When do two- or three-fund separation theorems hold?

Carole Bernard,*
Corrado De Vecchi†
and Steven Vanduffel‡§

June 27, 2020

Abstract

We show that when asset returns satisfy a location-scale property (possibly conditionally as e.g., for a multivariate generalized hyperbolic distribution) and the investor has law-invariant and increasing preferences, the optimal investment portfolio always exhibits two-fund or three-fund separation. As a consequence, we recover many of the three-fund (and two-fund) separation theorems that have been derived in the literature under very specific assumptions on preferences or distributions. These are thus merely special cases from the general characterization result of optimal portfolios that we provide.

Keywords: Two-fund theorem, Three-fund theorem, Law-invariant preferences, Stochastic dominance, Investment analysis, Decision analysis.

*Corresponding author: Carole Bernard, Department of Accounting, Law and Finance at Grenoble Ecole de Management (email: carole.bernard@grenoble-em.com) and Department of Economics and Political Sciences at Vrije Universiteit Brussel (VUB).
†Corrado De Vecchi, Department of Economics and Political Sciences at Vrije Universiteit Brussel (VUB). (e-mail: corrado.de.vecchi@vub.ac.be).
‡Steven Vanduffel, Department of Economics and Political Sciences at Vrije Universiteit Brussel (VUB). (e-mail: steven.vanduffel@vub.ac.be).
§This work was supported by FWO Odysseus project FWOODYS11 of Carole Bernard at VUB.
1 Introduction

James Tobin and Harry Markowitz laid out the foundation of modern portfolio theory. Specifically, Tobin (1958) was the first to provide a (two-fund) separation theorem. He argued that in a world with one safe asset and a large number of risky assets, investors should combine cash with a single portfolio of risky assets. Owen and Rabinovitch (1983) point out that Tobin’s separation result holds for any stochastic return generating process if the investor’s utility function is quadratic and for any concave increasing utility function if the returns are multivariate normally distributed. Chamberlain (1983) extends this last result to the class of elliptical distributions.

The assumption of quadratic utility to justify a two-fund separation is however highly problematic. Indeed, quadratic utility implies increasing absolute risk-aversion, which has an unrealistic behavioral implication in that an increase in available wealth leads to lower investments in risky assets and not more (Huang and Litzenberger (1988)). Moreover, utility theory itself has been criticized for not being consistent with real-world decision making and a series of alternative decision theories have emerged. The most prominent amongst these is the so-called Cumulative Prospect Theory (CPT) from Tversky and Kahneman (1992). Levy and Levy (2004) show that when returns are normally distributed, optimal portfolios for CPT-investors are to be found in the set of mean-variance efficient portfolios. This result was generalized by Pirvu and Schulze (2012) who show that a two-fund separation theorem holds under elliptically distributed returns. However, whilst each of these alternative decision theories has its own features and is of interest, none of them is deemed suitable for accommodating all possible investors’ preferences. Therefore, in this paper we do not make specific assumptions on the choice of the behavioral theory, rather we only assume some properties that are rarely disputed. For instance, most adopted theories agree that more is better than less (compliance with first-order stochastic dominance) and that in addition, a certain income is better than an uncertain one with the same mean (compliance with second-order stochastic dominance). The results we derive hold for all the preferences that satisfy at least one of these two key properties.

As for the assumption that returns can be described by a multivariate elliptical model, some discussion is needed. As yearly returns are in essence sums of daily returns, one may expect that they display a Gaussian pattern; see Cesari and Cremonini (2003) for formal empirical evidence. However, studies based on daily returns show that asset returns typically exhibit skewness; see for instance Eberlein and Keller (1995), Kühler, Neumann, Sørensen, and Streller (1999), and Carr, Geman, Madan, and Yor (2002), amongst others. The effect of skewness on optimal port-
folio choice (under various theories of choice under risk) has been explored in a series of papers. Maximizing expected exponential utility and assuming a generalized hyperbolic (GH) skewed Student t-distribution for the returns, Birge and Chavez-Bedoya (2016) derive a three-fund theorem. Vanduffel and Yao (2017) extend this result by characterizing optimal portfolios for risk averse expected utility maximizers when returns follow a so-called multivariate generalized hyperbolic (MGH) distribution\(^1\) (which includes the GH skewed Student t-distribution as a special case), that is they find that all risk-averse expected utility maximizers invest in three funds only. Kwak and Pirvu (2018) also obtain three-fund separation but model preferences with Cumulative Prospect Theory (CPT).

In this paper we make very weak assumptions on preferences and derive two-fund theorems assuming the returns have a location-scale distribution and three-fund theorems when the returns are assumed to have a conditional location-scale property. Specifically, we show first that when returns have a location-scale property (equivalently, they are elliptically distributed), all decision theories that are compliant with First-order Stochastic Dominance (FSD) yield optimal portfolios that exhibit two-fund separation. So, while these theories may be very different they essentially lead to similar portfolio compositions. Second, when the returns only need to satisfy a conditional location-scale property (e.g., when they follow a MGH distribution) a three-fund separation can still be obtained. The proofs of these results are rather straightforward and generalize all mentioned specific results. For instance, Cumulative prospect theory preferences are consistent with FSD and hence the results of Pirvu and Schulze (2012) (see also Levy and Levy (2004)) in the elliptical case are a consequence of ours.

The remainder of the paper is organized as follows. In Section 2 we formulate the optimal portfolio problem and provide our assumptions relative to the preferences. In Section 3 we prove a two-fund separation theorem when returns have a location scale distribution. In Section 4 we prove a three-fund separation theorem when returns have conditional location-scale property (MGH distribution). Section 5 illustrates the theoretical results with a numerical application. We conclude in Section 6.

---

\(^1\)The MGH distribution was introduced in the literature by Barndorff-Nielsen (1978), Barndorff-Nielsen (1997) and Blaesild and Jensen (1981) and has shown to be useful for modeling asset returns (Barndorff-Nielsen (1997), McNeil, Frey, and Embrechts (2010)).
2 Problem formulation and preferences

In this paper we study the optimal allocation of wealth among \( n \) assets for an investor under fairly weak assumptions on his preferences. We first describe the market setting and then discuss the weak assumptions that we make on the investors’ preferences.

We consider a single period economy in which there are \( n+1 \) assets available for investment. There is one risk-free asset yielding a fixed return \( r > 0 \) and there are \( n \) risky assets yielding stochastic returns that are described by the vector \( \mathbf{X} = (X_1, \ldots, X_n)^T \) and have a joint distribution \( F_X \). We denote by \( \mathbf{m}^T = (m_1, \ldots, m_n)^T \) the vector of expected returns and by \( \Delta \) their positive definite covariance matrix. In what follows, we tacitly assume that all these quantities exist and are finite.

Let \( W_0 \) denote the total fixed initial wealth and \( \mathbf{\omega} := (\omega_1, \ldots, \omega_n)^T \) be the vector of amounts invested in the \( n \) different risky assets (the remaining amount is thus invested in the risk-free asset). We call \( \mathbf{\omega} \) a portfolio. The final wealth \( W_{\mathbf{\omega}} \) of the portfolio writes as

\[
W_{\mathbf{\omega}} = \sum_{i=1}^{n} \omega_i (1 + X_i) + \left( W_0 - \sum_{i=1}^{n} \omega_i \right) (1 + r)
= W_0 (1 + r) + \sum_{i=1}^{n} \omega_i (X_i - r).
\]

Denote by \( \mathcal{W} \) the set of final wealths that can be purchased with initial wealth \( W_0 \) and denote by \( V(\cdot) \) the investor’s objective function. The investor’s goal is to determine the optimal portfolio \( \mathbf{\omega}^* \) (equivalently, the optimal terminal wealth \( W_{\mathbf{\omega}^*} \in \mathcal{W} \)) by solving the following optimization problem

\[
\max_{\mathbf{\omega}} V(W_{\mathbf{\omega}}).
\]  

In this paper, we do not explicitly specify\(^2\) the objective function \( V(\cdot) \). Nevertheless, we state some properties that appear very natural for “reasonable” objective functions to satisfy. In what follows we denote by \( F_W \) the cumulative distribution function of a random terminal wealth \( W \in \mathcal{W} \) and by \( F_W^{-1} \) its quantile function (defined as the left inverse of \( F_W \)).

**Definition 2.1** Let \( W_1, W_2 \in \mathcal{W} \). We say that \( W_1 \) is first-order stochastically dominated by \( W_2 \), denoted as \( W_1 \prec_{\text{FSD}} W_2 \), if for all \( p \in (0, 1) \), \( F_W^{-1}(p) \leq F_{W_2}^{-1}(p) \).

\(^2\)It could for instance be an expected utility, i.e., \( V(W_{\mathbf{\omega}}) := E[U(W_{\mathbf{\omega}})] \) in which \( U(x) \) is some specific utility function. It could also refer to a non-expected utility setting such as the decision theories of Yaari (1987), Tversky and Kahneman (1992) or Quiggin (1993).
It is intuitive that investors when choosing between $W_1$ and $W_2$ will prefer $W_2$ whenever $W_1 \prec_{FSD} W_2$.

**Assumption 2.1 (FSD-consistency on $W$)** Preferences $V(\cdot)$ are consistent with first-order stochastic dominance ($FSD$) on $W$. That is, for $W_1, W_2 \in W$, $W_1 \prec_{FSD} W_2$ implies $V(W_1) \leq V(W_2)$ and equality only holds when $W_1$ and $W_2$ have the same distribution.

It is known that Assumption 2.1 is also equivalent to having a law-invariant and increasing objective function $V(\cdot)$ (see Theorem 1 in Bernard, Chen, and Vanduffel (2015)). So, being consistent with FSD is equivalent to assuming that “more is preferred to less” ($a < b \Rightarrow V(a) < V(b)$) and that optimal choices are only driven by the distribution of final wealth and not by the states in which cash-flows are received (law-invariance). Clearly, Assumption 2.1 is completely natural and most decision theories comply with it. In fact, many economists consider a violation of the FSD property as grounds for refuting a particular model; see, for example, Birnbaum (1997), Birnbaum and Navarrette (1998) for more discussions. Recall also that although the original prospect theory by Kahneman and Tversky (1979) provides explanations for phenomena that were unexplained before, it violates first-order stochastic dominance. To overcome this potential issue, Tversky and Kahneman (1992) have developed the cumulative prospect theory. In what follows, investors with preferences that comply with Assumption 2.1 are called *FSD-investors*.

**Definition 2.2** Let $W_1, W_2 \in W$. We say that $W_1$ is second-order stochastically dominated by $W_2$, denoted as $W_1 \prec_{SSD} W_2$, if for all $p \in (0, 1)$,

$$\int_0^p F_{W_1}^{-1}(q)dq \leq \int_0^p F_{W_2}^{-1}(q)dq.$$

**Assumption 2.2 (SSD-consistency)** Preferences $V(\cdot)$ are consistent with second-order stochastic dominance ($SSD$). That is, $W_1 \prec_{SSD} W_2$ implies $V(W_1) \leq V(W_2)$ and equality only holds when $W_1$ and $W_2$ have the same distribution.

For instance, expected utility maximizers that employ an increasing and concave utility function to make decisions have preferences $V(\cdot)$ that are SSD-consistent. That is, $W_1 \prec_{SSD} W_2$ implies $E(u(W_1)) \leq E(u(W_2))$, where $u(x)$ is an increasing and concave utility function. Clearly, SSD-consistency implies FSD-consistency but the opposite does not hold true in general. In general, being SSD-preserving is quite a strong assumption. For instance, preferences that are consistent with rank dependent utility theory exhibits FSD-consistency but not SSD-consistency (Ryan (2006)) and the same holds true for cumulative prospect theory (see e.g., Baucells and Heukamp (2006)). In what follows, investors with preferences that comply with Assumption 2.2 are called *SSD-investors*. 

5

Electronic copy available at: https://ssrn.com/abstract=3703665
3 Two-fund Separation Theorems

We first study a market model in which the joint distribution \( F_X \) of the vector of asset returns \( X \) is assumed to belong to a so-called location-scale family of distributions. We then derive a two-fund theorem for FSD-investors. We point out that various two-fund theorems that have been derived in the literature under specific assumptions on preferences (e.g., preferences according to the cumulative prospect theory Tversky and Kahneman (1992)) and on distributions (e.g., following an elliptical model) comply with this setting and are merely particular cases of the characterization result we provide. Furthermore, we show that in this market setting, under this market model SSD-investors cannot be distinguished from FSD-investors, i.e., being SSD-consistent is equivalent to being FSD-consistent.

3.1 Distributional Assumption on Returns

Definition 3.1 (Location-scale property of \( F_X \)) Let \( Z \) be a real-valued random variable taking values having zero mean and unit variance. We say that \( F_X \) of \( X = (X_1, X_2, ..., X_n) \) has the location-scale property associated with \( Z \) if for any vector \( a = (a_1, ..., a_n)^T \) it holds that

\[
a^T X \overset{d}{=} a^T m + \sqrt{a^T \Delta a} Z,
\]

where "\( \overset{d}{=} \)" denotes the equality in distribution and where we recall that \( m \) is the vector of expected returns of \( X \) and \( \Delta \) is their positive definite covariance matrix.

The family \( \mathcal{F} \) of all multivariate distributions that have this location-scale property is then called the location-scale family of distributions associated with \( Z \).

Specifically, if the joint distribution \( F_X \) of \( X = (X_1, X_2, ..., X_n) \) is a member of \( \mathcal{F} \), then for all \( i \), there exist \( m_i \in \mathbb{R} \) and \( \delta_i > 0 \) such that \( X_i \overset{d}{=} m_i + \delta_i Z \), where \( m_i \) is the mean and \( \delta_i \) is the standard deviation of \( X_i \). The distributional constraint (3) is fairly restrictive as it imposes a condition on the distribution of all linear combinations. In fact, Chamberlain (1983) provides a characterization result that makes it possible to conclude that the only family \( \mathcal{F} \) satisfying the condition in Definition 3.1 is the multivariate elliptical family, that is when \( Z \) is an elliptically distributed random variable. The equivalence of (i) and (ii) in Proposition 3.2 can be found in Theorem 1 of Chamberlain (1983).
Proposition 3.2 (Chamberlain (1983)) Let $X$ be a random vector of $\mathbb{R}^n$ with invertible covariance matrix $\Delta$ (with Cholesky decomposition $\Delta = LL^T$) and mean $m$. The three following statements are equivalent:

(i) For all $a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, for all $c \in \mathbb{R}$, the distribution of $a^T X + c$ is determined by its mean $a^T m + c$ and its variance $a^T \Delta a$.

(ii) $F_X$ belongs to the multivariate elliptical family with covariance matrix $\Delta = LL^T$ and mean vector $m$, i.e., $Z = L^{-1}(X - m)$ is spherically distributed.

(iii) $F_X$ belongs to a location-scale family $\mathcal{F}$ associated to a random variable $Z := Z_1$ (where $Z := L^{-1}(X - m)$) as described in Definition 3.1.

Proof. (iii)$\Rightarrow$(i): Let the distribution of $X$ be in $\mathcal{F}$, as defined in Definition 3.1. Then, for any $a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, the mean and the standard deviation of $a^T X$ are equal to $a^T m$ and $\sqrt{a^T \Delta a}$, respectively. Moreover, the cdf of $a^T X$ can be expressed as $F_{a^T X}(x) = F_Z\left( \frac{x-a^T m}{\sqrt{a^T \Delta a}} \right)$. Therefore, the distribution of $a^T X$ is completely specified by its mean $a^T m$, and its standard deviation $\sqrt{a^T \Delta a}$.

(i)$\Rightarrow$(ii): Assume that the distribution of $a^T X + c$ is characterized by its mean and its variance for all $a \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Using the Cholesky decomposition, there exists a triangular invertible matrix $L$ such that $\Delta = LL^T$. Define $T = L^{-1}$ and $Z = T(X - m)$. Then $E[Z] = 0$ and the covariance matrix of $Z$ is the identity matrix $I_n$ (because $\text{cov}(Z) = \text{cov}(L^{-1}X) = L^{-1}\text{cov}(X)[L^{-1}]^T = L^{-1}\Delta(L^{-1})^T = L^{-1}LL^T(L^{-1})^T = I_n$). Let $R$ be an orthogonal matrix, i.e. $RR^T = I$. Define $w = RZ$ then $E[w] = 0$ and the covariance matrix of $w$ is the identity matrix $I_n$ (because $\text{cov}(RZ) = R\text{cov}(Z)R^T$). For any $a \in \mathbb{R}^n$, $a^T Z = a^T T(X - m) = a_1^T X + c_1$ and $a^T w = a^T R T(X - m) = a_2^T X + c_2$ are two portfolios with the same mean 0 and the same variance $a^T a$. But by (i), the distributions of $a_1^T X + c_1$ and $a_2^T X + c_2$ are characterized by their means and variance, thus $a^T Z$ and $a^T w$ have the same distribution for all $a$. Thus we conclude that $Z$ and $w$ have the same distribution (a distribution is characterized by the distribution of all linear combinations as then for all $t \in \mathbb{R}^n$, $E[e^{it^T X}] = E[e^{it^TY}]$, i.e. the vectors $X$ and $Y$ have the same characteristic function and thus must have the same distribution). Hence $Z$ is spherically distributed about 0.

Proof of (ii)$\Rightarrow$(iii): It is well-known that the elliptical family of distributions satisfies (iii). See, for example, Section in 3.3 McNeil, Frey, and Embrechts (2010).
3.2 Characterization of Optimal Portfolios for FSD-investors

When the joint distribution \( F_X \) of the random return vector \( X = (X_1, X_2, ..., X_n) \) has the location-scale property associated with \( Z \), we obtain from (3) that the terminal wealth \( W_\omega \) defined by (1) satisfies

\[
W_\omega \overset{d}{=} m_\omega + \delta_\omega Z
\]

with parameters \( m_\omega \) and \( \delta_\omega \) given as

\[
\begin{align*}
  m_\omega &:= \mathbb{E}[W_\omega] = W_0(1 + r) + \omega^T(m-r1) \\
  \delta_\omega &:= \text{std}[W_\omega] = \sqrt{\omega^T \Delta \omega},
\end{align*}
\]

where \( 1 \) is a vector of ones. The terminal wealth \( W_\omega \) that arises from the portfolio allocation \( \omega \) is thus characterized by the coefficients \( m_\omega \) and \( \delta_\omega \) given in (5) and we can thus reformulate our optimization problem (2) as

\[
\max_{(\delta_\omega, m_\omega) \in A} V(m_\omega + \delta_\omega Z)
\]

where \( A \) is the set of all couples \((\delta_\omega, m_\omega)\), as in (5), i.e., \( A \) is given as

\[
A := \left\{ \left( \sqrt{\omega^T \Delta \omega}, W_0(1 + r) + \omega^T(m-r1) \right) \right\}_{\omega \in \mathbb{R}^n}
\]

Given that the optimization (6) only deals with the mean \( m_\omega \) and the standard deviation \( \delta_\omega \) of portfolios \( \omega \), it becomes apparent that a connection to the mean-variance analysis developed by Markowitz (1952) holds.

**Definition 3.3 (Mean-variance efficiency frontier)** Consider a portfolio \( \omega \) with terminal wealth \( W_\omega \) having mean \( m_\omega \) and variance \( \delta_\omega^2 \). The portfolio \( \omega \) is mean-variance efficient if there is no portfolio yielding a terminal wealth with the same variance but a strictly larger mean. The set \( A^* \subseteq A \) containing all pairs \((\delta_\omega, m_\omega)\) for which \( \omega \) is a mean-variance efficient portfolio is called the mean-variance efficiency frontier.

**Proposition 3.4** The mean-variance efficiency frontier \( A^* \) is explicitly given as

\[
A^* = \left\{ (\delta_\omega, m_\omega) \mid \delta_\omega > 0 \text{ and } m_\omega = W_0(1 + r) + \delta_\omega \sqrt{h} \right\}_{\omega \in \mathbb{R}^n},
\]

where

\[
h = (m - r1)^T \Delta^{-1} (m - r1) > 0.
\]
Proof. Given $\delta \omega = \delta$, let us build the mean-variance efficient portfolio, i.e. the portfolio that solves the optimization problem

$$\max_{\omega} \omega m \text{ subject to } \delta \omega = \delta. \tag{10}$$

This is a standard problem and using Lagrange multipliers one readily obtains that the optimal portfolio is given as

$$\omega_* \ := \ \frac{\delta}{\sqrt{h}} \Delta^{-1}(m - r1). \tag{11}$$

Using the fact that $\Delta^{-1}$ is symmetric, the expected return of this portfolio is given by

$$m_{\omega_*} = W_0(1 + r) + \omega_*^T(m - r1)$$
$$= W_0(1 + r) + \frac{\delta \omega_*}{\sqrt{h}} (m - r1)^T \Delta^{-1}(m - r1)$$
$$= W_0(1 + r) + \delta \omega_* \sqrt{h}.$$ 

It is clear from the characterization of the set $A^*$ that for each $(\delta \omega, m \omega) \in A^*$, there exists exactly one portfolio $\omega$ yielding this specific mean $m_w$ and variance $\delta_w^2$. In what follows we sometimes identify such portfolio $\omega$ with the pair $(\delta \omega, m \omega)$ and correspondingly call $A^*$ also the set of mean-variance efficient portfolios.

Proposition 3.5 (Two-fund theorem) When $V(\cdot)$ satisfies Assumption 2.1 and when Problem (2) has a solution $\omega_*$, then $(\delta \omega_*, m \omega_*) \in A^*$. Furthermore, an optimum $\omega_*$ is of the form

$$\omega_* = \frac{\delta}{\sqrt{h}} \Delta^{-1}(m - r1) \tag{12}$$

for some $\delta > 0$ that is such that $V(\omega)$ is maximum when $\omega = \omega_*$. 

Proof. Let $\omega_*$ be an optimal solution to Problem (2). Note that $\omega_*$ must be mean-variance efficient, i.e., it must maximize the mean for a given standard deviation $\delta$ (Problem (10)). Indeed, if $\omega_*$ is not mean-variance efficient one can find another portfolio $\omega$ which is dominating in the sense of FSD and thus yields a higher objective value. Its expression was already derived above in (11), which ends the proof.
There are several important implications from Proposition 3.5. First, regardless of their specific objective function, the optimal allocation in risky assets of FSD-investors is always proportional to $\Delta^{-1}(m - r)1$ and the remaining funds are invested in the risk-free asset. We label $\Delta^{-1}(m - r)1$ as the market-fund. Hence, the optimal portfolio of FSD-investors ultimately always translate in an optimal proportion that is allocated to the market-fund. This observation is important for practical investment advice, as eliciting the proportion (a single number) that an investor is prepared to allocate to the market-fund is much easier than eliciting his preferences (i.e., the objective function $V(\cdot)$ that he aims at maximizing). Second, solving the high-dimensional optimal portfolio problem in (2) amounts to solving a one-dimensional problem in which the only unknown is the parameter $\delta$ in (12), which maximizes the objective function. Third, this proposition shows that various contributions in the literature are merely special cases of the general characterization we provide. Levy and Levy (2004) show that under normally distributed returns, CPT-investors select their portfolio on the mean-variance efficient frontier. This result was generalized by Pirvu and Schulze (2012) for elliptically distributed returns. Finally, Bertsimas, Lauprete, and Samarov (2004) obtains under an elliptical model that a two-fund theorem holds when investors minimize the expected shortfall for a given desired expected return. However, all these results are immediately implied by Proposition 3.5. In addition, from the proposition it also follows that under the distributional assumption we make, two-fund separation holds for investors with preferences described by Rank Dependent Utility Theory (Quiggin (1993)).

3.3 Characterization of Optimal Portfolios for SSD-investors

Recall that every SSD-investor is also an FSD-investor. Hence, Proposition 3.5 also applies to SSD-investors and their optimal portfolios thus also exhibit two-fund separation. However, by exploiting the specific characteristics of SSD-investors it might be possible to obtain a more refined characterization of their optimal portfolio. In this regard, it can be shown that the optimal portfolio of an SSD-investor must also solve the problem

$$\min_{\omega} \delta_{\omega} \text{ subject to } m_{\omega} = m.$$ (13)

for some given $m > W_0(1 + r)$. Observe next that for every couple $(\delta_{\omega}, m_{\omega})$ in $A^*$ the corresponding portfolio $\omega$ must be a solution to a problem of the form (13), since otherwise there exists $\omega'$ such that $\delta_{\omega'} < \delta_{\omega}$ and $m_{\omega'} = m_{\omega}$, but this is not possible as we proved that the maximum attainable expected value is a strictly decreasing function of the standard deviation and $(\delta_{\omega}, m_{\omega})$ is

Electronic copy available at: https://ssrn.com/abstract=3703665
mean-variance efficient by assumption. This means that using the extra information that optimal portfolios must also solve a problem of the form (13) does not lead to a reduction of the set $A^*$.

**Remark 3.1 (Rationalization of portfolios)** In general, one cannot expect that for every mean-variance efficient portfolio $\omega$ there exists a utility maximizer for whom this portfolio is optimum. However, when the distribution $F_X$ of the random return vector $X = (X_1, X_2, \ldots, X_n)$ has the location-scale property this holds true. Indeed, recall that in this case we consider the problem

$$\max_{(\delta, m) \in A} V(m + \delta Z)$$

where $A$ is the set that contains all couples $(\delta, m)$ as in (5). Consider now expected utility preferences $V(\cdot) = \mathbb{E}(u(\cdot))$ in which the utility function is given as $u(x) = x - \alpha x^2$ for some $\alpha > 0$. In this case, our maximization problem reads as

$$\max_{(\delta, m) \in A} m - \alpha(\delta^2 + m^2)$$

Clearly, if there is a solution $(\delta, m)$ it must be such that $\delta$ is minimum for the given value of $m$. As this property holds true for all $(\delta, m) \in A^*$, we only need to show that for every $(\delta, m) \in A^*$ there exists $\alpha > 0$ such that $(\delta, m)$ solves problem (15). On $A^*$ the optimization problem writes as

$$\max_{m > \frac{W_0(1+r)}{h}} m - \alpha \left[ \frac{(m - \frac{W_0(1+r)}{h})^2}{h} + m^2 \right]$$

in which $h$ is as in (9). Differentiation with respect to $m$ and equating to zero yields that $\alpha = \frac{1}{2(\frac{W_0(1+r)}{h})} > 0$. Note that for any $\alpha \geq \frac{1}{2W_0(1+r)}$ the risk-free investment is optimal (i.e., $m = W_0(1+r)$ and $\delta = 0$).

**Remark 3.2 (Two-fund theorems without distributional assumptions)** In the previous section we have shown that two-fund separation holds for general preferences under the key assumption that the joint distribution $F_X$ of the return vector $X$ has a location-scale property. Several contributions in the literature also derive a two-fund theorem without making distributional assumptions on asset returns. In this case, however, one requires specific preferences in that these solely balance the expected return (“reward”) and the variance (“risk”) of the terminal wealth. Specifically, De Giorgi, Hens, and Mayer (2011) (Theorem 1) provides a two-fund theorem when the investor preferences

Electronic copy available at: https://ssrn.com/abstract=3703665
can be described by
\[ V(W) = f(m(W), \rho(W)), \]
where \( f \) is monotonically decreasing in the risk \( \rho(W) \) and monotonically increasing in the reward \( m(W) \).

In the next section, we generalize the two-fund theorem in both directions. As compared to De Giorgi, Hens, and Mayer (2011), we do not require specific risk-reward preferences but only need preferences to be FSD-consistent. In contrast, we need to make some distributional assumptions. However, the distributional assumption is not very strong as it complies well with properties of observed financial returns data.

4 Three-fund Separation Theorems

In this section, we significantly relax the assumption of location-scale invariance for the joint distribution \( F \) of the return vector \( X = (X_1, X_2, ..., X_n) \). We derive a three-fund theorem and discuss the implications thereof. Specifically, we point out that our characterization of optimality implies various three-fund theorems that were derived in the literature under more restrictive assumptions.

**Definition 4.1 (Location-scale mixture property of \( F_X \))** Let \( Z \) be a random variable taking values in \( \mathbb{R} \) and having zero mean and unit variance. Let \( Y \geq 0 \) a.s. be a positive random variable that is independent of \( Z \). We say that the joint distribution \( F \) of \( X = (X_1, X_2, ..., X_n) \) has the location-scale mixture property associated with the random variables \( Z \) and \( Y \) if for any vector \( a = (a_1, ..., a_n)^T \) it holds that
\[
 a^T X \overset{d}{=} a^T \mu + Ya^T \gamma + \sqrt{Y} \sqrt{a^T \Sigma a} Z, \tag{17}
\]
for some vectors \( \mu \), and \( \gamma \) and for the positive definite symmetric matrix \( \Sigma \), which can be interpreted as parameters.

The family \( G \) of all multivariate distributions that have this location-scale mixture property is then called the location-scale family of distributions associated with \( Z \) and \( Y \).

If the joint distribution \( F_X \) of \( X = (X_1, X_2, ..., X_n) \) is a member of the location scale multivariate family associated with the random variables \( Z \) and \( Y \), then for all \( i \), \( X_i \) belongs to the same location-scale mixture family associated with \( Z \) and \( Y \), i.e., there exist \( \mu_i \in \mathbb{R}, \gamma_i \in \mathbb{R}, \) and \( \sigma_i > 0 \) such that \( X_i = \mu_i + Y \gamma_i + \sqrt{Y} \sigma_i Z \). The same is true for all univariate affine transformations of \( X \).
A prominent example of Definition 4.1 arises when $F_X$ belongs to the so-called multivariate generalized hyperbolic (MGH) family of distributions (Section 3.2 in McNeil, Frey, and Embrechts (2010)), which is a natural extension of the multivariate elliptical family of distributions. In this case, we find that $X = (X_1, ..., X_n) \sim F_X$ can be represented as

$$(X_1, ..., X_n) \overset{d}{=} \mu + Y \gamma + \sqrt{Y} AZ,$$

where $Z$ is a random vector that follows a multivariate normal distribution $\text{MVN}_k(0, I_k)$, $A \in \mathbb{R}^{n \times k}$ is a matrix to be chosen taking into account that $AZ \sim \text{MVN}_n(0, \Sigma)$ with $\Sigma = AA^T$, and the scalar factor in the mixture, $Y$, is a generalized inverse Gaussian distribution with parameters $\lambda$, $\chi$ and $\psi$ (Section 3.2 in McNeil, Frey, and Embrechts (2010)).

### 4.1 Characterization of Optimal Portfolios for FSD-investors

When the joint distribution $F_X$ of the random return vector $X = (X_1, X_2, ..., X_n)$ has the location-scale mixture property (17) associated with the random variables $Z$ and $Y$, then we obtain that the final wealth $W_\omega$ satisfies

$$W_\omega \overset{d}{=} \mu_\omega + Y \gamma_\omega + \sqrt{Y} \sigma_\omega Z$$

and has a distribution depending on the following three parameters

$$
\begin{align*}
\mu_\omega &= W_0(1 + r) + \omega^T(\mu - r 1) \\
\gamma_\omega &= \omega^T \gamma \\
\sigma_\omega &= \sqrt{\omega^T \Sigma \omega}
\end{align*}
$$

(19)

In this regard, we point out that unlike the case of a distribution $F_X$ with a location-scale property, the parameters $\mu_\omega$ and $\sigma_\omega$ can no longer be readily interpreted as the expected value and standard deviation of the terminal wealth $W_\omega$. It is straightforward to show that

$$
\begin{align*}
\mathbb{E}(W_\omega) &= \mu_\omega + \mathbb{E}(Y) \gamma_\omega, \\
\text{var}(W_\omega) &= \gamma_\omega^2 \text{var}(Y) + \sigma_\omega^2 \mathbb{E}(Y), \\
\text{skew}(W_\omega) &= \frac{\gamma_\omega \mathbb{E}[(Y - \mathbb{E}(Y))^3] + 3 \omega \sigma_\omega \text{var}(Y) + \sigma_\omega \mathbb{E}(Y^2) \mathbb{E}(Z^3)}{\text{var}(W_\omega)^{3/2}},
\end{align*}
$$

(20)

where we used the following formula for the skewness, $\text{skew}(W_\omega) = \frac{\mathbb{E}((W_\omega - \mathbb{E}(W_\omega))^3)}{\text{var}(W_\omega)^{3/2}}$.

In order to solve the portfolio optimization problem (2) under the new assumption on the return
distributions, we can reformulate this problem now as

$$\max_{(\mu_\omega, \sigma_\omega, \gamma_\omega) \in B} V(W_\omega)$$

where the set of triplets $B$ is given as

$$B := \left\{ \left( W_0(1 + r) + \omega^T(\mu - r1), \sqrt{\omega^T\Sigma\omega}, \omega^T\gamma \right) \right\}_{\omega \in \mathbb{R}^n}$$

**Definition 4.2 (“Mean-skewness-variance” efficiency frontier)** Consider a portfolio $\omega$ with terminal wealth $W_\omega$ having parameters $\mu_\omega$, $\gamma_\omega$ and $\sigma_\omega$. The portfolio $\omega$ is said to be “mean-skewness-variance” efficient if there is no portfolio that has the same value for $\sigma_\omega$ while having values for $\mu_\omega$ and $\gamma_\omega$ that are at least as big. In particular, the set $B^*$ containing all triplets $(\mu_\omega, \gamma_\omega, \sigma_\omega)$ for which $\omega$ is a “mean-skewness-variance” efficient portfolio is called the “mean-skewness-variance” efficient frontier.

**Proposition 4.3** The “mean-skewness-variance” efficiency frontier $B^*$ is explicitly given as

$$B^* = \left\{ (\mu_\omega, \sigma_\omega, \gamma_\omega) \mid \mu_\omega = W_0(1 + r) + \frac{kg_\omega}{g} + \frac{\sqrt{hg - kg^2} \sqrt{\sigma^2 g - \gamma^2}}{g}, \sigma_\omega > 0, \gamma_\omega \in \left( \frac{k}{\sqrt{g}} \sigma_\omega, \sqrt{g} \sigma_\omega \right) \right\}$$

where $h$, $g$ and $k$ are defined as follows

$$h = (\mu - r1)^T \Sigma^{-1} (\mu - r1) > 0, \quad g = \gamma^T \Sigma^{-1} \gamma > 0, \quad k = (\mu - r1)^T \Sigma^{-1} \gamma.$$
condensed way
\[
\omega_s = \frac{1}{\lambda_3} \Sigma^{-1} (\mu - r \mathbf{1}) - \frac{\lambda_4}{\lambda_3} \Sigma^{-1} \gamma
\]
where \( \lambda_3 \) and \( \lambda_4 \) are such that \( \omega^T_s \gamma = \gamma \) and \( \omega^T_s \Sigma \omega_s = \sigma^2 \). After rewriting this as a quadratic equation in \( \lambda_4 \) we obtain after some calculation that
\[
\lambda_3 = \frac{\sqrt{h g - k^2}}{\sqrt{2 g - \gamma^2}}, \quad \lambda_4 = \frac{k}{g} - \frac{\gamma}{\sqrt{2 g - \gamma^2}}.
\]
The portfolio \( \omega^*_{\sigma, \gamma} \) that solves Problem (25) is thus given as
\[
\omega^*_{\sigma, \gamma} = \frac{\sqrt{\sigma^2 g - \gamma^2}}{\sqrt{h g - k^2}} \Sigma^{-1} (\mu - r \mathbf{1}) - \left( \frac{k \sqrt{\sigma^2 g - \gamma^2}}{g \sqrt{h g - k^2}} - \frac{\gamma}{g} \right) \Sigma^{-1} \gamma.
\] (26)
To obtain the value for \( \mu_\omega \) that corresponds to the portfolio \( \omega^*_{\sigma, \gamma} \), note that
\[
\mu_\omega = W_0 (1 + r) + (\omega^*_{\sigma, \gamma})^T (\mu - r \mathbf{1}).
\]
We then obtain that
\[
(\omega^*_{\sigma, \gamma})^T (\mu - r \mathbf{1}) = \frac{\sqrt{\sigma^2 g - \gamma^2}}{\sqrt{h g - k^2}} (\mu - r \mathbf{1}) \Sigma^{-1} (\mu - r \mathbf{1}) - \left( \frac{k \sqrt{\sigma^2 g - \gamma^2}}{g \sqrt{h g - k^2}} - \frac{\gamma}{g} \right) \gamma^T \Sigma^{-1} (\mu - r \mathbf{1})
\]
\[
= \frac{\sqrt{\sigma^2 g - \gamma^2}}{\sqrt{h g - k^2}} \lambda_3 h_0 - \left( \frac{k \sqrt{\sigma^2 g - \gamma^2}}{g \sqrt{h g - k^2}} - \frac{\gamma}{g} \right) k = \frac{k \gamma}{g} + \frac{\sqrt{\sigma^2 g - \gamma^2}}{g \sqrt{h g - k^2}} \left( \frac{h g - k^2}{g} \right)
\]
\[
= \frac{k \gamma}{g} + \frac{\sqrt{\sigma^2 g - \gamma^2}}{g \sqrt{h g - k^2}} h_0
\]
where we used the fact that \( k = k^T = \gamma^T \Sigma^{-1} (\mu - r \mathbf{1}) \). We denote by \( B^*_1 \) the set of portfolios that have maximum value for \( \mu_\omega \) given \( \sigma_\omega \) and \( \gamma_\omega \). This set is thus explicitly given as
\[
B^*_1 = \left\{ (\mu_\omega, \sigma_\omega, \gamma_\omega) \mid \mu_\omega = W_0 (1 + r) + \frac{k \gamma}{g} + \frac{\sqrt{h g - k^2} \sqrt{\sigma^2 g - \gamma^2}}{g}, \sigma_\omega > 0, \gamma_\omega \in (-\sqrt{g \sigma_\omega}, \sqrt{g \sigma_\omega}) \right\}
\] (27)
Clearly, any “mean-skewness-variance” efficient portfolio belongs to \( B^*_1 \).
Next, we study for a given value for \( \sigma_\omega \), the functional relationship between \( \mu_\omega \) and \( \gamma_\omega \) on the set \( B^*_1 \). We compute the following first and second derivatives
\[
\frac{\partial \mu_\omega}{\partial \gamma_\omega} = \frac{k}{g} - \frac{\sqrt{h g - k^2}}{g} \frac{\gamma_\omega}{\sqrt{\sigma^2 g - \gamma^2}}
\] (28)
\[ \frac{\partial^2 \mu_\omega}{\partial^2 \gamma_\omega} = -\frac{\sqrt{hg - k^2}}{g} \left( \frac{\gamma_\omega^2}{(\sigma_\omega^2 g - \gamma_\omega^2)^2} + \left( \sigma_\omega^2 g - \gamma_\omega^2 \right)^{-\frac{1}{2}} \right). \]  

(29)

Clearly, the second-order derivative is always strictly negative for \( \gamma_\omega \in (\sigma_\omega \sqrt{g}, \sigma_\omega \sqrt{g}) \). Hence, we conclude that for a given value for \( \sigma_\omega \), \( \mu_\omega \) is a strictly concave function of \( \gamma_\omega \), and an easy calculation shows it attains its maximum when \( \gamma_\omega = \sigma_\omega \frac{k}{\sqrt{h}} \). Let now \( B^* \) be the subset of \( B^*_1 \) in which we restrict \( \gamma_\omega \) to the interval \( \left( \frac{k}{\sqrt{h}} \sigma_\omega, \sqrt{g} \sigma_\omega \right) \). Namely,

\[
B^* = \left\{ (\mu_\omega, \sigma_\omega, \gamma_\omega) \mid \begin{array}{l} 
\mu_\omega = W_0 (1 + r) + \frac{k \gamma_\omega}{g} + \frac{\sqrt{hg - k^2} \sigma_\omega^2 g - \gamma_\omega}{g}, \\
\sigma_\omega > 0, \gamma_\omega \in \left( \frac{k}{\sqrt{h}} \sigma_\omega, \sqrt{g} \sigma_\omega \right)
\end{array} \right\}. 
\]

(30)

A graphical illustration of the set \( B^* \) is presented in Figure 1.

Figure 1: Set \( B^* \). Using the market parameters given in Table 1, this graph shows the shape of the “mean-skewness-variance” frontier \( B^* \).

Any “mean-skewness-variance” portfolio must strictly belong to \( B^* \). Indeed it cannot belong to \( B^*_1 \setminus B^* \), as in this case one can always find a portfolio in \( B^* \) with the same \( \sigma_\omega \), and higher values for \( \gamma_\omega \) and \( \mu_\omega \). Conversely, every portfolio in \( B^* \) is “mean-skewness-variance” efficient (note that \( \mu_\omega \) is decreasing in \( \gamma_\omega \)).

Proposition 4.4 (Three-fund theorem for FSD-investors) When \( V(\cdot) \) satisfies Assumption 2.1 and when Problem (21) has a solution \( \omega^* \), then \( (\mu_{\omega^*}, \sigma_{\omega^*}, \gamma_{\omega^*}) \) must be in the set \( B^* \). Further-
more, the optimum $\omega^*$ is given as

$$
\omega^* := \omega^{\sigma_{\omega^*}, \gamma_{\omega^*}} := \frac{\sqrt{\sigma_{\omega^*}^2 g - \gamma_{\omega^*}^2}}{\sqrt{h g - k^2}} \Sigma^{-1} (\mu - r 1) - \left( \frac{k \sqrt{\sigma_{\omega^*}^2 g - \gamma_{\omega^*}^2}}{g \sqrt{h g - k^2}} - \frac{\gamma_{\omega^*}}{g} \right) \Sigma^{-1} \gamma,
$$

where $\sigma_{\omega^*} > 0$ and $\gamma_{\omega^*} \in \left( \frac{k}{\sqrt{\sigma_{\omega^*}}}, \sqrt{g \sigma_{\omega^*}} \right)$ are chosen such that $V(W_{\omega^*})$ is maximum.

**Proof.** Let $\omega^*$ be the solution to Problem (21), i.e., the optimal portfolio for the objective function $V(\cdot)$, that is consistent with FSD. Let $(\mu_{\omega^*}, \sigma_{\omega^*}, \gamma_{\omega^*})$ be the parameters of the terminal wealth determined by $\omega^*$. If $(\mu_{\omega^*}, \sigma_{\omega^*}, \gamma_{\omega^*})$ is not in $B^*_1$, then it is possible to construct a portfolio $\omega'$, that has the same parameters $\sigma_{\omega^*}$ and $\gamma_{\omega^*}$, but a strictly higher parameter $\mu_{\omega'} > \mu_{\omega^*}$. The portfolio $\omega'$ strictly dominates $\omega^*$ in FSD. As $V(\cdot)$ is consistent with FSD, this implies $V(W_{\omega'}) > V(W_{\omega^*})$, which violates the hypothesis of optimality of $\omega^*$. If $\omega^*$ is in $B^*_1 \setminus B^*$ then we can find a portfolio in $B^*$ with the same $\sigma_{\omega^*}$, a higher $\gamma_{\omega}$ and a higher or equal $\mu_{\omega}$, which violates again the hypothesis of optimality of $\omega^*$. Finally, the expression of a “mean-skewness-variance” efficient portfolio $\omega^*$ is given in (26), which ends the proof.

From Proposition 4.4, the optimal portfolio thus consists in investing part of the initial wealth in the risk-free asset, and another part in a linear combination of two funds, $\Sigma^{-1} (\mu - r 1)$ and $\Sigma^{-1} \gamma$. Moreover, the composition of these two funds does not depend on the investor’s preferences. The preferences of the investor translate into optimal weights that are allocated to both funds. Thus, the proposition also implies that solving the high-dimensional optimal portfolio problem in (2) can be reduced to solving a two-dimensional problem in $\mathbb{R}^2$ (i.e., the two optimal weights to be determined). Moreover, Proposition 4.4 shows that a three-fund theorem holds in any setting where the distribution of assets returns exhibits a location-scale mixture property and where preferences are FSD-consistent. Therefore, we recover various results in the literature in which three-fund theorems have been derived under specific assumptions on distributions and preferences. For example, Birge and Chavez-Bedoya (2016) derive a three-fund theorem under expected utility theory with exponential utility and $t$-skewed returns (special case of the MGH multivariate distributions). As yet another example, when preferences are according to cumulative prospect theory and returns follow a $t$-skewed distribution Kwak and Pirvu (2018) show three-fund separation. All these results now immediately follow from Proposition 4.4.
## 4.2 Optimal Portfolios for SSD-investors

Recall that SSD-investors are also FSD-investors and Proposition 3.5 thus also applies to SSD-investors. In this section, we explore whether for SSD-investors a more specific characterization can be derived for their optimal portfolio. In this regard, it can be shown (see e.g., Appendix A) that SSD-investors choose a portfolio that belongs to the set $C^*$ of portfolios that solve the auxiliary problem

$$
\min_{\omega} \sigma_\omega \text{ subject to } \mu_\omega = \mu, \gamma_\omega = \gamma.
$$

(31)

Hence the optimal portfolios for SSD-investors are to be found in the intersection $C^* \cap B^*$, limiting the set of admissible portfolios and perhaps also leading to a more refined characterization for the optimal portfolio under SSD-consistent preferences. However, for every couple $(\mu_\omega, \gamma_\omega)$, the value of $\sigma_\omega$ such that $(\mu_\omega, \sigma_\omega, \gamma_\omega) \in B^*$ is uniquely determined (see Appendix B), and thus we get that $B^*$ is a subset of $C^*$ and hence $B^* = C^* \cap B^*$ and it does not seem useful to further characterize $C^*$.

Nevertheless, the following result shows that under certain market conditions, $B^*$ is actually too broad in that it contains portfolios that cannot be optimal for an SSD-investor.

**Proposition 4.5** If $k + \mathbb{E}(Y)g < 0$, there exist portfolios in $B^*$ having an expected return that is lower than the return given by the risk-free investment. Specifically, If $k + \mathbb{E}(Y)g < 0$ then for $(\mu_\omega, \sigma_\omega, \gamma_\omega) \in B^*$ it holds that

$$
\mathbb{E}(W_\omega) < W_0(1 + r) \iff \gamma_\omega \in \left( \sigma_\omega \sqrt{g} \frac{\sqrt{h g - k^2} \sigma_\omega g - \gamma_\omega^2}{\sqrt{(k + \mathbb{E}(Y)g)^2 + h g - k^2} \sigma_\omega \sqrt{g}} \right).
$$

(32)

**Proof.** Assume $k + \mathbb{E}(Y)g < 0$. Proposition 4.3 shows that in $B^*$ there exists a specific relationship between the parameter $\mu_\omega$ and the parameters $(\sigma_\omega, \gamma_\omega)$. Here we use it to characterize portfolios in $B^*$ with an expected return that is lower than the return given by the risk-free investment.

$$
\mathbb{E}(W_\omega) < W_0(1 + r) \iff \mu_\omega + \mathbb{E}(Y)\gamma_\omega < W_0(1 + r)
$$

$$
\iff W_0(1 + r) + \frac{k \gamma_\omega}{g} + \frac{\sqrt{h g - k^2} \sigma_\omega^2 g - \gamma_\omega^2}{g} + \mathbb{E}(Y)\gamma_\omega < W_0(1 + r)
$$

$$
\iff \frac{k \gamma_\omega}{g} + \frac{\sqrt{h g - k^2} \sigma_\omega^2 g - \gamma_\omega^2}{g} + \mathbb{E}(Y)\gamma_\omega < 0
$$

$$
\iff \gamma_\omega (k + \mathbb{E}(Y)g) < -\sqrt{h g - k^2} \sigma_\omega^2 g - \gamma_\omega^2.
$$

Since the right side of the last inequality is negative, it follows from the assumption $k + \mathbb{E}(Y)g < 0$
that if $\gamma_\omega \leq 0$ then $E(W_\omega) \geq W_0(1 + r)$. Next, we consider the case $\gamma_\omega > 0$,

$$E(W_\omega) < W_0(1 + r) \iff \gamma_\omega > -\frac{\sqrt{hg - k^2} \sigma_\omega^2 g - \gamma_\omega^2}{k + \mathbb{E}(Y)g} > 0$$

$$\iff \gamma_\omega^2 > \frac{(h g - k^2)(\sigma_\omega^2 g - \gamma_\omega^2)}{(k + \mathbb{E}(Y)g)^2}$$

$$\iff \gamma_\omega^2 > \left( \frac{(k + \mathbb{E}(Y)g)^2}{(k + \mathbb{E}(Y)g)^2 + h g - k^2} \right) \sigma_\omega^2 g \left( \frac{h g - k^2}{k + \mathbb{E}(Y)g} \right)$$

$$\iff \gamma_\omega^2 > \frac{h g - k^2}{(k + \mathbb{E}(Y)g)^2 + h g - k^2}.$$ 

Therefore, for the portfolios $W_\omega$ in $B^*$,

$$E(W_\omega) < W_0(1 + r) \iff \gamma_\omega \in \left( \sigma_\omega \sqrt{g} \frac{\sqrt{hg - k^2}}{\sqrt{(k + \mathbb{E}(Y)g)^2 + h g - k^2}}, \sigma_\omega \sqrt{g} \right).$$

5 Application

Bellman (2015) explained that to optimize a $n$-dimensional function on a continuous domain by exhaustively searching a discrete grid (obtained by a crude discretization), one could easily end up with making trillions of evaluations of the function. This is what he called “curse of dimensionality.” Specifically, in the context of portfolio optimization, the initial dimensionality of the problem (i.e., the number $n$ of assets) is typically in the range 30-1000. Hence, if one considers a grid of spacing $1/100$ on the unit cube in 30 dimensions, we already have $100^{30}$ evaluations to make, which is not feasible in practice and out-rules the use of exhaustive enumeration strategies. It is well-known that computational tractability is greatly enhanced if the optimization problem at hand is convex, as in this case local optima are global optima, a feature, which allows local search algorithms to guarantee optimal solutions. However, demonstrating convexity is not always straightforward nor always true or desirable. For instance, if one aims to maximize a lower quantile of terminal wealth then one is using a non-convex objective that is however FSD-consistent.

A main contribution of this paper is to show that under the rather flexible assumption of a multivariate location-scale mixture distribution of the asset returns, any optimization problem that is FSD-consistent can be readily approached using exhaustive search in a two-dimensional grid. Moreover, as the optimization problem is essentially of a two-dimensional nature, explicit solutions for concave objectives might be in reach or at least they can be easily obtained numerically.

19
We illustrate both features by revisiting a portfolio optimization problem that was also considered in Birge and Chavez-Bedoya (2016). Specifically, we show that the framework developed in this paper makes it indeed possible to transform their $n$-dimensional (concave) portfolio optimization problem into a concave two-dimensional problem, which is easier to deal with. Furthermore, we show that an exhaustive search can also provide the solution in a very fast way and is straightforward to implement.

5.1 Concave optimization and exhaustive search

Birge and Chavez-Bedoya (2016) assume that the vector $X$ of asset returns follows a so-called Generalized Hyperbolic Skew-t distribution; that is, the variables $Y$ and $Z$ in the model (17) follow, respectively, an Inverse Gaussian distribution with parameter $\nu > 3$, and a Gaussian distribution. Hence, $W_\omega \sim Skew-t(\nu, \mu_\omega, \sigma_\omega, \gamma_\omega)$ in which $(\mu_\omega, \sigma_\omega, \gamma_\omega)$ are as given in (19). Furthermore, it is assumed that the investor preferences are described by an exponential utility, i.e. $U(x) = -e^{-ax}$, in which $a > 0$ is the risk-aversion coefficient. So either for their assumptions on preferences or on the assets returns distribution, their setting is included in ours. The moment generating function of $W_\omega$ is well-known and given as

$$E(e^{sW_\omega}) = e^{\mu_\omega s \frac{1-\frac{\nu}{2}}{\Gamma(\frac{\nu}{2})} \left( -\nu(\sigma_\omega^2 s^2 + 2s\gamma_\omega) \right)^\frac{\nu}{2} K_{\frac{\nu}{2}} \left( \sqrt{-\nu(\sigma_\omega^2 s^2 + 2s\gamma_\omega)} \right) },$$

(33)

which only exists for $s \in \mathbb{R}$ such that $2s\gamma_\omega + \sigma_\omega^2 s^2 \leq 0$. The expected exponential utility $E(U(W_\omega))$ is thus given as

$$E(U(W_\omega)) = \frac{-e^{-\nu a_2 s \frac{1-\frac{\nu}{2}}{\Gamma(\frac{\nu}{2})} \left( -\nu(\sigma_\omega^2 a_2^2 - 2a_\gamma_\omega) \right)^\frac{\nu}{2} K_{\frac{\nu}{2}} \left( \sqrt{-\nu(\sigma_\omega^2 a_2^2 - 2a_\gamma_\omega)} \right) }}{} ,$$

(34)

in which $a$ is such that $\sigma_\omega^2 a_2^2 - 2a_\gamma_\omega \leq 0$.

To find the optimal portfolio Birge and Chavez-Bedoya (2016) show the concavity of the objective function w.r.t the weights $\omega_i$. Sometimes, explicit solutions for the optimal vector of weights $\omega^*$ can be obtained. In contrast, we directly find the vector $(\mu_\omega^*, \sigma_\omega^*, \gamma_\omega^*)$ in the set $B^*$ that yields maximum expected utility and from this we infer the optimal vector of weights $\omega^*$.

For the ease of exposition we further omit the subscripts $\omega$ and $\omega^*$. Using the functional relationship between $\mu$ and $(\sigma, \gamma)$ in $B^*$ (see equation (23)), we obtain that the objective function
(34) can be written as a function of \((\sigma, \gamma)\) only. We denote it by \(f(\sigma, \gamma)\), where

\[
f(\sigma, \gamma) = -e^{-a \left( W_0 (1+r) + \frac{k_2}{g} + \frac{\sqrt{\gamma}}{g} \sqrt{\sigma^2 + \gamma^2} \right)} \frac{2^{1-\frac{\nu}{4}}}{\Gamma(\frac{\nu}{2})} (-\nu(\sigma^2 a^2 - 2a\gamma))^\frac{\nu}{2} K_{\frac{\nu}{2}} \left( \sqrt{-\nu(\sigma^2 a^2 - 2a\gamma)} \right) .
\]

(35)

In Appendix C, we show that this problem can be alternatively formulated as

\[
\max_{(\beta, \gamma) \in D} q(\beta, \gamma),
\]

(36)

where \(q(\beta, \gamma)\) is two-dimensional concave function and the domain \(D\) is given as

\[
D = \left\{ (\beta, \gamma) \mid \gamma^2 - 2g^2 \gamma + \beta \leq 0, \ \beta \geq 0 \right\}.
\]

The definition and properties of \(q(\beta, \gamma)\) are illustrated in Appendix C. Once the optimal values \(\beta^*\) and \(\gamma^*\) are found, we obtain the optimal value \(\sigma^* = \sqrt{\frac{\beta^* + \gamma^*}{2}}\). The optimal values for \(\mu^*\) and the weights \(\omega_i^*\) as functions of \(\sigma^*\) and \(\gamma^*\) follow from equation (30) and Proposition 4.4, respectively.

Hereafter, we illustrate that the solution obtained using the concave optimization corresponds closely to the one that we would obtain using exhaustive search.

Specifically, we consider two risky assets, \(X = (X_1, X_2) \sim Skew-t(\nu, \mu, \Sigma, \gamma)\), with parameters as described in Table 1. The initial wealth is \(W_0 = 1\) and risk-free rate is set \(r = 0.02\). For the risk-aversion parameter \(a\), we consider \(a = 2\).

Table 1: Market parameters for the two-dimensional market.

<table>
<thead>
<tr>
<th></th>
<th>(\mu)</th>
<th>(\gamma)</th>
<th>(\sigma)</th>
<th>(\rho)</th>
<th>(\nu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>0.05</td>
<td>-0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>5</td>
</tr>
<tr>
<td>(X_2)</td>
<td>0.08</td>
<td>0.7</td>
<td>0.5</td>
<td>0.3</td>
<td>5</td>
</tr>
</tbody>
</table>

In order to find the maximum of the objective function \(f(\sigma, \gamma)\) given in equation (35) using an exhaustive search approach, we first describe the set of feasible solutions.

From equation (35), a couple \((\sigma, \gamma)\) that belongs to the domain of \(f(\sigma, \gamma)\) must satisfy \(\gamma \geq \frac{1}{2} \sigma^2 \frac{a}{g}\). Furthermore, from the definition of \(B^*\) in (23), we have the additional constraint that \(\gamma \in \left( \sigma k \sqrt{h}, \sigma \sqrt{g} \right)\). Putting these conditions together, we obtain the set of feasible solutions \(S\):

\[
S = \left\{ (\sigma, \gamma) \mid \sigma \in \left( 0, \frac{2\sqrt{g}}{a} \right), \ \gamma \in \left( \max \left\{ \frac{2\sigma^2}{a} ; \frac{\sigma}{\sqrt{h}} \right\}, \sigma \sqrt{g} \right) \right\}
\]

(37)

Observe that \(S\) is a bounded convex set in \(\mathbb{R}^2\), which significantly eases the implementation of an
exhaustive search.

In Table 2 we display the optimal values for $\mu^*$, $\sigma^*$ and $\gamma^*$ obtained using exhaustive search and obtained using concave optimization. A two-dimensional grid is considered, with 2,000 values for $\sigma$ and for each $\sigma$ we obtain 2,000 values for $\gamma$, i.e., a total number of 4,000,000 points to be evaluated.

<table>
<thead>
<tr>
<th></th>
<th>$\mu_\omega$</th>
<th>$\gamma_\omega$</th>
<th>$\sigma_\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concave optimization</td>
<td>1.0487</td>
<td>2.1159</td>
<td>1.0437</td>
</tr>
<tr>
<td>Exhaustive search</td>
<td>1.0518</td>
<td>2.1114</td>
<td>1.0437</td>
</tr>
</tbody>
</table>

This example shows how the implementation of exhaustive search on a two-dimensional grid merely requires the derivation of the set of feasible points. Therefore, this approach can in principle be applied to any optimization problem within our framework, and this turns out to be particularly useful when the objective function is not concave.

5.2 Mean-variance approximations

Levy and Markowitz (1979) and Markowitz (2014) provide some theoretical support for the observation that the optimal portfolio of an investor who maximizes expected utility can always be approximated by a mean-variance efficient portfolio; see also Birge and Chavez-Bedoya (2016) for some numerical evidence. Note, however, that under the assumption of a location-scale mixture for the assets returns, the optimal portfolio is a combination of three funds and thus not two, like in the case of mean-variance efficient portfolios. Hence, it is not obvious that one always find a mean-variance efficient portfolio (two funds) that is very close to an EUT-optimal portfolio (three funds).

Here, we show that under the assumption of a location-scale mixture for the asset returns there exist portfolios that are optimal for an exponential utility investor and that cannot be well approximated by a mean-variance efficient portfolio. To measure the distance $d(\omega^1, \omega^2)$ between two portfolios $\omega^1$ and $\omega^2$, we use the Euclidean distance, i.e.,

\[
d(\omega^1, \omega^2) = \sqrt{\sum_{i=1}^{n} (\omega^1_i - \omega^2_i)^2}
\]  \hspace{1cm} (38)

In Appendix D, we explain how for a given portfolio $\omega^*_{EU}(a)$ that is optimal for an expected utility maximizing investor with risk aversion coefficient $a$, one can determine the mean-variance
efficient portfolio $\omega_{MV}^*(a)$ that is closest, i.e., such that $d(\omega_{EU}^*(a), \omega_{MV}^*(a))$ becomes minimum among all mean-variance efficient portfolios $\omega_{MV}$.

Figure 2 displays, the behavior of this minimum distance, first, for a range of values for the risk aversion parameter $a$, $d(\omega_{EU}^*(a), \omega_{MV}^*(a))^2$ in Panel A and then as a function of the parameter $\gamma_2$ of the second risky asset in Panel B in the case of $a = 1$. Unless otherwise specified, all market parameters are set as in Table 1.

Both panels in Figure 2 show that the minimum distance between the exponential utility maximizing portfolio and a mean-variance efficient portfolio can be significant. This situation occurs when the risk-aversion parameter $a$ approaches 0 or when the parameter $\gamma_2$ of the second risk asset is big enough.

![Figure 2: Minimum distance between the optimal exponential utility maximizing portfolio and the closest mean-variance efficient portfolio as a function of the risk-aversion parameter $a$ in Panel A and as a function of $\gamma_2$ in Panel B.](image)

In Figures 3 and 4 we compare in more detail the composition of the EUT-optimal portfolio with the composition of the closest mean-variance efficient portfolio. On the one hand, the portfolios exhibit similarities in that the decision of going long or short in an asset appears to be the same in both cases. On the other hand, the amounts invested in the various assets are clearly different and they are particularly different when the parameter $a$ approaches zero or when $\gamma_2$ is increasing. This feature is consistent with the observations made in Figure 2. Note also that in all considered
cases the exponential utility maximizer appears to invest more in the risk-free asset than the corresponding mean-variance maximizer.

Figure 3: Comparing the portfolios as the parameter \( a \) takes three values \( a = 0.3, a = 1 \) and \( a = 5 \). Each optimal exponential utility portfolio is compared with the mean-variance portfolio that minimizes the Euclidean distance (38). The market parameters are given in Table 1. The respective weights in the risk-free asset and in the two risky assets are displayed as bars.

All in all, the counterexamples provided in this numerical study suggest that not all optimal portfolios for expected utility investors can be efficiently approximated by a mean-variance portfolio.

Figure 4: Comparing the portfolios as \( \gamma_2 \) takes three values \( \gamma_2 = 0.3, \gamma_2 = 0.6 \) and \( \gamma_2 = 1 \). Each optimal exponential utility portfolio is compared with the mean-variance portfolio that minimizes the Euclidean distance (38). The other market parameters are given in Table 1 and the risk aversion parameter is set to \( a = 5 \). The respective weights in the risk-free asset and in the two risky assets are displayed as bars.
6 Final Remarks

Under very weak conditions on the investor’s preferences we provide conditions on the multivariate
distribution of asset returns that lead to two-fund or three-fund separation. Specifically, a location-
scale of the multivariate distribution (elliptical distributions) leads to a two-fund separation for
all investors with law-invariant and increasing preferences whereas a weaker conditional location-
scale property implies three-fund separation. The latter condition is for instance satisfied by a
multivariate hyperbolic distribution and is known to be fairly realistic for modeling real-world
asset returns. Thus the assumptions made on the assets’ returns distribution are the key element
to switch from a two-fund theorem to a three-fund theorem. The specificity of the objective
function, which reflects the specific assumptions on the investor’s preferences, is only used to select
the optimal weights that are allocated to each of the two, respectively three funds.

Using this characterization of the optimal portfolio, it is then possible to significantly reduce
the complexity of finding an optimal portfolio, as only a two-dimensional optimization in the case
of two-fund separation (respectively a three-dimensional optimization in the case of three-fund
separation) is needed even in a market setting in which there are thousands of assets.

Using our theoretical approach and general characterization of optimal portfolio under very
general preferences, we are able to show that two-fund and three-fund theorems that have appeared
in the literature and that were derived under specific assumptions on the investor’s preferences
and on the market setting are special cases of our general characterization results. Finally, we
provide evidence that the the optimal portfolio of an expected utility maximizer cannot always be
approximated by a mean-variance efficient portfolio even though this was claimed in the literature
(e.g., Levy and Markowitz (1979) and Markowitz (2014)).
References


Management Science, 52(9), 1409–1423.

press.

of Mathematical Economics, 59, 10–23.

BERTSIMAS, D., G. J. LAUPRETE, AND A. SAMAROV (2004): “Shortfall as a risk measure: prop-
erties, optimization and applications,” Journal of Economic Dynamics and control, 28(7), 1353–
1381.

BIRGE, J. R., AND L. CHAVEZ-BEDOYA (2016): “Portfolio optimization under a generalized hy-
perbolic skewed t distribution and exponential utility,” Quantitative Finance, 16(7), 1019–1036.

decision, and measurement, pp. 73–100.

of stochastic dominance and cumulative independence,” Journal of Risk and Uncertainty, 17, 
49–78.


CESARI, R., AND D. CREMONINI (2003): “Benchmarking, portfolio insurance and technical anal-
ysis: a Monte Carlo comparison of dynamic strategies of asset allocation,” Journal of Economic 
Dynamics and Control, 27(6), 987–1011.

CHAMBERLAIN, G. (1983): “A characterization of the distributions that imply mean-variance 

two-fund separation,” Finance Research Letters, 8(2), 52–58.

281–299.


Appendix : Supplementary Material

A  SSD-investors optimal portfolios are in $C^*$

Recall that any objective function $V(\cdot)$ that satisfies Assumption 2.2 is also consistent with convex order, i.e., given two portfolios $W_1$ and $W_2$ such that $W_1 \prec_{cx} W_2$, then $V(W_1) \geq V(W_2)$. Hence, to prove that all SSD-optimal portfolios are in $C^*$ it is sufficient to prove that given two portfolios with location-scale mixture property, same parameter $\mu_\omega$ and $\gamma_\omega$ but different $\sigma_\omega$, then the one with a lower $\sigma_\omega$ is also lower in convex order, as stated in the next proposition.

**Proposition A.1** Let $Z$ be a random variable with $E(Z) = 0$ and $\text{std}(Z) = 1$, and let $Y \geq 0$ a.s. be a random variable that is independent of $Z$. Let $W_1$ and $W_2$ be two random variables such that $W_1 \overset{d}{=} \mu + Y\gamma + \sqrt{Y}\sigma_1 Z$ and $W_2 \overset{d}{=} \mu + Y\gamma + \sqrt{Y}\sigma_2 Z$ with $\mu \in \mathbb{R}$, $\gamma \in \mathbb{R}$ and $0 < \sigma_1 < \sigma_2$. Then, for all convex functions $f$,

$$E(f(W_1)) \leq E(f(W_2))$$  \hspace{1cm} (39)

**Proof.** First, observe that $E(W_1) = E(W_2)$. For $i = 1, 2$ and for all $y \geq 0$, let us denote with $W_{iy}$ the random variable $W_i | Y = y$. The expected value and the standard deviation of $W_{iy}$, will be denoted with $m_{iy} = \mu + y\gamma$ and $\delta_{iy} = \sqrt{y}\sigma_i$, respectively. Under our hypothesis on the distribution of $W_1$ and $W_2$, we can write

$$W_{iy} \overset{d}{=} \mu + y\gamma + \sqrt{y}\sigma_i Z$$

$$\overset{d}{=} m_{iy} + \delta_{iy} Z$$

It is clear that for all convex functions $f$,

$$E(f(W_1)) = E(E(f(W_1|Y = y))) \leq E(E(f(W_2|Y = y))) = E(f(W_2))$$

\[\blacksquare\]

B  Uniqueness of $\sigma_\omega^2$ in $B^*$

**Proof.** From the definition of the set $B^*$ (Proposition 4.3) we can deduce that for any couple $(\mu_\omega, \gamma_\omega)$, there exists a unique value of $\sigma_\omega^2$ such that $(\mu_\omega, \sigma_\omega, \gamma_\omega) \in B^*$. To see this point, we can simply invert the functional relationship given in equation (23).

$$(\mu_\omega, \sigma_\omega, \gamma_\omega) \in B^* \implies \mu_\omega = W_0(1 + r) + \frac{k\gamma_\omega}{g} + \frac{\sqrt{hg - k^2} \sqrt{\sigma_\omega^2 g - \gamma_\omega^2}}{g}$$

$$\implies \frac{(\mu_\omega - W_0(1 + r)) g - k\gamma_\omega}{\sqrt{hg - k^2}} = \sqrt{\sigma_\omega^2 g - \gamma_\omega^2}$$

28
Thus

\[(\mu, \sigma, \gamma) \in B^* \implies \frac{(\mu - W_0(1 + r)g - k\gamma)}{hg - k^2} = \sigma^2 g - \gamma^2 \]

\[\implies \frac{(\mu - W_0(1 + r))\sqrt{g} - \frac{k\gamma}{\sqrt{g}}}{hg - k^2} = \frac{\sigma^2}{g} \]

\[\implies \frac{(W_0(1 + r) - \mu)\sqrt{g} + \frac{k\gamma}{\sqrt{g}}}{hg - k^2} + \frac{\gamma^2}{g} = \sigma^2.\]

\[\blacksquare\]

C Proof of Concavity of the problem (36)

**Proof.** Considering \(f(\sigma, \gamma)\) in (35), we switch the parametrization from \((\sigma, \gamma)\) to \((\beta, \gamma)\), with \(\beta = \sigma^2 g - \gamma^2\) and obtain the following portfolio optimization problem:

\[
\max_{(\beta, \gamma) \in D} f(\beta, \gamma),
\]

where

\[
f(\beta, \gamma) = e^{-a(W_0(1+r) + \frac{k\gamma}{g} + \sqrt{hg - k^2\sqrt{\beta}})} \left( -\nu \left( \frac{\beta + \gamma^2 a^2 - 2a\gamma}{g} \right) \right)^\frac{\nu}{2} K_{\frac{\nu}{2}} \left( \sqrt{-\nu \left( \frac{\beta + \gamma^2 a^2 - 2a\gamma}{g^2 - 2a\gamma} \right)} \right),
\]

and where the domain \(D\) is given as

\[
D = \left\{ (\beta, \gamma) \mid \gamma^2 - 2\frac{g}{a} \gamma + \beta \leq 0, \ \beta \geq 0 \right\}.
\]

To find a solution to problem (40), we will in a first step rewrite \(f(\beta, \gamma)\) as an easier-to-deal-with function with the same maximum. To this end, consider the function

\[
\theta(y) = -\frac{\ln(-y) + \ln(2^{1-\frac{y}{2}}) - \ln(\Gamma(\frac{\nu}{2}))}{a} - W_0(1 + r).
\]

Since \(\theta(y)\) is increasing for \(y < 0\), \(q(\beta, \gamma) := \theta(f(\beta, \gamma))\) has the same maximum and domain as \(f(\beta, \gamma)\). Furthermore, \(q(\beta, \gamma)\) writes as

\[
q(\beta, \gamma) = \frac{k\gamma}{g} + \frac{\sqrt{hg - k^2\sqrt{\beta}}}{g} - \frac{g(\beta, \gamma)}{a},
\]

in which

\[
g(\beta, \gamma) = \ln \left( \left( -\nu \left( \frac{\beta + \gamma^2 a^2 - 2a\gamma}{g} \right) \right)^\frac{\nu}{2} K_{\frac{\nu}{2}} \left( \sqrt{-\nu \left( \frac{\beta + \gamma^2 a^2 - 2a\gamma}{g^2 - 2a\gamma} \right)} \right) \right).
\]

In a second step we prove that \(q(\beta, \gamma)\) is concave w.r.t. the variables \((\beta, \gamma)\). We start by proving
that $-g(\beta, \gamma)$ is concave. In this regard, note that $-g(\beta, \gamma) = r(A(\beta, \gamma))$, where

$$r(y) = -\ln \left(y^{\frac{\gamma}{2}} K_{\frac{\epsilon}{\gamma}}(\sqrt{y})\right), \quad A(\beta, \gamma) = -\nu \left(\frac{\beta + \gamma^2 a^2 - 2a\gamma}{g}\right).$$

In their Appendix C1, Birge and Chavez-Bedoya (2016) prove that $r(y)$ is an increasing and concave function. Furthermore, as $A(\beta, \gamma)$ is concave as well, $-g(\beta, \gamma)$ is concave. Finally, $q(\beta, \gamma)$ can be seen as a sum of concave functions, so it is concave itself.

\section{Minimum distance to mean-variance efficient portfolios}

**Proof.** In our numerical study, each portfolio $\omega_{EU}^*$ optimal for an exponential utility investor is compared with the mean-variance portfolio that minimizes the Euclidean distance (38) with $\omega_{EU}^*$. In this appendix, we illustrate how, given a generic $n$-dimensional vector $x$ interpretable as a benchmark portfolio, it is possible to determine the mean-variance efficient portfolio that has minimum Euclidean distance with $x$.

Considering $n$ risky assets whose returns have a positive definite covariance matrix $\Delta$ and an expected value $m$, all mean-variance efficient portfolios can be written as

$$\omega_\delta^* = \delta \omega_1^* \frac{1}{\sqrt{n}} (m - r1).$$

Therefore, $\omega_\delta^* = \delta \omega_1^*, \forall \delta > 0$. Our aim is to solve the following problem:

$$\min_\delta \sum_{i=1}^{n} (x_i - \delta \omega_1^*)^2$$

Let $d(\delta)$ be the function we aim to minimize, i.e. $d(\delta) = \sum_{i=1}^{n} (x_i - \delta \omega_1^*)^2$. The first and the second-order derivatives of $d(\delta)$ are

$$d'(\delta) = \sum_{i=1}^{n} -2 \omega_1^* (x_i - \delta \omega_1^*), \quad d''(\delta) = 2 \sum_{i=1}^{n} \omega_1^{*2}$$

Assuming $\omega_1^* \neq 0$, $d(\delta)$ is strictly convex. To find the minimum we look for the value of $\delta$ such that $d''(\delta) = 0$, namely

$$d'(\delta) = 0 \iff \sum_{i=1}^{n} -2 \omega_1^* (x_i - \delta \omega_1^*) = 0 \iff \sum_{i=1}^{n} \omega_1^{*2} \delta = \sum_{i=1}^{n} \omega_1^* x_i \iff \delta = \frac{\sum_{i=1}^{n} \omega_1^* x_i}{\sum_{i=1}^{n} \omega_1^{*2}}$$

Hence, $\sum_{i=1}^{n} \omega_1^* x_i / \sum_{i=1}^{n} \omega_1^{*2} \omega_1^*$ is the mean-variance efficient portfolio with minimum Euclidean distance with the vector $x$.