \mathbf{A}^T\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{M}}\boldsymbol{\lambda} - 2\mathbf{b}$$

Substituting $\mathbf{A}$ for $\mathbf{T}$ in Lemma 1, $\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A}$ is invertible. Premultiplying the equation above by $(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}$, we obtain $\tilde{\boldsymbol{\ell}} = (\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{M}}\boldsymbol{\lambda} - 2\mathbf{b})$. Substituting $\tilde{\boldsymbol{\ell}}$ into $\mathbf{x} = \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{M}}\boldsymbol{\lambda} - \mathbf{A}\tilde{\boldsymbol{\ell}})$ gives us

$$\begin{aligned}
\mathbf{x} &= \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{M}}\boldsymbol{\lambda} - \mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{M}}\boldsymbol{\lambda} - 2\mathbf{b})) \\
&= \frac{1}{2}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{M}}\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{M}}\boldsymbol{\lambda}) + \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{b} \\
&= \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{I}_n(\tilde{\mathbf{M}}\boldsymbol{\lambda}) - \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{M}}\boldsymbol{\lambda}) + \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{b} \qquad (10)
\end{aligned}$$

where $\mathbf{I}_n$ is the identity matrix of order n. Combining the first two terms of (10), we have the optimal solution of (8) as a function of the risk tolerance vector $\boldsymbol{\lambda}$ in the form of

$$\mathbf{x} = \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{I}_n - \mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1})\tilde{\mathbf{M}}\boldsymbol{\lambda} + \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{b} \qquad (11)$$

Expressing the first term of the above in columns as

$$\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{I}_n - \mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1})\tilde{\mathbf{M}} = \left[\mathbf{g}^2 \ \ldots \ \mathbf{g}^k\right]_{n\times(k-1)} \qquad (12)$$

allows us to write (11) as

$$\mathbf{x} = \lambda_2\mathbf{g}^2 + \ldots + \lambda_k\mathbf{g}^k + \mathbf{x}^{mv} \qquad (13)$$

where

$$\mathbf{x}^{mv} = \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{b} \qquad (14)$$

is the minimum-variance portfolio. Note that $\mathbf{x}^{mv}$ is the optimal solution of (8) when $\lambda_2 = 0$, ..., $\lambda_k = 0$. We now have Theorem 4.

Theorem 4 *The efficient set E is a translated polyhedral cone.*

Proof From (13), the efficient set can be written as

$$E = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \lambda_2 \mathbf{g}^2 + \ldots + \lambda_k \mathbf{g}^k + \mathbf{x}^{mv}, \lambda_2 \ldots \lambda_k \geq 0\} \tag{15}$$

This is recognized as a polyhedral cone whose generators are the $\mathbf{g}^i$, $i = 2 \ldots k$, translated to the point $\mathbf{x}^{mv}$. □

The nondominated set N is obtained by substituting the points in E into the objective functions of (2). Moreover, the nondominated set is a paraboloidal subset of the (paraboloidal) minimum-variance surface. Since by Theorem 4 the efficient set is a polyhedral cone, in the next section we investigate the dimensionality of this cone.

5 Determining an efficient set's dimensionality

In this section we make use of the result that if we pre- or post-multiply a matrix with an unknown rank by an invertible matrix, the resultant matrix will have the same unknown rank (Lax 2007). This is done when it is easier to determine the rank of a resultant matrix.

It is clear from (15) that the dimensionality of the efficient set E is the rank of $\left[\mathbf{g}^2 \ \ldots \ \mathbf{g}^k\right]$, but the rank of $\left[\mathbf{g}^2 \ \ldots \ \mathbf{g}^k\right]$ may be less than $k-1$. With the rank of $\tilde{\mathbf{M}}$ being $k-1$, a sufficient condition for the rank of $\left[\mathbf{g}^2 \ \ldots \ \mathbf{g}^k\right]$ to be $k-1$ is, from the left-hand side of (12), for,

$$\mathbf{\Sigma}^{-1}(\mathbf{I}_n - \mathbf{A}(\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}) \tag{16}$$

to be an invertible $n \times n$ matrix. To compute the rank of this matrix, we pre-multiply it by $\mathbf{\Sigma}$ and then post-multiply the product by $\mathbf{\Sigma}$ to get

$$\mathbf{\Sigma} - \mathbf{A}(\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T \tag{17}$$

Sometimes, the matrix of (17), and hence of (16), may not be invertible. Consider the example in which $k = 3$, $n = 4$, $m = 2$, $\mathbf{\Sigma} = \mathbf{I}_4$ and $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$. Here, the matrix of (17) $= \begin{bmatrix} .5 & -.5 & 0 & 0 \\ -.5 & .5 & 0 & 0 \\ 0 & 0 & .5 & -.5 \\ 0 & 0 & -.5 & .5 \end{bmatrix}$, is singular, and has a rank of 2. Hence, the dimensionality of E in this case would be less than or equal to 2.

However, there is the special case of (2) below

$$\begin{aligned} \min \{z_1 &= \mathbf{x}^T\mathbf{\Sigma}\mathbf{x}\} \\ \max \{z_2 &= \boldsymbol{\mu}^{2^T}\mathbf{x}\} \\ &\vdots \\ \max \{z_k &= \boldsymbol{\mu}^{k^T}\mathbf{x}\} \\ \text{s.t.} \quad & \mathbf{1}^T\mathbf{x} = 1 \end{aligned} \tag{18}$$

which is recognized as Merton when $k = 2$, and as an extension of Merton when $k \geq 3$. In this problem, with just the one constraint of $\mathbf{1}^T\mathbf{x} = 1$, the efficient set is always of dimensionality of $k - 1$. Establishing this, we have the following.

Result 1 *The efficient set of* (18) *for $k \geq 2$ is a $(k-1)$-dimensional translated cone.*

Proof With $\mathbf{A} = \mathbf{1}$, let $a = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$, which is a scalar. Re-writing (10), we have

$$
\begin{aligned}
\mathbf{x} &= \frac{1}{2}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{M}}\boldsymbol{\lambda} - \frac{1}{2a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{M}}\boldsymbol{\lambda})\boldsymbol{\Sigma}^{-1}\mathbf{1} + \frac{1}{a}\boldsymbol{\Sigma}^{-1}\mathbf{1} \\
&= \frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\mu}^2 \ \dots \ \boldsymbol{\mu}^k\right]\begin{bmatrix}\lambda_2 \\ \vdots \\ \lambda_k\end{bmatrix} - \frac{1}{2a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\mu}^2 \ \dots \ \boldsymbol{\mu}^k\right]\begin{bmatrix}\lambda_2 \\ \vdots \\ \lambda_k\end{bmatrix})\boldsymbol{\Sigma}^{-1}\mathbf{1} + \frac{1}{a}\boldsymbol{\Sigma}^{-1}\mathbf{1} \\
&= \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\lambda_2\boldsymbol{\mu}^2 + \dots + \lambda_k\boldsymbol{\mu}^k) - \frac{1}{2a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}(\lambda_2\boldsymbol{\mu}^2 + \dots + \lambda_k\boldsymbol{\mu}^k))\boldsymbol{\Sigma}^{-1}\mathbf{1} + \frac{1}{a}\boldsymbol{\Sigma}^{-1}\mathbf{1} \\
&= \frac{1}{2}(\lambda_2\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2 + \dots + \lambda_k\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^k) - \frac{1}{2a}(\lambda_2\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2 + \dots + \lambda_k\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^k)\boldsymbol{\Sigma}^{-1}\mathbf{1} + \frac{1}{a}\boldsymbol{\Sigma}^{-1}\mathbf{1} \\
&= \frac{\lambda_2}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2 - \frac{1}{a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2)\boldsymbol{\Sigma}^{-1}\mathbf{1}) + \dots + \frac{\lambda_k}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^k - \frac{1}{a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^k)\boldsymbol{\Sigma}^{-1}\mathbf{1}) + \frac{1}{a}\boldsymbol{\Sigma}^{-1}\mathbf{1}
\end{aligned}
$$

Thus,

$$
\mathbf{g}^i = \tfrac{1}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^i - \tfrac{1}{a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^i)\boldsymbol{\Sigma}^{-1}\mathbf{1}) \qquad \text{for } i = 2\dots k
$$

Suppose that there exist $\gamma_i \in \mathbb{R}$, $i = 2\dots k$ such that $\sum_{i=2}^k \gamma_i \mathbf{g}^i = 0$, that is,

$$
\sum_{i=2}^{k} \gamma_i \tfrac{1}{2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^i - \tfrac{1}{a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^i)\boldsymbol{\Sigma}^{-1}\mathbf{1}) = 0
$$

$$
\sum_{i=2}^{k} \gamma_i \tfrac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^i - \sum_{i=2}^{k} \gamma_i \tfrac{1}{2a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^i)\,\boldsymbol{\Sigma}^{-1}\mathbf{1} = 0
$$

Upon observing that $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2 \dots \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^k$ and $\boldsymbol{\Sigma}^{-1}\mathbf{1}$ are linearly independent, it follows that the coefficients of the equation above are all zeros, that is, $\gamma_i\frac{1}{2} = 0, i = 2\dots k$ (thus $\gamma_i = 0, i = 2\dots k$), and $\sum_{i=2}^k \gamma_i \frac{1}{2a}(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^i) = 0$. With $\gamma_i = 0, i = 2\dots k$, $\mathbf{g}^2 \dots \mathbf{g}^k$ are linearly independent by definition. In other words, after making the substitution of $\mathbf{A} = \mathbf{1}$ in (12), $\boldsymbol{\Sigma}^{-1}(\mathbf{I}_n - \frac{1}{a}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1})\tilde{\mathbf{M}}$ has full column rank $k-1$. □

After specifying the efficient set and investigating its dimensionality, we turn our intention to an interesting question in the next section. As additional objectives are added to (2), what is the relationship between the existing efficient set and the next efficient set?

6 Subsuming relationship

In Steuer (1986), there is a multiple objective programming example in which the efficient set becomes smaller as an additional linear objective is added to the model. The question is: Can this happen in (2) when additional linear objectives are added one after the other? In this section we show that the answer is "no". That is, after mean and variance, the efficient set, if anything, only subsumes its former self as additional linear objectives are added. To demonstrate that efficient portfolios continue to remain efficient as additional linear objectives are added, we have Theorem 5.

Theorem 5 *Starting with the efficient set of the $k = 2$ mean-variance problem and with $S = \{\mathbf{x} \mid \mathbf{A}^T\mathbf{x} = \mathbf{b}\}$, the efficient set becomes a superset of itself whenever an additional linear objective is added to the formulation.*

Proof To demonstrate the superset relationship, we add an additional linear objective to (2) to form

$$\begin{aligned} \min\{z_1 &= \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x}\} \\ \max\{z_2 &= {\boldsymbol{\mu}^2}^T \mathbf{x}\} \\ &\vdots \\ \max\{z_{k+1} &= {\boldsymbol{\mu}^{k+1}}^T \mathbf{x}\} \\ \text{s.t.} \quad & \mathbf{A}^T \mathbf{x} = \mathbf{b} \end{aligned} \tag{19}$$

Substituting for $\tilde{\mathbf{M}}$ and $\boldsymbol{\lambda}$ in (11), the efficient points of (2) are given by

$$\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{I}_n - \mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1})\left[\boldsymbol{\mu}^2 \ \ldots \ \boldsymbol{\mu}^k\right]\begin{bmatrix} \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix} + \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{b}$$

whereas efficient points of (19) are given by

$$\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{I}_n - \mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1})\left[\boldsymbol{\mu}^2 \ \ldots \ \boldsymbol{\mu}^{k+1}\right]\begin{bmatrix} \lambda_2 \\ \vdots \\ \lambda_{k+1} \end{bmatrix} + \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{b}$$

Thus, the efficient sets of (2) and (19) are respectively

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \lambda_2\mathbf{g}^2 + \ldots + \lambda_k\mathbf{g}^k + \mathbf{x}^{mv}, \lambda_2 \ldots \lambda_k \geq 0\}$$

and

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \lambda_2\mathbf{g}^2 + \ldots + \lambda_{k+1}\mathbf{g}^{k+1} + \mathbf{x}^{mv}, \lambda_2 \ldots \lambda_{k+1} \geq 0\}$$

By comparing the two sets above, the efficient set of (19) is seen to be a superset of the efficient set of (2). □

To show how the efficient sets of (2) are translated cones that grow to supersets of themselves as additional objectives are added, we have Fig. 2. On the left is a representation of the efficient set when $k = 2$. In this case the efficient set is a half-ray translated to the minimum-variance portfolio $\mathbf{x}^{mv}$. In the middle of the figure is a representation of the efficient set after a second linear objective is added to make $k = 3$. Here the efficient set is a 2-dimensional cone translated to $\mathbf{x}^{mv}$. After a further linear objective is added to the model, we have the efficient set as on the right in the figure. In this case, with $k = 4$, the efficient set is, as seen, a 3-dimensional polyhedral cone translated to $\mathbf{x}^{mv}$. In this way, we see how the efficient sets subsume their former selves as additional objectives are added.

7 Illustrations

In terms of an instance of the *extended* mean-variance formulation (2) of this paper with $k = 4$ and $n = 5$, let us numerically illustrate the results of this problem using the results of this paper as follows. For the five stocks, let us select from the Dow Jones Industrial Average Index: 1. AXP (American Express), 2. DIS (Walt Disney), 3. JNJ (Johnson & Johnson), 4. KO (Coca Cola), and 5. (WMT) Wal-Mart Stores. For portfolios constructed out of these stocks, let the criteria (objectives) to be considered be

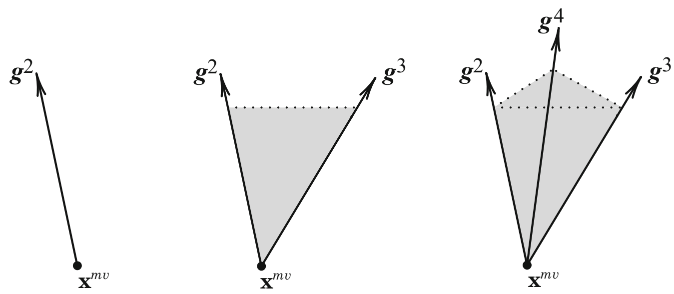

Fig. 2 Translated cones of the efficient set of (2) as linear objectives are successively added to the model

1. min { portfolio variance }
2. max { portfolio expected return without dividends }
3. max { portfolio expected dividend yield }
4. max { expected corporate social responsibility }

For the constraints of the problem, let

1. $\mathbf{1}^T\mathbf{x} = 1$
2. portfolio expected P/E (price to earnings) ratio = 17.

The data for the problem come from the five stocks' monthly returns without dividends, monthly dividend yields, annual CSR (corporate social responsibility), and annual P/E ratios from January 2010 to December 2013.[2] We compute the sample means and sample covariance matrix of the stocks' returns, and the sample means of the dividend yields. We then assume the sample statistics as population parameters and annualize by the method of Bodie et al. (2017, Chapter 13) in order to report $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}^2$, and $\boldsymbol{\mu}^3$. We also compute the sample means of the stocks' annual CSRs and report the sample means as population parameters in $\boldsymbol{\mu}^4$. After computing the sample means of the P/E ratios, we are able to report $\mathbf{A}$ and $\mathbf{b}$ below as well.

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.0413 & 0.0225 & 0.0110 & 0.0094 & 0.0076 \\ 0.0225 & 0.0440 & 0.0100 & 0.0109 & 0.0121 \\ 0.0110 & 0.0100 & 0.0198 & 0.0078 & 0.0102 \\ 0.0094 & 0.0109 & 0.0078 & 0.0162 & 0.0070 \\ 0.0076 & 0.0121 & 0.0102 & 0.0070 & 0.0217 \end{bmatrix},$$

$$\boldsymbol{\mu}^2 = \begin{bmatrix} 0.2239 \\ 0.2395 \\ 0.0981 \\ 0.1013 \\ 0.1078 \end{bmatrix} \boldsymbol{\mu}^3 = \begin{bmatrix} 0.0148 \\ 0.0132 \\ 0.0342 \\ 0.0286 \\ 0.0246 \end{bmatrix}, \boldsymbol{\mu}^4 = \begin{bmatrix} 8.0000 \\ 6.7500 \\ 10.5000 \\ 3.7500 \\ -1.7500 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 1 & 14.3273 \\ 1 & 17.7308 \\ 1 & 17.7038 \\ 1 & 16.9214 \\ 1 & 13.6995 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 17 \end{bmatrix}$$

[2] Data source of this paper: databases "CRSP" and "MSCI (formerly KLD and GMI)" via Wharton Research Data Services, https://wrds-web.wharton.upenn.edu/wrds/.

Using (6), the inverse image set of the minimum-variance surface is

$$\{\mathbf{x} \in \mathbb{R}^5 \mid \mathbf{x} = e_2 \begin{bmatrix} -45.4372 \\ 45.9736 \\ 59.0867 \\ -122.1046 \\ 62.4815 \end{bmatrix} + e_3 \begin{bmatrix} -358.4168 \\ 306.4312 \\ 482.8894 \\ -913.7268 \\ 482.8229 \end{bmatrix} + e_4 \begin{bmatrix} 0.2928 \\ -0.2482 \\ -0.2104 \\ 0.515 \\ -0.3492 \end{bmatrix}$$

$$+1 \begin{bmatrix} 16.2347 \\ -14.7916 \\ -16.9579 \\ 32.168 \\ -15.6532 \end{bmatrix} + 17 \begin{bmatrix} -0.1465 \\ 0.136 \\ -0.1211 \\ 0.3193 \\ -0.1877 \end{bmatrix}, \text{ for all } e_2, e_3, e_4 \in \mathbb{R}\}$$

Using (7), the equation of the minimum-variance surface (paraboloid) is

$$z_1 = \begin{bmatrix} z_2 & z_3 & z_4 & 1 & 17 \end{bmatrix} \begin{bmatrix} 339.1450 & 2568.8379 & -1.6541 & -95.8324 & -0.4015 \\ 2568.8379 & 19547.3448 & -12.5322 & -726.1840 & -3.1784 \\ -1.6541 & -12.5322 & 0.0084 & 0.4751 & 0.0014 \\ -95.8324 & -726.1840 & 0.4751 & 27.4817 & 0.0877 \\ -0.4015 & -3.1784 & 0.0014 & 0.0877 & 0.0025 \end{bmatrix} \begin{bmatrix} z_2 \\ z_3 \\ z_4 \\ 1 \\ 17 \end{bmatrix}$$

In this problem, the minimum-variance portfolio (14) is

$$\mathbf{x}^{mv} = \begin{bmatrix} -0.0121 \\ 0.0614 \\ 0.3698 \\ 0.4903 \\ 0.0905 \end{bmatrix}$$

With this as background, let us say that the decision maker, in looking over this problem in the form of (8), is curious about what would be his optimal portfolio if his weighting vector over the four objectives were $\boldsymbol{\alpha} = (.4, .3, .2, .1)$. Then, converting the weighting vector to a risk tolerance vector, we have $\boldsymbol{\lambda} = (.75, .50, .25)$. Then, in terms of the three risk tolerances, the efficient set, via (15), is given by

$$\{\mathbf{x} \in \mathbb{R}^5 \mid \mathbf{x} = \lambda_2 \begin{bmatrix} 1.073 \\ 1.6373 \\ -0.8019 \\ -1.261 \\ -0.6473 \end{bmatrix} + \lambda_3 \begin{bmatrix} -0.0755 \\ -0.2341 \\ 0.201 \\ 0.0577 \\ 0.0508 \end{bmatrix} + \lambda_4 \begin{bmatrix} 116.4001 \\ -41.7032 \\ 129.9236 \\ -131.9752 \\ -72.6453 \end{bmatrix} + \mathbf{x}^{mv},\ \lambda_2, \lambda_3, \lambda_4 \geq 0\}$$

and the decision maker's optimal portfolio is

$$\mathbf{x}^* = .75 \begin{bmatrix} 1.073 \\ 1.6373 \\ -0.8019 \\ -1.261 \\ -0.6473 \end{bmatrix} + .50 \begin{bmatrix} -0.0755 \\ -0.2341 \\ 0.201 \\ 0.0577 \\ 0.0508 \end{bmatrix} + .25 \begin{bmatrix} 116.4001 \\ -41.7032 \\ 129.9236 \\ -131.9752 \\ -72.6453 \end{bmatrix} + \mathbf{x}^{mv}$$

As a second illustration of the material of this paper, we demonstrate how the formula of (14) for the minimum-variance portfolio of this paper

$$\mathbf{x}^{mv} = \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{b}$$

collapses to Merton's formula for the minimum-variance portfolio given on page 1854 of his paper. This happens because in Merton, $\mathbf{A} = \mathbf{1} \in \mathbb{R}^n$ and $\mathbf{b} = 1$. Substituting into the above, we have

$$= \mathbf{\Sigma}^{-1}\mathbf{1}(\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1})^{-1}1$$
$$= \frac{1}{\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1}}\mathbf{\Sigma}^{-1}\mathbf{1}$$

which is precisely Merton's result but in matrix notation.

8 Future directions

With the analytical derivation literature at the $k = 2$ level remaining relatively stable over about the last 20 years until recently, things have now begun to change. Principally among these things is the idea that criteria beyond mean and variance may well be present in a portfolio problem which, as documented by Zopounidis et al. (2015) and Aouni et al. (2018), is a prospect that only appears to be gaining momentum. This has led to the $k = 3$ mean-variance analytical derivation paper by Qi et al. (2017) for when the number of objectives in portfolio selection is three.

With that paper, to our knowledge, being the only analytical derivation paper at the $k = 3$ level, and with no analytical derivation papers known to exist at higher levels of k, in this paper we have taken it upon ourselves to target the $k \geq 4$ case. Moreover, in our treatment of the $k \geq 4$ case, S is allowed to be defined by more than just one linear equality constraint. The only restriction is that the coefficient vectors of the linear objectives and of the constraints, when taken as a group, constitute a linearly independent set of vectors. Note that in this paper, nothing in the $k \geq 4$ case destroys anything in the $k = 3$ case of Qi et al. or in the $k = 2$ case of Merton.

As for future directions, one idea, suggested by a referee, is to use analytical methods to develop closed-form formulas for sensitive analysis. This could include formulas for changes in a decision maker's weighting vector and changes in the inputs to a problem. Something else that is a thought for the future is the task of working on getting the possibility of multiple criteria approaches in portfolio analysis better accepted in mainstream finance. There are several obstacles. One is that many people are of the position that additional criteria are "priced", that is, they are built into the price so there is no need to treat them explicitly. Another is that even if a criterion is not priced, a question is: Why can't the criterion be modeled as a constraint? Admittedly, overcoming obstacles is hard to do when theory is in short supply, as when $k > 3$, but it is hoped that this paper will change some of that.

In applying the research of this paper, one must be aware of the limitation that all objective functions are assumed to be monotonic. In this regard, it is to be noted that there is a growing literature on non-monotonic objective functions as described in Ghaderi et al. (2017), and as pointed out in Greco et al. (2014), as a result of inteactions among the criteria, non-monotonic objective functions can find their way into advanced applications in finance. We need to be aware of this.

Our hope in this paper is to show how there can be substantial theoretical support for idea of criteria beyond risk and return in finance. Certainly, one area in which research is welcome would be on how to incorporate multiple risk tolerance factors into portfolio analysis. Ultimately, some kind of multiple criteria CAPM (capital asset pricing model) formula would be highly desirable. However, in contrast to today's CAPM (see Bodie et al.

2017, Chapter 13), it would probably have to be made to some degree customizable by the decision maker because of the many different ways in which investors would likely wish to treat their additional criteria. In any event, any changes would probably have to get into the educational process first. One way to help in this regard would be to show students how the analytical treatments of conventional $k = 2$ mean-variance portfolio selection can be, as in the paper, extended to include additional criteria as in this paper.

Acknowledgements The first author acknowledges support from the National Social Science Fund of China 2018 (Grant No. 18BGL063). Also, the authors wish to thank two anonymous referees for their helpful comments.

References

Ben Abdelaziz, F., Aouni, B., & El-Fayedh, R. (2007). Multi-objective stochastic programming for portfolio selection. *European Journal of Operational Research*, *177*(3), 1811–1823.

Aouni, B., Colapinto, C., & La Torre, D. (2014). Financial portfolio management through the goal programming model: Current state-of-the-art. *European Journal of Operational Research*, *234*(2), 536–545.

Aouni, B., Doumpos, M., Pérez-Gladish, B., & Steuer, R. E. (2018). On the increasing importance of multiple criteria decision aid methods for portfolio selection. *Journal of the Operational Research Society*, *69*(10), 1525–1542.

Ballestero, E., Bravo, M., Pérez-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.

Bodie, Z., Kane, A., & Marcus, A. J. (2017). *Investments* (11th ed.). Boston: McGraw-Hill Irwin.

Brockwell, P. J., & Davis, R. A. (1987). *Time series: Theory and methods* (1st ed.). New York: Springer.

Calvo, C., Ivorra, C., & Liern, V. (2011). The geometry of the efficient frontier of the portfolio selection problem. *Journal of Financial Decision Making*, *7*(1), 27–36.

Campbell, J. Y., Lo, A. W., & Mackinlay, A. C. (1997). *The econometrics of financial markets*. Princeton: Princeton University Press.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. New York: Wiley.

Ehrgott, M., Klamroth, K., & Schwehm, C. (2004). An MCDM approach to portfolio optimization. *European Journal of Operational Research*, *155*(3), 752–770.

Geoffrion, A. M. (1968). Proper efficiency and the theory of vector maximization. *Journal of Matheatical Analysis and Applications*, *22*(3), 618–630.

Ghaderi, M., Ruiz, F., & Agell, N. (2017). A linear programming approach for learning non-monotonic additive value functions in multiple criteria decision aiding. *European Journal of Operational Research*, *259*(3), 1073–1084.

Greco, S., Mousseau, V., & Słowiński, R. (2014). Robust ordinal regression for value functions handling interacting criteria. *European Journal of Operational Research*, *239*(3), 711–730.

Guerard, J. B., & Mark, A. (2003). The optimization of efficient portfolios: The case for an R&D quadratic term. *Research in Finance*, *20*, 213–247.

Hirschberger, M., Qi, Y., & Steuer, R. E. (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., Steuer, R. E., Utz, S., Wimmer, M., & Qi, Y. (2013). Computing the nondominated set in tri-criterion portfolio selection. *Operations Research*, *61*(1), 169–183.

Huang, C. F., & Litzenberger, R. H. (1988). *Foundations for financial economics*. Englewood Cliffs: Prentice-Hall.

Ingersoll, J. E. (1987). *Theory of financial decision making*. New York: Rowman & Littlefield.

Lax, P. D. (2007). *Linear algebra and its applications*. Hoboken: Wiley.

Lo, A. W., 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.

Luenberger, D. G. (1997). *Investment science*. New York: Oxford University Press.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77–91.

Merton, R. C. (1972). An analytical derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, *7*(4), 1851–1872.

Niedermayer, A., & Niedermayer, D. (2010). Applying Markowitz's critical line algorithm. In J. B. Guerard (Ed.), *Handbook of portfolio construction* (pp. 383–400). Berlin: Springer.

Qi, Y., Steuer, R. E., & Wimmer, M. (2017). An analytical derivation of the efficient surface in portfolio selection with three criteria. *Annals of Operations Research*, *25*(1), 161–177.

Roll, R. (1977). A critique of the asset pricing theory's tests. *Journal of Financial Economics*, *4*(2), 129–176.

Stein, M., Branke, J., & Schmeck, H. (2008). Efficient implementation of an active set algorithm for large-scale portfolio selection. *Computers and Operations Research*, *35*(12), 3945–3961.

Steuer, R. E. (1986). *Multiple criteria optimization: Theory, computation, and application*. New York: Wiley.

Steuer, R. E., Qi, Y., & Hirschberger, M. (2007). Suitable-portfolio investors, nondominated frontier sensitivity, and the effect of multiple objectives on standard portfolio selection. *Annals of Operations Research*, *152*(1), 297–317.

Utz, S., Wimmer, M., & Steuer, R. E. (2015). Tri-criterion modeling for creating more-sustainable mutual funds. *European Journal of Operational Research*, *246*(1), 331–338.

Woodside-Oriakhi, M., Lucas, C., & Beasley, J. E. (2011). Heuristic algrithms for the cardinality constrained efficient frontier. *European Journal of Operational Research*, *213*(3), 538–550.

Zopounidis, C., Galariotis, E., Doumpos, M., Sarri, S., & Andriosopoulos, K. (2015). Multiple criteria decision aiding for finance: An updated bibliographic survey. *European Journal of Operational Research*, *247*(2), 339–348.