XAIA INVESTMENT

THE BASICS OF QUANTITATIVE PORTFOLIO SELECTION

Jan-Frederik Mai
XAIA Investment GmbH
Sonnenstraße 19, 80331 München, Germany
jan-frederik.mai@xaia.com

Date: January 29, 2020

Abstract The seminal work of Harry Markowitz from the 1950s is the first scientific approach towards portfolio selection based on the idea of diversification, constituting a quantitative setup whose core ideas are still prominent in today's financial industry. The content of the present article consists of three parts. First, the Markowitz theory is summarized, with an emphasis on its relation with the concept of the Sharpe ratio, and also recalling its relation with the Black–Scholes model via power utility maximization. Second, its limitations and potential generalizations are discussed. Third, it is demonstrated in the particular case of our fund XAIA Credit Curve Carry how the related concept of Sharpe ratio maximization can assist with managing daily portfolio adjustments.

1 Introduction

Regarding notations, for an arbitrary natural number n the elements of $\mathbb{R}^n$ are always interpreted as column vectors, and for $\mathbf{x} \in \mathbb{R}^n$ we write $\mathbf{x}^T$ for the associated row vector, i.e. the transpose of $\mathbf{x}$. This explicit distinction between row and column vectors sometimes leads to seemingly excessive uses of the transpose-symbol, but is important because of the matrix algebra involved in Markowitz theory. We consider an investment universe of $d \in \mathbb{N}$ risky assets, and write $\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ for the vector of invested capitals corresponding to these assets in our portfolio. Furthermore, by $\mathbf{1}$ and $\mathbf{0}$ we denote the elements in $\mathbb{R}^d$ with all entries equal to one and zero, respectively. It is furthermore assumed that there exists no portfolio in $\mathbb{R}^d \setminus \{\mathbf{0}\}$ that is risk-free, which can intuitively be seen as a no arbitrage hypothesis. The existence of a unique risk-free asset is included in the analysis separately by enlarging the space of possible (risky) portfolios $\mathbb{R}^d$ by an additional dimension to $\mathbb{R} \times \mathbb{R}^d$. We represent the first component of an element $(x_0, \mathbf{x})$ in this set as the cash amount invested into the risk-free asset. For a non-trivial[1] portfolio $(x_0, \mathbf{x})$ the value $1 - x_0 / \sum_{k=0}^{d} |x_k|$ is called the *investment ratio*. Notice that the investment ratio is by definition non-negative, and equals zero if and only if $\mathbf{x} = \mathbf{0}$. If the investment ratio equals one (i.e. $x_0 = 0$), we say the portfolio is *fully invested*. If it is larger than one (i.e. $x_0 < 0$), we say the portfolio is *leveraged*.

In this article by *quantitative portfolio selection* we mean a two-step portfolio selection approach of the following kind.

(1) For each potential portfolio $(x_0, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^d$ a finite number of key figures forms the basis for portfolio selection. Examples of such key figures are expected return measurements,

[1] By "non-trivial" we mean any portfolio $(x_0, \mathbf{x})$ except for the case $x_0 = x_1 = \ldots = x_d = 0$.

default likelihoods, risk measures like Value-at-Risk or standard deviation, skewness and kurtosis, the investment ratio, etc..

(2) Portfolio selection tasks are exclusively based on an algorithm that uses only the key figures from Step 1 as input. A typical example is that one defines a preference function $p(.)$ from the set of the key figures to $\mathbb{R}$ and prefers portfolio $(x_0, \mathbf{x})$ to $(y_0, \mathbf{y})$ if

$$p\big(\text{key figures of } (x_0, \mathbf{x})\big) \geq p\big(\text{key figures of } (y_0, \mathbf{y})\big).$$

Maximization of $p(.)$ is then a natural, generic mathematical wrapping to organize portfolio selection.

The seminal work of Harry Markowitz, cf. Markowitz (1952, 1959), is one particular portfolio selection approach that falls within this category, and is recapped in Section 2, where an important connection between Markowitz-optimality and the concept of a Sharpe ratio is particularly highlighted[2]. Section 3 discusses shortcomings and potential generalizations of the Markowitz paradigm, and Section 4 presents a concrete application of the Markowitz technique to daily portfolio rebalancing tasks in the particular case of our fund XAIA Credit Curve Carry. Section 5 finally concludes.

2 Markowitz theory

The core idea in the seminal work of Harry Markowitz, cf. Markowitz (1952, 1959), is to describe each potential asset in the investment universe by just two key figures, one representing an expected return measurement, henceforth denoted by a function $\mu : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$, and the other representing a risk measurement, henceforth denoted by a function $\sigma : \mathbb{R} \times \mathbb{R}^d \to [0, \infty)$. The expected return measurement of a specific portfolio $(x_0, \mathbf{x})$ is denoted by $\mu(x_0, \mathbf{x})$, and the risk measurement by $\sigma(x_0, \mathbf{x})$. It is assumed that $\sigma(x_0, \mathbf{x}) > 0$ whenever $\mathbf{x} \neq \mathbf{0}$ and $\sigma(x_0, \mathbf{0}) = 0$ for arbitrary $x_0 \in \mathbb{R}$. This means that any portfolio with positive investment ratio is risky, and one with zero investment ratio is risk-free. This two-dimensional viewpoint naturally defines a partial order on the set of all portfolios, i.e. on $\mathbb{R} \times \mathbb{R}^d$, namely in that a rational investor is assumed to prefer $(x_0, \mathbf{x})$ to $(y_0, \mathbf{y})$ if the expected return of $(x_0, \mathbf{x})$ is greater or equal to the one of $(y_0, \mathbf{y})$ while the risk of $(x_0, \mathbf{x})$ is smaller or equal than the one of $(y_0, \mathbf{y})$, i.e.

$$\begin{aligned} &(x_0, \mathbf{x}) \geq (y_0, \mathbf{y}) \\ &\quad \Leftrightarrow \ \mu(x_0, \mathbf{x}) \geq \mu(y_0, \mathbf{y}) \text{ and } \sigma(x_0, \mathbf{x}) \leq \sigma(y_0, \mathbf{y}) \\ &\quad \Leftrightarrow \ \text{the point } \big(\sigma(x_0, \mathbf{x}), \mu(x_0, \mathbf{x})\big) \text{ lies northwest of} \\ &\qquad\quad \text{the point } \big(\sigma(y_0, \mathbf{y}), \mu(y_0, \mathbf{y})\big) \text{ in the } (\sigma, \mu)\text{-plane.} \end{aligned}$$

Basing portfolio selection on this paradigm corresponds to a search for portfolios $(x_0, \mathbf{x})$ that are maximal with respect to this partial

[2] The interested reader is referred to the standard textbook Elton, Gruber (1995) for more elaborate background on Markowitz theory and its applications.

order on a subset $D \subset \mathbb{R} \times \mathbb{R}^d$. Typical subsets of interest are

$$D(N) := \big\{(0, \mathbf{x}) \,:\, \mathbf{1}^T \mathbf{x} = N\big\},$$
$$D_+(N) := \{(0, \mathbf{x}) \in D(N) \,:\, x_1, \dots, x_d \geq 0\},$$

which intuitively represent fully invested portfolios whose total invested capital[3] is N, with or without the possibility for asset short-selling. Thanks to the two-dimensional risk-return perspective this optimization problem can be solved graphically by finding the portfolios that are "most northwest" in the (σ, μ)-plane. Under the assumption that $\mu(.)$ and $\sigma(.)$ are continuous, the projection of a certain set $D \subset \mathbb{R} \times \mathbb{R}^d$ of considered portfolios is connected whenever D itself is connected in $\mathbb{R} \times \mathbb{R}^d$.
The traditional Markowitz setup defines the key figures μ, σ as

$$\mu(x_0, \mathbf{x}) := r\, x_0 + \boldsymbol{\mu}^T \mathbf{x}, \quad \sigma(x_0, \mathbf{x}) := \sqrt{\mathbf{x}^T \Sigma\, \mathbf{x}}, \qquad (1)$$

with a risk-free rate $r \in \mathbb{R}$, a vector $\boldsymbol{\mu} \in \mathbb{R}^d$ of expected returns for the risky assets, and a symmetric and positive definite matrix $\Sigma \in \mathbb{R}^{d \times d}$. This definition is backed by any model for the unknown future log-returns of the d risky assets over the next year as random variables $R_1, \dots, R_d$. Denoting by $\boldsymbol{\mu}$ the mean vector and by Σ the covariance matrix of the random vector $\mathbf{R} := (R_1, \dots, R_d)^T$, the definition (1) relies on the assumption that the unknown future log-return of the fully invested portfolio $(0, \mathbf{x})$ equals the linear combination $\mathbf{x}^T \mathbf{R}$ of the single risky assets' log-returns. Definition (1) corresponds to considering mean and standard deviation of the portfolio log-return $\mathbf{x}^T \mathbf{R}$ as expected return and risk measurements. The parameters $\boldsymbol{\mu}$ and Σ are typically retrieved from historical data and/or expert opinions, see also paragraph 3.4 below.
Subsection 2.1 summarizes the major findings within the Markowitz setting, in particular with regards to portfolio optimality. Subsection 2.2 further highlights the close connection of Markowitz optimality with the concept of a Sharpe ratio. Subsection 2.3 briefly recalls the seminal work of R.C. Merton, who shows that Markowitz-optimal portfolios maximize the power utility function in a dynamic multivariate Black–Scholes model.
To abbreviate notation we henceforth denote by

$$[\mathbf{x}, \mathbf{y}] := \mathbf{x}^T \Sigma^{-1} \mathbf{y}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,$$

the scalar product induced by the symmetric and positive definite matrix Σ^{-1}, because it appears several times in what follows.

2.1 The major statements

The first main target in Markowitz theory is to find optimal portfolios within the set $D(1)$ of fully invested portfolios, and to describe their location in the (σ, μ)-plane. These findings are gathered in the following theorem, and are visualized in Figure 1.

Theorem 2.1 (Optimal portfolios in $D(1)$)
We assume $\boldsymbol{\mu} \notin \{\mathbf{0}, \mathbf{1}\}$, since these two cases of less interest are treated separately in Remark 2.2 below.

[3] The total invested capital is often normalized to one, i.e. $N = 1$.

(a) For each given expected return $c \in \mathbb{R}$ the function $\mathbf{x} \mapsto \sqrt{\mathbf{x}^T \Sigma \mathbf{x}}$ takes a unique minimum $s(c)$ on the set

$$D(1, c) := \{\mathbf{x} \in \mathbb{R}^d \,:\, (0, \mathbf{x}) \in D(1),\ \boldsymbol{\mu}^T \mathbf{x} = c\}$$

of fully invested portfolios with expected return equal to c.

(b) The minimum in part (a) is given by

$$s(c) = \sqrt{\frac{c^2 [\mathbf{1}, \mathbf{1}] - 2c [\mathbf{1}, \boldsymbol{\mu}] + [\boldsymbol{\mu}, \boldsymbol{\mu}]}{[\boldsymbol{\mu}, \boldsymbol{\mu}] [\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2}},$$

and is attained at the portfolio

$$\mathbf{x}(c) = \frac{c [\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]}{[\boldsymbol{\mu}, \boldsymbol{\mu}] [\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2} \Sigma^{-1} \boldsymbol{\mu} + \frac{[\boldsymbol{\mu}, \boldsymbol{\mu}] - c [\mathbf{1}, \boldsymbol{\mu}]}{[\boldsymbol{\mu}, \boldsymbol{\mu}] [\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2} \Sigma^{-1} \mathbf{1}.$$

(c) The minimal value of $s(c)$ is attained for $c = [\mathbf{1}, \boldsymbol{\mu}]/[\mathbf{1}, \mathbf{1}]$, is given by $s(c) = 1/\sqrt{[\mathbf{1}, \mathbf{1}]}$, and the respective portfolio

$$\mathbf{x}_m := \mathbf{x}\big([\mathbf{1}, \boldsymbol{\mu}]/[\mathbf{1}, \mathbf{1}]\big) = \frac{\Sigma^{-1} \mathbf{1}}{[\mathbf{1}, \mathbf{1}]}$$

is called the *minimum variance portfolio*.

(d) For each attainable standard deviation $s \geq 1/\sqrt{[\mathbf{1}, \mathbf{1}]}$ there is an interval $[c_-(s), c_+(s)]$ of expected returns that can be attained by portfolios in the set

$$\{\mathbf{x} \in \mathbb{R}^d \,:\, (0, \mathbf{x}) \in D(1),\ \sqrt{\mathbf{x}^T \Sigma \mathbf{x}} = s\}$$

of fully invested portfolios with standard deviation s. The endpoints of these intervals describe a hyperbola given by

$$c_\pm(s) = \frac{[\mathbf{1}, \boldsymbol{\mu}]}{[\mathbf{1}, \mathbf{1}]} \pm \frac{\sqrt{[\boldsymbol{\mu}, \boldsymbol{\mu}] [\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2}}{[\mathbf{1}, \mathbf{1}]} \sqrt{s^2 [\mathbf{1}, \mathbf{1}] - 1}.$$

Proof
See Appendix A. □

The upper branch $c_+(s)$ of the hyperbola in part (d) is called *efficient frontier*, because it represents the “most northwest” portfolios in the (σ, μ)-plane, i.e. the ones that are maximal with respect to the aforementioned partial order.

Remark 2.2 (Non-interesting cases)
In the case $\boldsymbol{\mu} = \mathbf{0}$ the minimum variance portfolio $\mathbf{x}_m$ is clearly the unique optimal portfolio with respect to the Markowitz partial order, and the (σ, μ)-projection of the set $D(1)$ boils down to a part of the σ-axis. In the case $\boldsymbol{\mu} = \mathbf{1}$, every fully invested portfolio $(0, \mathbf{x}) \in D(1)$ satisfies $\boldsymbol{\mu}^T \mathbf{x} = 1$, so that $D(1, c) = \emptyset$ for $c \neq 1$. This means that the (σ, μ)-projection boils down to a line parallel to the σ-axis and the minimum variance portfolio is the unique optimal portfolio with respect to the Markowitz partial order. In all other cases, which are considered in Theorem 2.1, the Cauchy-Schwarz inequality implies $[\boldsymbol{\mu}, \boldsymbol{\mu}] [\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2 > 0$. This guarantees that we never divide by zero in all presented formulas.

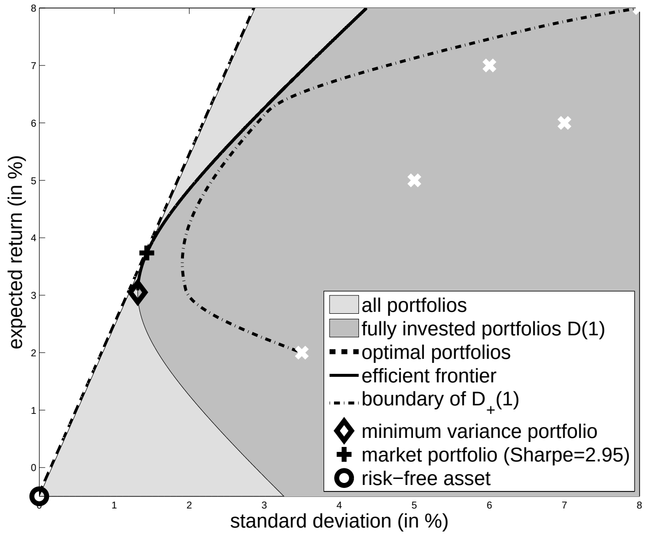

Fig. 1: Visualization of the (σ, μ)-plane, and the projections of the respective portfolios. The dark gray area represents the set $D(1)$ of fully invested portfolios, while the light gray area represents the set $\mathbb{R} \times \mathbb{R}^d$ of all possible portfolios. The upper boundary of the light gray area (the dotted line) represents the set of optimal portfolios, and its point of tangency with the dark gray area represents the market portfolio, i.e. the unique optimal fully invested portfolio. The dotted line within the dark gray area sketches the boundary of the set $D(1)_+$ of fully invested portfolios when short-selling is prohibited. The white crosses represent single risky assets.

While Theorem 2.1 only treats fully invested - and thus truly risky - portfolios, it is clearly also of interest to include portfolios with investment ratios different from one into the analysis. The portfolio $(1, \mathbf{0})$ keeps all money in the risk-free cash account and is represented by the point $(0, r)$ in the (σ, μ)-plane. It is not difficult to enhance the analysis from Theorem 2.1 from $D(1)$ to the larger set

$$\Big\{\big(1-\lambda,\, \lambda\,\mathbf{x}\big) \in \mathbb{R} \times \mathbb{R}^d \;:\; \lambda \in [0, \infty),\, \mathbf{x} \in D(1)\Big\}$$

of portfolios whose total wealth is normalized to one, just like in $D(1)$, but which might have an investment ratio different from one (i.e. essentially all portfolios modulo normalization). By definition (1) of $\mu(.)$ and $\sigma(.)$, the projection of an element $(1-\lambda,\, \lambda\,\mathbf{x})$ of this set into the (σ, μ)-plane lies on the straight line from $(0, r)$ to the point $\big(\sigma(0, \mathbf{x}), \mu(0, \mathbf{x})\big)$, and it wanders along this line with the parameter λ. It is thus clear that there is exactly one such straight line, associated with some $(0, \mathbf{x}_M) \in D(1)$, that has exactly one point of tangency with the efficient frontier. While every portfolio represented by some point on the straight line is optimal, the so-called *market portfolio* $(0, \mathbf{x}_M)$ is the unique fully invested portfolio that is optimal.

Theorem 2.3 (The market portfolio)
Unless $[\mathbf{1}, \boldsymbol{\mu}] = r\,[\mathbf{1}, \mathbf{1}]$, the market portfolio is attained for

$$c = \frac{[\boldsymbol{\mu}, \boldsymbol{\mu}] - r\,[\mathbf{1}, \boldsymbol{\mu}]}{[\mathbf{1}, \boldsymbol{\mu}] - r\,[\mathbf{1}, \mathbf{1}]} \tag{2}$$

in Theorem 2.1, and is given by the formula

$$\mathbf{x}_M = \mathbf{x}(c) = \frac{\Sigma^{-1}\,(\boldsymbol{\mu} - r\,\mathbf{1})}{[\mathbf{1}, \boldsymbol{\mu}] - r\,[\mathbf{1}, \mathbf{1}]}. \tag{3}$$

Proof
See Appendix B. □

Remark 2.4 (Special case)
Notice that the denominator $[\mathbf{1}, \boldsymbol{\mu}] - r\,[\mathbf{1}, \mathbf{1}]$ in Formulas (2) and (3) can be zero, and this case is thus ruled out in Theorem 2.3. In general, the efficient frontier $c_+(s)$ is asymptotically parallel to a straight line through $(0, [\mathbf{1}, \boldsymbol{\mu}]/[\mathbf{1}, \mathbf{1}])$. If $[\mathbf{1}, \boldsymbol{\mu}] = r\,[\mathbf{1}, \mathbf{1}]$, this point becomes $(0, r)$ and the point of tangency defining the market portfolio would be (∞, ∞). The set $D(1)$ is unbounded and there is a sequence $(0, \mathbf{x}_n) \in D(1)$ with $||\mathbf{x}_n|| \to \infty$ as $n \to \infty$ such that the associated sequence of (σ, μ)-projections of $(0, \mathbf{x}_n)$, which is also an unbounded sequence in the plane, may be viewed as an approximation of the optimal portfolio, see also Algorithm 1 below.

Finally, let us gather a few remarks regarding the portfolios $D_+(1)$ with short-selling restriction. The inclusion $D_+(1) \subset D(1)$ can also be observed in the (σ, μ)-plane. The set $D_+(1)$ is not necessarily projected onto the interior of a hyperbola like $D(1)$, as can be seen from Figure 1, where its boundary is visualized. This boundary is always a curve between the two single risky assets with minimal and maximal expected return. In particular, there can be, but need not be, an intersection with the efficient frontier. In classical economic theory, the case of no intersection can be interpreted as the CAPM not working[4]. The article Brennan, Lo (2010) discusses this issue of so-called *impossible frontiers*.

2.2 Connection with the Sharpe ratio

Interestingly, the market portfolio $\mathbf{x}_M$ has a close connection with the function

$$p(x_0, \mathbf{x}) := \frac{(\boldsymbol{\mu} - r\,\mathbf{1})^T\,\mathbf{x}}{\sqrt{\mathbf{x}^T\,\Sigma\,\mathbf{x}}}, \quad (x_0, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^d \setminus \{\mathbf{0}\}. \tag{4}$$

Notice that we write this as function of $(x_0, \mathbf{x})$ despite the fact that it is independent of x_0 in order to indicate that it is a preference function in the sense mentioned in the introduction. The function $p(x_0, .)$ in (4), which is defined only for portfolios with positive investment ratio, is called *Sharpe ratio*.
The Sharpe ratio is scale-invariant in the sense that

$$p(x_0, \mathbf{x}) = p\big(\eta\,x_0, \lambda\,\mathbf{x}\big) \tag{5}$$

[4] The capital asset pricing model (CAPM) is an equilibrium model. Under a set of idealized assumptions, including that all market participants invest according to the Markowitz paradigm, it implies that the weights in the market portfolio equal the shares of the respective assets' value in the total value of all assets, i.e. are non-negative in particular.

for each $\lambda > 0$, $\eta \in \mathbb{R}$. Unless $\boldsymbol{\mu} = r\,\mathbf{1}$ (and hence $p(x_0, .)$ is identically zero and non-interesting), the scale-invariance of the Sharpe ratio implies that there is no way to continuously extend $p(x_0, .)$ to portfolios with zero investment ratio, because this would imply for arbitrary $\mathbf{x} \in \mathbb{R}^d$ that

$$p(x_0, \mathbf{0}) = \lim_{\epsilon \searrow 0} p(x_0, \epsilon\,\mathbf{x}) = p(x_0, \mathbf{x}),$$

which leads to a contradiction for arbitrary $x_0 \in \mathbb{R}$. For a similar reason, the assumption of positive-definiteness of Σ is crucial; the interested reader is referred to Appendix C. The scale invariance of the Sharpe ratio (5) has two important consequences. First, the portfolio $(x_0, \mathbf{x})$ has the same preference as the portfolio $\lambda\,(x_0, \mathbf{x})$ for arbitrary $\lambda > 0$, i.e. portfolio size is not of importance for portfolio preference. This is a property which is not always realistic in applications, thinking for instance of blocking minority stakes in bond prospectuses. Second, and even more importantly in the current context, all portfolios whose (σ, μ)-projections lie on a straight line through $(0, r)$ have the same Sharpe ratio. This can be seen by setting $\eta = 1 - \lambda$ in (5). This already implies the following remarkable statement, manifesting a strong relationship between Markowitz optimality and the Sharpe ratio.

Theorem 2.5 (Sharpe ratio maximality = Markowitz optimality)
The market portfolio $(0, \mathbf{x}_M)$ of Theorem 2.3 is the unique maximizer of $p(x_0, .)$ on the set $D(1)$ of fully invested portfolios.

Proof
Clear by definition of the market portfolio and the remark preceding this theorem. □

Having a closer look at the maximization of the Sharpe ratio, it is further useful to observe that the gradient of $p(x_0, \mathbf{x})$ with respect to $\mathbf{x}$ is given by

$$\nabla_{\mathbf{x}}\, p(x_0, \mathbf{x}) = p(x_0, \mathbf{x}) \left(\frac{(\boldsymbol{\mu} - r\,\mathbf{1})}{(\boldsymbol{\mu} - r\,\mathbf{1})^T \mathbf{x}} - \frac{\Sigma\,\mathbf{x}}{\mathbf{x}^T\,\Sigma\,\mathbf{x}} \right). \qquad (6)$$

By the scale invariance property it suffices to consider $p(.)$ on the set $S_d := \{(0, \mathbf{x}) \,:\, \mathbf{x} \in \mathbb{R}^d,\ ||\mathbf{x}|| = 1\}$, and it is not difficult to see with the help of Equation (6) that $p(.)$ takes its maximum on S_d at $(0, \mathbf{x}_*)$, where

$$\mathbf{x}_* := \frac{\Sigma^{-1}\,(\boldsymbol{\mu} - r\,\mathbf{1})}{||\Sigma^{-1}\,(\boldsymbol{\mu} - r\,\mathbf{1})||}$$

is precisely the normalized version of $\mathbf{x}_M$ according to Theorem 2.3. In other words, the optimal fully invested portfolio - i.e. the market portfolio - can be found analytically by the following algorithm.

Algorithm 1 (Market portfolio via Sharpe ratio maximization)

(i) Maximize the preference function $p(.)$ on S_d to obtain $\mathbf{x}_*$.

(ii) Unless $[\mathbf{1}, \boldsymbol{\mu}] = r\,[\mathbf{1}, \mathbf{1}]$, we have that $\mathbf{x}_M = \mathbf{x}_*/(\mathbf{1}^T\,\mathbf{x}_*)$.

(iii) If $[\mathbf{1}, \boldsymbol{\mu}] = r\,[\mathbf{1}, \mathbf{1}]$, the Sharpe ratio $p(.)$ has no maximum on $D(1)$, but we can find a sequence $(0, \mathbf{x}_n) \in D(1)$ such that $p(0, \mathbf{x}_n)$ converges to $p(0, \mathbf{x}_*)$ by defining $\mathbf{x}_n := \mathbf{y}_n / (\mathbf{1}^T\,\mathbf{y_n})$ for an arbitrary sequence $\mathbf{y}_n \in S_d \setminus \{\mathbf{x}_*\}$ converging to $\mathbf{x}_*$.

With this background in mind, Equation (6) is of great help in daily portfolio management. To explain how, let $(x_0, \mathbf{x})$ denote an existing portfolio. The gradient $\nabla_{\mathbf{x}}\, p(x_0, \mathbf{x})$ tells us precisely which portfolio holdings should be decreased or increased in order to improve the Sharpe ratio of $(x_0, \mathbf{x})$. More precisely, under the assumption that $p(x_0, \mathbf{x}) > 0$, i.e. the current portfolio has positive expected excess return, for $k = 1, \ldots, d$ we have

$$\frac{\partial}{\partial x_k} p(x_0, \mathbf{x}) > 0 \;\Leftrightarrow\; \frac{(\mu_k - r\,\mathbf{1})}{(\boldsymbol{\mu} - r\,\mathbf{1})^T \mathbf{x}} > \frac{\mathbf{e}_k^T\,\Sigma\,\mathbf{x}}{\mathbf{x}^T\,\Sigma\,\mathbf{x}}, \tag{7}$$

where $\mathbf{e}_k$ denotes the k-th unit vector in $\mathbb{R}^d$. Notice in particular that $\mathbf{e}_k^T\,\Sigma\,\mathbf{x}$ may be considered the covariance of the current portfolio (without risk-free cash account) and a portfolio that is fully invested exclusively in asset k. In words, (7) means that the holding of asset k needs to be increased if and only if the ratio of expected excess return of asset k and the current portfolio is larger than the ratio of the covariance of asset k with the current portfolio and the current portfolio variance. Thus, if the excess return contribution of an asset exceeds its variance contribution, its holdings should be increased, and vice versa. The vector within the brackets in (6) may be used to rank the assets in the portfolio in ascending order, with the ones at the bottom of the resulting list being the most urgent ones to be reduced, and the ones on top of the list the most urgent ones to be increased. Obviously, such a ranking is quite useful in daily portfolio monitoring, cf. Section 4.3 for further remarks.

2.3 Connection to power utility maximization

Apparently, the market portfolio $\mathbf{x}_M$ of Theorem 2.3 satisfies the equation

$$(1 - p)\,\Sigma\,\mathbf{x}_M = \boldsymbol{\mu} - r\,\mathbf{1},$$

where the constant p is defined to be $p := 1 - \sum_{k=1}^d (\mu_k - r)$, which is smaller than one under the reasonable assumption that $\mu_k > r$ for each k. This fact implies that the market portfolio $\mathbf{x}_M$ is optimal in the sense of power utility maximization within a multivariate Black-Scholes model. This link between Markowitz optimality and the Black–Scholes model has been established in the seminal papers Merton (1969, 1971). Concretely, the price process $\mathbf{S}(t) = (S_1(t), \ldots, S_d(t))$, $t \in [0, T]$, is modeled as a multivariate geometric Brownian motion with means $\boldsymbol{\mu}$ and covariance matrix Σ. A trading strategy $\boldsymbol{\pi}(t)$ is a stochastic process with the meaning that $\pi_i(t)$ gives the proportion of wealth spent on asset i at time t. It is reasonable to restrict one's focus on self-financing trading strategies, which means that $\boldsymbol{\pi}(t)$ receives an initial amount of wealth at time $t = 0$ but requires no further wealth inflows at later time points. The function $U_p(x) = p^{-1}\,x^p$ is increasing and concave for $p < 1$, called power utility function (notice that for the seemingly ill-defined case $p = 0$ it is consistently defined to equal $U_0(x) = \log(x)$). A reasonable paradigm for portfolio optimization within this dynamic Black–Scholes framework is to maximize the functional

$$\mathbb{E}\Big[U_p\Big(\text{portfolio value at } T \text{ when using trading strategy } \boldsymbol{\pi}(t)\Big)\Big]$$

with respect to $\boldsymbol{\pi}(t)$. Up to a scalar depending on the chosen risk aversion parameter p - that only controls how much money is kept in the risk-free bank account - the optimizer is given by $\boldsymbol{\pi}(t) = \mathbf{x}_M$ for arbitrary $t \geq 0$. In words, the portfolio allocation should be held constant and equal to the market portfolio from Markowitz theory. The fact that the optimal trading strategy keeps the portfolio allocation identically constant over time relies heavily on the assumption that the log-returns in the Black–Scholes model are independent and identically distributed.

3 Limitations and generalizations

The beauty and usefulness of the Markowitz approach towards portfolio selection rely strongly on the fact that optimal portfolios can be computed in closed form. In the following, we discuss some limitations of the Markowitz approach that have to be kept in mind when applying it. For some of them, we indicate potential generalizations, which naturally come along with a loss of mathematical tractability.

3.1 Optimal investment ratio?

In practice, the following question is clearly of interest in daily portfolio management: What is the optimal investment ratio? Unfortunately, the Markowitz setting is not rich enough to help answering this question. This is because the Sharpe ratio is invariant with respect to the investment ratio. Consequently, the task of optimizing the investment ratio is decoupled from the Markowitz setup and has to be carried out separately. Either the investment ratio has to be chosen in a fully discretionary manner by portfolio management, or it might be found by quantitative methods that are different from standard Markowitz theory. A prominent approach is *utility maximization*, which itself is a huge field of research and its description lies outside the scope of the present article.

3.2 Standard deviation as risk measure?

Under the hypothesis that the future portfolio return is normally distributed, there are plausible mathematical arguments as to why the standard deviation is a reasonable portfolio risk measure. Within the Markowitz setting, this hypothesis is justified when assuming that the random vector of the future returns of all d single risky assets is multivariate normally distributed. However, unfortunately this asumption is not always justified. Think of a credit-risky asset which has a non-negligible probability of defaulting. The future return of such an asset can obviously not be modeled well by a normal distribution, since the non-negligible likelihood of a default implies a discontinuity in the return distribution function. In such situations, i.e. when the normality assumption is not justified, standard deviation is not a good measure of risk with its drawbacks being well-understood. For instance, it is not a *coherent risk measure* in the sense of Artzner et al. (1999).
To quickly grasp the point, it is educational to imagine a portfolio consisting of d credit-risky assets, each one exposed to *wipeout*

risk[5], and compare it with a portfolio with identical total invested capital but consisting only of one of the credit-risky assets, say asset XY. Typically, the diversified portfolio is expected to have a smaller standard deviation, since the wipeout risk is reduced significantly on portfolio level (since it is unlikely that all names in the portfolio are wiped out). However, it is easy to imagine a situation in which the asset XY has higher expected return than the diversified portfolio; so high that both portfolios have identical Sharpe ratios. In this case, the Markowitz setting does not distinguish between the two portfolios. However, most portfolio managers, not only the most risk averse ones, would typically prefer the diversified portfolio because the single-asset portfolio faces a significantly larger wipeout risk. Such additional portfolio preferences might be incorporated using the notion of so-called *utility functions* and/or by enhancing the list of key figures to be used for portfolio selection - but the Markowitz setup alone is not rich enough to accomplish this.

3.3 Risk-return perspective sufficient?

Related with the problems associated with the use of standard deviation is the general concern about the loss of essential information when narrowing down something as complex as a multivariate probability distribution of d asset returns to a two-dimensional risk-return measurement. For example, two portfolio returns might have identical first and second moments, but different skewness and kurtosis measurements. The Markowitz setting does not distinguish between the two, even though portfolio management might wish to include selection criteria that can. Of course, one might include even more key figures than just moments into the analysis, e.g. default probabilities in the case of credit-risky assets. However, these generalizations come along with a huge loss in mathematical tractability. The high level of mathematical tractability in the Markowitz setting relies strongly on the fact that the elegant apparatus of linear algebra fits together smoothly with the concept of reducing random variables to their first two moments only. Generally speaking, it is an applied mathematician's most delicate routine to decide whether a model's level of tractability outweighs its lack of realism, or not.

3.4 Estimation of Σ and $\boldsymbol{\mu}$?

It stands to reason to estimate the required inputs Σ and $\boldsymbol{\mu}$ from historical asset return data, which is anything but a straightforward thing to do.

On the one hand, the estimation of Σ is quite challenging in general but at least theoretically possible, with canonical estimators improving with the length of the time series, cf. Lindskog (2000); Ledoit, Wolf (2004) for prominent references on this topic. Theoretically, the more assets are considered (i.e. the larger the matrix dimension d) the longer is the time series length required to obtain reliable estimates. In practice, this is typically not satisfied so that a non-negligible estimation noise remains. Laloux et al. (1999) point out that this estimation noise affects the accuracy of small eigenvalues and associated eigenvectors of Σ more than it does affect large eigenvalues and associated eigenvectors. This

[5] We mean the risk of loosing all money invested into the asset. An example in practice would be a subordinated bond.

is unfortunate because optimal Markowitz portfolios are naturally dominated by eigenvectors associated with small eigenvalues. If $\mathbf{x}$ is an eigenvector associated with a small eigenvalue $\lambda > 0$ satisfying $||\mathbf{x}|| = 1$, the standard deviation $\sqrt{\mathbf{x}^T\,\Sigma\,\mathbf{x}} = \sqrt{\lambda}$ is small. Consequently, either $\mathbf{x}$ or $-\mathbf{x}$ (depending on the sign of $(\boldsymbol{\mu} - r\,\mathbf{1})^T\,\mathbf{x}$) leads to a large Sharpe ratio, inducing a portfolio with large preference in the Markowitz-sense according to Theorem 2.5.

On the other hand, it is well-known that the estimation of the expected future asset returns $\boldsymbol{\mu}$ is impossible based on historical data, even if a normal distribution assumption is well-justified. To see this, denote the price of an asset over time by X_t. Observing this price over the period $[0, T]$ at the n distinct time points $t_i := i/n \cdot T$, $i = 0, 1 \ldots, n$, the sample average of the n observed log returns $\log(X_{t_i}/X_{t_{i-1}})$, $i = 1, \ldots, n$, multiplied by n/T, yields the best estimator according to statistical theory under the hypothesis that log-returns are independent and identically distributed, i.e.

$$\hat{\mu}_n := \frac{1}{T}\sum_{i=1}^{n}\log(X_{t_i}/X_{t_{i-1}}) = \frac{1}{T}\,\log(X_T/X_0).$$

Obviously, this optimal estimator $\hat{\mu}_n$ does not improve with increasing n. No matter how many data points we have available, the best estimator relies solely on two data points: the most recent and the least recent. Needless to say that the confidence interval for $\hat{\mu}_n$ does not shrink at all with increasing number of observations n.

As a consequence of this theoretical impossibility to estimate $\boldsymbol{\mu}$ from historical return data, it is common to derive it from expert opinion and/or scenario assumptions. A typical scenario assumption could be to assume that all essential fundamentals remain constant over the considered time period, consider this the most likely base case scenario, and compute an expected return measurement under this hypothesis. Needless to mention that such an approach entails a massive lack of mathematical rigor. Rather, it has to be viewed as a subjective opinion of portfolio management. A concrete example of such a derivation of $\boldsymbol{\mu}$ based on scenario assumptions is provided in Section 4.

4 An application of the theory to our fund XAIA Credit Curve Carry

We demonstrate how the Markowitz setting provides a relatively simple quantitative tool to optimize the portfolio weights in our fund XAIA Credit Curve Carry. In the light of the preceding section it goes without saying that this approach is not a panacea, but can at best be an indicative assistance tool for an active portfolio management.

A single position in our fund XAIA Credit Curve Carry tries to profit from the shape of an observed CDS curve in the following way. If the CDS curve associated with some reference entity is very steep, the idea is to sell long-dated protection on that name, and buy short-dated protection with the same nominal on the same name. For such a so-called *flattener position*, there are two sources of potential income that can be quantified and that play an important role for the investment idea:

(i) If there is a credit event with respect to the reference entity prior to the maturity of the short-dated CDS, the initial upfront difference between long-dated and short-dated CDS is gained, which is typically positive for steep curves.

(ii) If the CDS curve shape does not change at all over time, that is if today's CDS par spread for an x-year CDS equals the par spread for an x-year CDS in one year (for all x), then the curve steepness implies a "roll-down gain" that is earned, called *roll-down yield* hereafter.

In the complimentary cases when no credit event takes place and the CDS curve shape changes over time (e.g. flattens or steepens), the position makes profits and losses that can be both positive (if flattening) or negative (if steepening). The major assumption of portfolio management is that the aforementioned roll-down yield can be earned over time on statistical average in case no credit event takes place. The fund concept relies on this conviction.

4.1 Definition of r, $\boldsymbol{\mu}$ and Σ

The risk-free rate of return r is retrieved from observed interest rate swap prices based on a 3-month tenor EURIBOR rate, for instance according to one of the bootstrapping algorithms described in Hagan, West (2006, 2008).
According to the investment idea described above, the k-th component μ_k of $\boldsymbol{\mu}$ is defined as

$$\mu_k := p_k \cdot \text{roll-down yield} + (1 - p_k) \cdot \text{upfront difference} + r,$$

with p_k denoting the probability that no credit event takes place before maturity of the short-dated CDS (typically about one year). The probability p_k might be extracted from the CDS curve, for instance by the standard algorithm that bootstraps a piecewise constant default intensity function from CDS prices with different maturities. We add r to the definition of μ_k, because only a small fraction (if any) of the money we receive from our investors is actually spent due to the derivative nature of the involved instruments. We distribute the lion's share of the received capital over a battery of low-risk cash instruments, which we identify with the risk-free asset within the Markowitz model.
The derivation of the matrix Σ can be based on historically observed CDS prices. Each single position in the fund consists of a short CDS with maturity $T^{(s)}$ and a long CDS with maturity $T^{(l)} < T^{(s)}$. Historically, we do observe the price of this flattener position under the assumption that we have rolled both CDS at each past roll date (even though in reality we might not have done so). Consequently, for each single position we observe a price time series of the form $-u_t^{(s)} + u_t^{(l)}$, where $u_t^{(s)}$ denotes the upfront of the (long-dated) short CDS at time t, and $u_t^{(l)}$ the upfront of the (short-dated) long CDS at time t. The potential worst case for such a position at time $t < T^{(l)}$ is the release of information to the marketplace which makes certain that a credit event will take place between $T^{(l)}$ and $T^{(s)}$, and the subsequent CDS auction will yield zero recovery for sure. The worst case loss in this case is $1 - u_t^{(s)} + u_t^{(l)}$, which is always positive and

which we define as the invested capital at time t. Consequently, the log return over the period $[t_1, t_2]$ is reasonably defined by $\log\big((1-u_{t_2}^{(s)}+u_{t_2}^{(l)})/(1-u_{t_1}^{(s)}+u_{t_1}^{(l)})\big)$. From observed time series of daily log returns of all positions in the portfolio according to this definition, we estimate Σ as n times their associated sample covariance matrix, with n denoting the length of the time series. This estimation strategy relies on the assumption that the historically observed vectors of daily log returns are independent and identically distributed, in which case this estimator is well-known to be optimal in the sense of standard statistical theory. However, the remarks in paragraph 3.4 discuss the unavoidable flaws of this estimation problem.

4.2 Looking at our portfolio through Markowitz goggles

Figure 2 visualizes the Markowitz (σ, μ)-plane for our fund XAIA Credit Curve Carry on 25 September 2017. It is observed that our portfolio has a Sharpe ratio of 1.1, while the market portfolio has a significantly better Sharpe ratio of 2.39. However, it can already be observed from the boundary of $D_+(1)$ in Figure 2 that the market portfolio contains short-selling, which is not reasonable in our particular case. In theory, of course one could as well sell the short-dated CDS protection and buy the long-dated CDS protection, i.e. short-sell a flattener position, but this sign change, seemingly innocent in theory, in practice comes along with typically huge bid-ask spreads on both involved CDS, so that the mathematical model, ignoring these transaction costs, becomes unrealistic and useless. Consequently, we can only apply the Markowitz setup with short-selling restriction, i.e. consider $D_+(1)$ instead of $D(1)$. The only interesting part of Figure 2 is hence the location of the white dot (representing our portfolio's (σ, μ)-projection) within the dotted line within the dark gray area (bounding the projection of the set $D_+(1)$ of fully invested portfolios without short-selling). Following the Markowitz paradigm thoroughly, we would like to see that the straight line through $(0, r)$ and the white dot (i.e. the white dotted line) is tangent to the boundary of the (σ, μ)-projection of $D_+(1)$. This not being the case means that we can actually increase the Sharpe ratio of our portfolio by changing our portfolio weights. We even see that one single position is located northwest to the red line, meaning it has a higher Sharpe ratio than our portfolio. This means if we sell our portfolio and instead put all our money into this particular flattener position, we obtain a portfolio that is superior with respect to Markowitz theory. This shows quite clearly how the aforementioned limitations of Markowitz theory lead to essential problems in practice. Our portfolio consists of 25 names, whose weightings deviate from an equally weighted basket only within pre-described boundaries. This discretionary management decision, which is not naturally incorporated in the Markowitz setting, relies on the conviction that a portfolio concentration on too few names leads to an amount of drawdown risk which cannot be outweighed by an increased Sharpe ratio.

4.3 Daily portfolio rebalancing

Even though we have learned about the limitations of Markowitz theory in Section 3, we believe that the related concept of Sharpe ratio maximization can still be useful to assist with certain portfo-

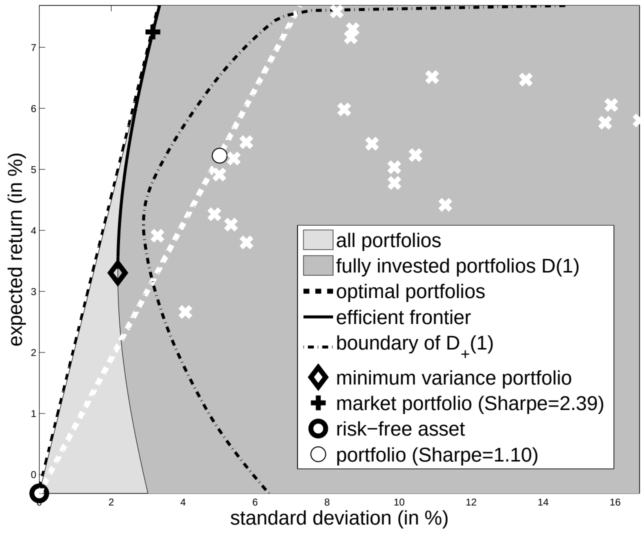

Fig. 2: Visualization of the (σ, μ)-plane for our fund XCCC, when only the d assets we already hold define the cosmos of possible investments. The white crosses represent the single positions in the portfolio, the white dot represents the portfolio.

lio selection decisions that have to be made on a regular basis. This will be explained here, starting with the following question that appears frequently:

(Q) "Should we add a certain new position into our existing portfolio?"

A quantitative answer to this question could, at least partially, be based on the Sharpe ratio, leading to inclusion (rejection) of the new position if it increases (decreases) the portfolio Sharpe ratio. Making use of Equation (7), this can be accomplished by comparing the potential position's contribution to the portfolio return with its contribution to portfolio variance. Based precisely on this idea the XAIA institute article Mai (2017) explores when a macro hedge improves portfolio performance.
Similarly, the following question of daily portfolio weight fine-tuning is frequently of interest:

(Q) "Which holdings in our existing portfolio should be increased/ decreased?"

Again, an answer to this question could be based on the Sharpe ratio to some extent. As already mentioned in paragraph 2.2, Equation (6) induces a ranking of all current portfolio positions, with the top (bottom) names of the ranking being the most urgent ones to be increased (decreased). From a risk-return perspective, performing the related portfolio adjustments moves the white dot in Figure 2 (representing the existing portfolio) into the direction of the market portfolio's (σ, μ)-projection. Despite the aforementioned limitations of Sharpe ratio maximization, the concept of such a Sharpe-ratio based ranking in daily portfolio re-balancing might be useful, since the proposed adjustments could be performed within the bounds of further, discretionary management rules. On a high level, this means that Sharpe ratio maximization is used as one, but not the only one, tool to manage the portfolio.

5 Conclusion

The quantitative basics of modern portfolio theory à la Markowitz have been summarized. In particular, it has been emphasized that Markowitz optimality stands in close connection to Sharpe ratio maximization. Several shortcomings of Markowitz theory have been discussed, and the whole concept has been demonstrated in an application to our fund XAIA Credit Curve Carry.

References

P. Artzner, F. Delbaen, J. Eber, D. Heath, Coherent measures of risk, *Mathematical Finance* **9:3** (1999) pp. 203–228.

T.J. Brennan, W. Lo, Impossible frontiers, *Management Science* **56:6** (2010) pp. 905–923.

E.J. Elton, M.J. Gruber, Modern portfolio theory and investment analysis, 5th edition *John Wiley & Sons, Inc., New York* (1995).

P.S. Hagan, G. West, Interpolation methods for curve construction, *Applied Mathematical Finance* **13:2** (2006) pp. 89–129.

P.S. Hagan, G. West, Methods for constructing a yield curve, *Wilmott magazine* **May** (2008) pp. 70–81.

L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, *Physical Review Letters* **83:7** (1999) pp. 1467–1470.

O. Ledoit, M. Wolf, Honey, I shrunk the covariance matrix, *Journal of Portfolio Management* **31** (2004) pp. 110–119.

F. Lindskog, Linear correlation estimation, *RiskLab report, available from www.risklab.ch/Papers.htm#LCELindskog* (2002).

J.-F. Mai, When does a macro hedge improve portfolio performance?, *XAIA homepage article* (2017).

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

H. Markowitz, Portfolio selection: efficient diversification of investments, *Wiley, New York* (1959).

R.C. Merton, Lifetime portfolio selection under uncertainty: The continuous-time case, *Review of Economics and Statistics* 51:3 (1969) pp. 247–257.

R.C. Merton, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3:4 (1971) pp. 373–413.

A Proof of Theorem 2.1

Instead of minimizing the function $\mathbf{x} \mapsto \sqrt{\mathbf{x}^T \Sigma \mathbf{x}}$, we might as well minimize half of its square, as $x \mapsto x^2/2$ is monotone on $[0, \infty)$. Resorting to the Lagrange multiplier method to accomplish the optimization under side constraints, the Lagrange function under consideration is thus

$$F(\mathbf{x}, \lambda_1, \lambda_2) := \frac{1}{2}\mathbf{x}^T \Sigma \mathbf{x} - \lambda_1 (\mathbf{x}^T \mathbf{1} - 1) - \lambda_2 (\mathbf{x}^T \boldsymbol{\mu} - c).$$

Setting its gradient equal to zero to check for critical values, we obtain the equation system

$$\text{(i) } \Sigma \mathbf{x} = \lambda_1 \mathbf{1} + \lambda_2 \boldsymbol{\mu}, \quad \text{(ii) } \mathbf{x}^T \mathbf{1} = 1, \quad \text{(iii) } \mathbf{x}^T \boldsymbol{\mu} - c.$$

From (i) it is already observed that any critical point $\mathbf{x}$ lies in the span of $\Sigma^{-1} \mathbf{1}$ and $\Sigma^{-1} \boldsymbol{\mu}$. Plugging (i) into (ii) and (iii) gives a linear equation system for the two unknowns λ_1, λ_2, which determines them uniquely as

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} [\boldsymbol{\mu}, \mathbf{1}] & [\mathbf{1}, \mathbf{1}] \\ [\boldsymbol{\mu}, \boldsymbol{\mu}] & [\boldsymbol{\mu}, \mathbf{1}] \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ c \end{pmatrix} = \begin{pmatrix} \frac{c\,[\mathbf{1},\mathbf{1}] - [\mathbf{1},\boldsymbol{\mu}]}{[\boldsymbol{\mu},\boldsymbol{\mu}]\,[\mathbf{1},\mathbf{1}] - [\mathbf{1},\boldsymbol{\mu}]^2} \\ \frac{[\boldsymbol{\mu},\boldsymbol{\mu}] - c\,[\mathbf{1},\boldsymbol{\mu}]}{[\boldsymbol{\mu},\boldsymbol{\mu}]\,[\mathbf{1},\mathbf{1}] - [\mathbf{1},\boldsymbol{\mu}]^2} \end{pmatrix}.$$

Since the set $D(1, c)$ is convex and the function $\mathbf{x} \mapsto \mathbf{x}^T \Sigma \mathbf{x}/2$ is strictly convex on $\mathbb{R}^d$, the unique critical point of the Lagrange function equals the unique minimizer of portfolio standard deviation on $D(1, c)$, which is observed to be $\mathbf{x}(c)$, as claimed. The respective minimum $s(c)$ is computed easily, finishing the proofs of parts (a) and (b). Part (c) is immediate, and part (d) is obtained by inverting the function $s(c)$ of part (b).

B Proof of Theorem 2.3

By part (d) of Theorem 2.1, the slope of a point s on the efficient frontier $c_+(s)$ is given by

$$\frac{\partial}{\partial s} c_+(s) = \frac{\sqrt{[\boldsymbol{\mu}, \boldsymbol{\mu}]\,[\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2}\, s}{\sqrt{s^2\,[\mathbf{1}, \mathbf{1}] - 1}}.$$

By definition, the market portfolio has standard deviation s such that $\big(s, c_+(s)\big)$ equals the unique point of tangency of the efficient frontier with a line through $(0, r)$. Thus, s is characterized by the equation

$$r + \frac{\partial}{\partial s} c_+(s)\, s = c_+(s),$$

which is an equation that can be re-arranged to

$$\sqrt{s^2\,[\mathbf{1}, \mathbf{1}] - 1} = \frac{\sqrt{[\boldsymbol{\mu}, \boldsymbol{\mu}]\,[\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2}}{[\mathbf{1}, \boldsymbol{\mu}] - r\,[\mathbf{1}, \mathbf{1}]}. \tag{8}$$

Using part (d) of Theorem 2.1, this yields the expected return

$$c_+(s) = \frac{[\mathbf{1}, \boldsymbol{\mu}]}{[\mathbf{1}, \mathbf{1}]} + \frac{\sqrt{[\boldsymbol{\mu}, \boldsymbol{\mu}]\,[\mathbf{1}, \mathbf{1}] - [\mathbf{1}, \boldsymbol{\mu}]^2}}{[\mathbf{1}, \mathbf{1}]} \sqrt{s^2\,[\mathbf{1}, \mathbf{1}] - 1} \overset{(8)}{=} \frac{[\boldsymbol{\mu}, \boldsymbol{\mu}] - r\,[\mathbf{1}, \boldsymbol{\mu}]}{[\mathbf{1}, \boldsymbol{\mu}] - r\,[\mathbf{1}, \mathbf{1}]},$$

which implies the claim.

C Well-definedness of the Sharpe ratio requires positive definite Σ

If Σ was singular, there was a non-zero portfolio $\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ of risky assets with zero variance $\mathbf{x}^T \Sigma \mathbf{x} = 0$, i.e. a risk-free asset combined by risky assets. If further $\boldsymbol{\mu}^T \mathbf{x} \neq 0$, then the (non-zero) risk-free asset $\text{sgn}((\boldsymbol{\mu} - r\,\mathbf{1})^T \mathbf{x})\,\mathbf{x}$ would have infinite preference $p\big(x_0, \text{sgn}((\boldsymbol{\mu} - r\,\mathbf{1})^T \mathbf{x})\,\mathbf{x}\big) = \infty$, hence could be considered an arbitrage. Finally, left to discuss is the case that $(\boldsymbol{\mu} - r\,\mathbf{1})^T \mathbf{x} = 0$ for all $\mathbf{x}$ in the kernel of Σ, denoted by ker(Σ), in which well-definedness of $p(x_0, \mathbf{x})$ is a priori unclear due to the fact that we divide zero by zero. However, this turns out to be an irrelevant case, as will be briefly explained. By the assumption in this case, the kernel of Σ is contained in the vector space $\langle \boldsymbol{\mu} - r\,\mathbf{1} \rangle^{\perp}$ consisting of all vectors orthogonal to $\boldsymbol{\mu} - r\,\mathbf{1}$, so is contained in a $(d-1)$-dimensional subspace. There exists an orthogonal basis of $\mathbb{R}^d$ of the form $(\boldsymbol{\mu} - r\,\mathbf{1}, \mathbf{x}_2, \ldots, \mathbf{x}_d)$, with $(\mathbf{x}_m, \ldots, \mathbf{x}_d)$ an orthogonal basis of ker(Σ), for some $m \in \{2, \ldots, d\}$. Any $\mathbf{x} \in \langle \boldsymbol{\mu} - r\,\mathbf{1}_d \rangle \oplus \text{ker}(\Sigma)$ has a unique representation of the form $\mathbf{x} = \lambda_1^{(x)}\,(\boldsymbol{\mu} - r\,\mathbf{1}_d) + \sum_{k=m}^{d} \lambda_k^{(x)}\,\mathbf{x}_k$ with $\lambda_k^{(x)} \in \mathbb{R}$, implying that

$$p(x_0, \mathbf{x}) = \text{sgn}\big(\lambda_1^{(x)}\big)\,\frac{(\boldsymbol{\mu} - r\,\mathbf{1})^T\,(\boldsymbol{\mu} - r\,\mathbf{1})}{\sqrt{(\boldsymbol{\mu} - r\,\mathbf{1})^T\,\Sigma\,(\boldsymbol{\mu} - r\,\mathbf{1})}}.$$

Consequently, $p(x_0, .)$ takes exactly two different values (namely with different signs) on ker(Σ), except for the trivial case $\boldsymbol{\mu} = r\,\mathbf{1}$, in which it is identically zero. Thus, there is no way to extend $p(x_0, .)$ continuously from $\mathbb{R}^d \setminus \text{ker}(\Sigma)$ to all of $\mathbb{R}^d \setminus \{\mathbf{0}\}$ except in the trivial case $\boldsymbol{\mu} = r\,\mathbf{1}$.