Working Paper

Statistical Properties of Estimators for the Log-Optimal Portfolio

Gabriel Frahm

January 22, 2020

Statistical Properties of Estimators for the Log-Optimal Portfolio

Gabriel Frahm

Helmut Schmidt University
Faculty of Economic and Social Sciences
Department of Mathematics and Statistics
Chair of Applied Stochastics and Risk Management
Holstenhofweg 85, D-22043 Hamburg, Germany

URL: www.hsu-hh.de/stochastik
Phone: +49 (0)40 6541-2791
E-mail: frahm@hsu-hh.de

Supervised by: Prof. Dr. Gabriel Frahm
Chair of Applied Stochastics and
Risk Management

URL: www.hsu-hh.de/stochastik

Statistical Properties of Estimators for the Log-Optimal Portfolio

Gabriel Frahm[*]

Helmut Schmidt University

Department of Mathematics and Statistics

Chair of Applied Stochastics and

Risk Management

January 22, 2020

Abstract

The best constant re-balanced portfolio represents the standard estimator for the log-optimal portfolio. It is shown that a quadratic approximation of log-returns works very well on a daily basis and a mean-variance estimator is proposed as an alternative to the best constant re-balanced portfolio. It can easily be computed and the numerical algorithm is very fast even if the number of dimensions is high. Some small-sample and the basic large-sample properties of the estimators are derived. The asymptotic results can be used for constructing hypothesis tests and for computing confidence regions. For this purpose, one should apply a finite-sample correction, which substantially improves the large-sample approximation. However, it is shown that the impact of estimation errors concerning the expected asset returns is serious. The given results confirm a general rule, which has become folklore during the last decades, namely that portfolio optimization typically fails on estimating expected asset returns.

Keywords: Best constant re-balanced portfolio, estimation risk, growth-optimal portfolio, log-optimal portfolio, mean-variance optimization.

JEL Subject Classification: C13, G11

---

[*]Phone: +49 40 6541-2791, e-mail: frahm@hsu-hh.de.

1. Motivation

D URING the last decades, the log-optimal portfolio (LOP) has become increasingly important in portfolio theory. There is a significant number of publications related to the LOP—or to the growth-optimal portfolio (GOP), which is often treated synonymously. The reader can find a huge number of articles in MacLean et al. (2011). For an overview on the subject matter see Christensen (2005). In spite of a controversial debate (Merton and Samuelson, 1974), it cannot be denied that the LOP has a number of nice and beautiful properties. For example, it is asymptotically optimal among all portfolios that share the same constraints on the portfolio weights (Cover and Thomas, 1991, Chapter 15). Moreover, the LOP can be considered a discrete-time approximation of the GOP, which serves as a numéraire portfolio and thus plays a major role in financial mathematics (Karatzas and Kardaras, 2007, Platen and Heath, 2006). The GOP provides a link between financial mathematics, neoclassical finance, and financial econometrics (Frahm, 2016). Hence, the LOP is of particular interest for a variety of reasons.

In this work, the statistical properties of LOP estimators are investigated. To the best of my knowledge, this is not done so far in the literature. We will consider the standard estimator for the LOP, i.e., the best constant re-balanced portfolio (BCRP), and the mean-variance estimator (MVE), which is based on a quadratic approximation of log-returns. The question of whether or not the BCRP or the MVE outperforms any other investment strategy is not discussed in this work. This seems to be well-investigated in the literature. In particular, the BCRP and the MVE are not compared with one another in order to clarify whether maximizing the logarithmic utility or the mean-variance objective function is more preferable (Hakansson, 1971). Instead, the MVE is used only to approximate the BCRP. Correspondingly, the mean-variance optimal portfolio (MVOP) does not end in itself. Here, it just represents an approximation of the LOP.

The main conclusions of this work are as follows:

(i) The MVE provides a very good approximation of the BCRP if re-balancing takes place on a daily basis. The numerical implementation of the MVE is quite easy and the corresponding algorithm is very fast even if the number of dimensions is high.

(ii) One typically overestimates the expected out-of-sample log-return on the BCRP and even the expected log-return on the LOP. Similar statements hold true for the expected out-of-sample performance of the MVE and the performance of the MVOP.

(iii) The BCRP exists and is unique under mild regularity conditions. Moreover, it is strongly consistent, which holds true also for the expected out-of-sample log-return on the BCRP and its in-sample average log-return. Similar results are obtained for the MVE.

(iv) Although both the BCRP and the MVE are affected by short-selling constraints, they are $\sqrt{n}$-consistent. The asymptotic results can be used in order to construct hypothesis tests and to compute confidence regions.

(v) Due to the constraints on the portfolio weights, the asymptotic results are inaccurate in most practical applications. Nonetheless, a finite-sample correction exists. It substantially improves the large-sample approximation of the MVE (and thus of the BCRP).

(vi) However, the impact of estimation risk that comes from estimating expected asset returns is tremendous in most real-life situations. This problem is so serious that estimating the LOP becomes a futile endeavour if we have no prediction power.

The rest of this work is organized as follows: In Section 2 the basic assumptions are made and the mathematical notation is explained. Section 3 contains some elementary results and provides a simple characterization of the LOP. In Section 4 the small-sample and large-sample properties of the BCRP are derived, which includes its existence, uniqueness, and consistency. That section contains also the asymptotic distribution of the BCRP. The reader can find the corresponding results for the MVE in Section 5. In Section 6 some computational issues that are related to the BCRP are discussed and the finite-sample correction for the MVE is demonstrated. Section 7 concludes this work. Finally, the appendix contains an important but quite tedious derivation.

2. Basic Assumptions and Notation

Throughout this work, $\mathbb{N}$ denotes the set of positive integers, i.e., $\mathbb{N} := \{1, 2, \dots\}$, and the symbol "log" stands for the natural logarithm. The symbol $\mathbf{0}$ denotes a vector of zeros and $\mathbf{1}$ is a vector of ones. The dimensions of $\mathbf{0}$ and $\mathbf{1}$ should always be clear from the context. Any tuple $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ is understood to be a *column* vector and $x' = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}$ is the transpose of x. It is implicitly assumed that we have an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where Ω is some state space, $\mathcal{A}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on $\mathcal{A}$, which is often referred to as the physical or real-world probability measure in financial mathematics (Frahm, 2016). A random quantity is a (measurable) real-valued function on Ω. According to probability theory, two random quantities are considered identical if and only if they coincide with probability 1. Analogously, any statement about a random quantity is meant to be true almost surely (a.s.). Hence, we can drop the additional remarks "$\mathbb{P}(\cdot) = 1$" and "a.s." for convenience. For example, if X is a random variable, "$X = x$" with $x \in \mathbb{R}$ means that $\mathbb{P}(X = x) = 1$ and if Y is a random variable, too, then "$X > Y$" means that X is greater than Y with probability 1, etc. Further, if $\{X_n\}_{n \in \mathbb{N}}$ is a random sequence, "$X_n \to x$" means that X_n converges a.s. to x as n tends to infinity. Thus, we may drop also the notation "$n \to \infty$."

Consider an asset universe with one riskless asset and $N \in \mathbb{N}$ risky assets. It is assumed that the assets are infinitely divisible and any market frictions are ignored. Let $S_t = (S_{0t}, S_{1t}, \dots, S_{Nt})$ be the vector of asset prices at time $t = 0, 1, \dots$, where S_{0t} denotes the price of the riskless asset. The unit of time is one trading day. It is assumed that $S_{0t} = 1$ for $t = 0, 1, \dots$ and that $S_{i0} = 1$ for $i = 1, 2, \dots, N$. In the following, each statement that contains the index i or t is meant to be true for all i and t that are appropriate in the given context. The price process $\{S_t\}_{t=0,1,\dots}$ shall be positive. The time-index set is always $\{0, 1, \dots\}$ and thus the subscript in "$\{\cdot\}_{t=0,1,\dots}$" will be omitted for notational convenience.

Let $X_t := S_t / S_{t-1}$ be the vector of price relatives after the trading day t, where the division of S_t by S_{t-1} is understood to be componentwise. Any capital appreciation for Asset i during Day t, e.g., interest or dividend income, is considered part of the asset price S_{it}. The portfolio weights of the risky assets are denoted by $w_1, w_2, \dots, w_N$, whereas w_0 is the weight of the riskless asset.

Hence, $w = (w_0, w_1, \ldots, w_N) \in \mathbb{R}^{N+1}$ is a portfolio that consists of the riskless asset and N risky assets. Each single asset is considered a portfolio, i.e., a canonical vector in $\mathbb{R}^{N+1}$. In order to distinguish the weights of the risky assets, $w_1, w_2, \ldots, w_N$, from the weight of the riskless asset, w_0, the notation $\tilde{w} = (w_1, w_2, \ldots, w_N) \in \mathbb{R}^N$ is used. This means that $w = (w_0, \tilde{w})$. Analogously, $\tilde{X}_t$ indicates the risky part of X_t, i.e., we have that $X_t = (1, \tilde{X}_t)$. Finally, the return on Asset i after Day t is given by $R_{it} := X_{it} - 1$ and $R_t = \tilde{X}_t - \mathbf{1} = (R_{1t}, R_{2t}, \ldots, R_{Nt})$ denotes the vector of risky asset returns. Since we assume that $S_{0t} = 1$ for $t = 0, 1, \ldots$, the risk-free interest rate is supposed to be zero. This assumption is made without loss of generality, which will be explained below.

Although the following terms will be defined later on, their notation is used throughout this work and so it shall be clarified beforehand: The symbol $w^* = (w_0^*, w_1^*, \ldots, w_N^*)$ denotes the LOP, whereas $w^\star = (w_0^\star, w_1^\star, \ldots, w_N^\star)$ is the MVOP. Note that the former superscript, "$*$," has 6 spikes, whereas the latter, "$\star$," consists of 5 spikes. This shall symbolize a key observation of this work, namely that the LOP and the MVOP are almost indistinguishable in most practical applications, which hopefully does not hold for the symbols themselves. The symbols $\tilde{w}^* = (w_1^*, w_2^*, \ldots, w_N^*)$ and $\tilde{w}^\star = (w_1^\star, w_2^\star, \ldots, w_N^\star)$ denote the "risky parts" of w^* and $w^\star$, respectively. Further, w_i is the portfolio weight of Asset i. By contrast, w_n is an estimator for the *portfolio w*, where $n \in \mathbb{N}$ is the number of observations.[1] Consequently, w_{in} is the estimator for the portfolio weight of Asset i.

At the end of each trading day, the investor re-balances his portfolio according to a *constant* vector of portfolio weights w satisfying the budget constraint $\mathbf{1}'w = 1$. The portfolio value at Day $n \in \mathbb{N}$ amounts to $V_{wn} := \prod_{t=1}^n w'X_t$. The investment capital might vanish during some trading day if we do not pose any additional constraints on the portfolio weights. In fact, if we allow the investor to enter short positions, the probability of going bankrupt, i.e., $V_{wn} \leq 0$, is positive unless we make some additional assumption about $\{X_t\}$, but this is omitted in this work. Hence, the portfolio w must be an element of the (unit) simplex

$$\mathcal{S} := \left\{ w \in \mathbb{R}^{N+1} \colon w \geq \mathbf{0} \ \wedge \ \mathbf{1}'w = 1 \right\}.$$

The assumption that $w \in \mathcal{S}$ is crucial. It guarantees that $w'X_t > 0$ so that $V_{wn} > 0$ for all $n \in \mathbb{N}$. Hence, the log-value process $\log V_{wn} = \sum_{t=1}^n \log w'X_t$ exists for all $w \in \mathcal{S}$ and $n \in \mathbb{N}$, where $\log w'X_t$ is referred to as the log-return on the portfolio after Day t. The short-selling constraints are indispensable because otherwise $\log w'X_t$ might not be defined.

As already mentioned above, we can assume without loss of generality that the risk-free interest rate is zero: Let $r \equiv R_0 > -1$ be the risk-free interest rate and Y_t the vector of relative prices with $Y_{0t} = 1 + r$. Then we could use the *discounted* relative-price vector $X_t := Y_t / (1 + r)$. The log-return on any portfolio $w \in \mathcal{S}$ amounts to $\log(w'Y_t) = \log(1 + r) + \log(w'X_t)$. We can ignore the first term, $\log(1 + r)$, provided we are interested only in maximizing the expected log-return on the portfolio w, which is our main focus here. Throughout this work, we suppose that X_t contains the discounted relative prices but omit the word "discounted" for convenience.

Now, the following basic assumptions are made:

A1. The relative-price process $\{X_t\}$ is strictly stationary,

---

[1] Actually, it is assumed that $n > N$ and so there should be no confusion.

A2. the expected value of $\log w'X_t$ is finite for all $w \in \mathcal{S}$, and

A3. $w_1'X_t$ and $w_2'X_t$ do not coincide for any $w_1, w_2 \in \mathbb{R}^{N+1}$ with $w_1 \neq w_2$.

A1 is a fundamental assumption in econometrics and implies that the elements of $\{X_t\}$ are identically distributed, but it is *not* assumed that they are serially independent. Further, **A2** guarantees that we can work with the quantity $\mathbf{E}(\log w'X_t)$ and **A3** requires the relative prices to span $(0, \infty)^{N+1}$. It follows that no risky asset can be replicated by a convex combination of other assets. More precisely, since $\mathbb{P}(w_1'X_t = w_2'X_t) \neq 1$ for all $w_1, w_2 \in \mathbb{R}^{N+1}$ with $w_1 \neq w_2$, it holds that $\mathbb{P}(w'X_t = 0) \neq 1$ for all $w \in \mathbb{R}^{N+1}$ with $w \neq \mathbf{0}$ and vice versa. Hence, it cannot happen that $\tilde{w}'R_t = c \in \mathbb{R}$ for any $\tilde{w} \in \mathbb{R}^N$ with $\tilde{w} \neq \mathbf{0}$. Otherwise, we could define $w_0 := -(c + \mathbf{1}'\tilde{w})$ so that $w'X_t = -(c + \mathbf{1}'\tilde{w}) + (\mathbf{1}'\tilde{w} + c) = 0$ with $w \neq \mathbf{0}$. Conversely, if there is no $\tilde{w} \in \mathbb{R}^N$ with $\tilde{w} \neq \mathbf{0}$ such that $\tilde{w}'R_t = c \in \mathbb{R}$, then we cannot have that $w'X_t = 0$ with $w \neq \mathbf{0}$ because otherwise $\tilde{w}'R = -(w_0 + \mathbf{1}'\tilde{w}) \in \mathbb{R}$ with $\tilde{w} \neq \mathbf{0}$.

To sum up, it holds that $\mathbb{P}(w'X_t = 0) \neq 1 \Leftrightarrow \mathbb{P}(\tilde{w}'R_t = c) \neq 1$ with $\tilde{w} \neq \mathbf{0}$. How can we interpret this basic condition from an economical point of view? Suppose that $\tilde{w}'R_t = c \neq 0$. Then we have an arbitrage opportunity, which is not possible if the market is in equilibrium and the market participants are rational (Frahm, 2018). By contrast, in the case that $\tilde{w}'R_t = 0$, at least one risky asset is redundant. For example, let us assume that $\tilde{w}_N \neq 0$ without loss of generality. Then we can construct the portfolio $\tilde{v} := -\tilde{w}/\tilde{w}_N$ of risky assets so that $\tilde{v}'R_t = 0$ with $\tilde{v}_N = -1$. Now, we are able to replicate Asset N by a linear combination of all other assets, including the riskless asset, by using the portfolio $(1 - \mathbf{1}'(\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{N-1}), \tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{N-1}) \in \mathbb{R}^N$, which satisfies the budget constraint. Hence, we can ignore Asset N and reduce the asset universe to $N - 1$ risky assets. Of course, also the converse is true. That is, if we are able to replicate a risky asset by linear combination of all other assets, we must have that $\tilde{w}'R_t = 0$ with $\tilde{w} \neq \mathbf{0}$.

Throughout this work, it is assumed that the dimension reduction has already been made in advance. Note also that without the dimension reduction the covariance matrix of R_t would not be positive definite because $\tilde{w}'R_t = c$ implies that $\tilde{w}'\mathbf{Var}(R_t)\tilde{w} = \mathbf{Var}(\tilde{w}'R_t) = 0$ for $\tilde{w} \neq \mathbf{0}$ and vice versa. Hence, **A3** is indispensable also for statistical reasons. In fact, this assumption plays a major role in portfolio theory, where it is typically required that $\mathbf{Var}(R_t) > 0$, i.e., that the covariance matrix of the risky asset returns is positive definite.

3. The Log-Optimal Portfolio

Definition 1. *A log-optimal portfolio is a portfolio $w^* \in \mathcal{S}$ that maximizes the expected log-return, i.e.,*

$$w^* \in \arg\max_{w \in \mathcal{S}} \mathbf{E}(\log w'X_t).$$

The LOP is often associated with the "Kelly criterion" (Kelly, 1956). Its asymptotic optimality properties are elaborated by Algoet and Cover (1988), Bell and Cover (1980) as well as Breiman (1961).[2] Although it was originally studied in information theory, it became of growing interest

---

[2]See also Chapter 15 in Cover and Thomas (1991).

to the finance community over the last decades. As already mentioned in Section 1, the LOP is sometimes referred to as the GOP. However, the GOP is typically studied in a continuous-time framework, whereas the LOP is based on a discrete-time setting.[3]

The Lagrange function of the optimization problem given by Definition 1 is

$$\mathcal{L}(w, \kappa, \lambda) = -\mathbf{E}\big(\log w' X_t\big) - \kappa' w + \lambda(\mathbf{1}' w - 1)$$

with $\kappa = (\kappa_0, \kappa_1, \ldots, \kappa_N) \geq \mathbf{0}$ and $\lambda \in \mathbb{R}$. The corresponding Karush–Kuhn–Tucker (KKT) conditions are quite nice (Cover and Thomas, 1991, Theorem 15.2.1). The following theorem establishes also the existence and uniqueness of the LOP.

Theorem 1. *The LOP exists and is unique. It is characterized by $w^* \in \mathcal{S}$ such that*

$$\mathbf{E}\left(\frac{X_{it}}{w^{*\prime} X_t}\right) \begin{cases} = 1, & w_i^* > 0 \\ \leq 1, & w_i^* = 0 \end{cases}.$$

Proof: The simplex $\mathcal{S}$ is compact and convex. Further, the random variables $v' X_t$ and $w' X_t$ do not coincide for any $v, w \in \mathcal{S}$ with $v \neq w$. Hence, since the natural logarithm is strictly concave, for each $0 < \pi < 1$ and $v, w \in \mathcal{S}$ with $v \neq w$ it holds that

$$\log \big[\pi v + (1 - \pi) w\big]' X_t = \log \big(\pi v' X_t + (1 - \pi) w' X_t\big) \geq \pi \log v' X_t + (1 - \pi) \log w' X_t$$

and

$$\mathbb{P}\Big(\log \big[\pi v + (1 - \pi) w\big]' X_t > \pi \log v' X_t + (1 - \pi) \log w' X_t \Big) > 0.$$

This means that the objective function $w \mapsto \mathbf{E}\big(\log w' X_t\big)$ is strictly concave, which implies that w^* exists and is unique. Further, the partial difference quotient

$$
\begin{aligned}
\frac{\log \big(w' x + \Delta w_i x_i\big) - \log w' x}{\Delta w_i} &= \frac{1}{\Delta w_i} \log \frac{w' x + \Delta w_i x_i}{w' x} \\
&= \frac{1}{\Delta w_i} \log \left(1 + \frac{\Delta w_i x_i}{w' x}\right) > 0, \qquad \Delta w_i > 0,
\end{aligned}
$$

increases monotonically and tends to $x_i / w' x > 0$ as $\Delta w_i \searrow 0$ for each $x > 0$. From the Monotone Convergence Theorem, we conclude that

$$\frac{\partial}{\partial w} \mathbf{E}\big(\log w' X_t\big) = \mathbf{E}\left(\frac{X_t}{w' X_t}\right).$$

Hence, we have that $\mathbf{E}\big(X_t / w^{*\prime} X_t\big) = \lambda \mathbf{1} - \kappa$ with $w^* \in \mathcal{S}$, $\lambda \in \mathbb{R}$, $\kappa = (\kappa_0, \kappa_1, \ldots, \kappa_N) \geq \mathbf{0}$, and $w_i^* \kappa_i = 0$. From $w^{*\prime} \mathbf{E}\big(X_t / w^{*\prime} X_t\big) = \mathbf{E}\big(w^{*\prime} X_t / w^{*\prime} X_t\big) = 1$, $w^{*\prime} \mathbf{1} = 1$, and $w^{*\prime} \kappa = 0$, we conclude that $\lambda = 1$. Thus, we obtain

$$\mathbf{E}\left(\frac{X_t}{w^{*\prime} X_t}\right) = \mathbf{1} - \kappa,$$

which leads to the given expression in the theorem. Q.E.D.

---

[3]See Karatzas and Kardaras (2007) for a detailed analysis of the GOP.

The portfolio weight w_i^* is bounded by $\mathcal{S}$ if and only if $\mathbf{E}(X_{it}/w^{*\prime}X_t) < 1$, whereas the partial derivative equals 1 whenever $w_i^* > 0$. If all (optimal) portfolio weights are positive, the solution of the optimization problem, w^*, lies in the interior of $\mathcal{S}$, which is denoted by $\mathcal{S}^{\mathrm{o}}$. Since we have that

$$(w - w^*)'\mathbf{E}\left(\frac{X_t}{w^{*\prime}X_t}\right) = w'\underbrace{\mathbf{E}\left(\frac{X_t}{w^{*\prime}X_t}\right)}_{=\,1} - \underbrace{\mathbf{E}\left(\frac{w^{*\prime}X_t}{w^{*\prime}X_t}\right)}_{=1} = 1 - 1 = 0\,, \qquad \forall\, w \in \mathcal{S}^{\mathrm{o}}\,,$$

the expected log-return stays constant after a local change of the portfolio weights.[4] This could be true even on the boundary of $\mathcal{S}$, $\partial\mathcal{S}$, as long as $\mathbf{E}(X_t/w^{*\prime}X_t) = \mathbf{1}$. In this case, all portfolio weights are still unbounded by $\mathcal{S}$. By contrast, if (at least) one partial derivative is lower than 1, some portfolio weight must be zero, i.e., $w^* \in \partial\mathcal{S}$. Then the expected log-return decreases after a local change of a portfolio weight that is bounded by $\mathcal{S}$. These basic considerations will be important later on when deriving the asymptotic properties of the LOP estimators.

4. The Best Constant Re-Balanced Portfolio

Definition 2. *A best constant re-balanced portfolio is a portfolio $w_n^* \in \mathcal{S}$ that maximizes the in-sample average log-return, i.e.,*

$$w_n^* \in \arg\max_{w \in \mathcal{S}} \ \frac{1}{n}\sum_{t=1}^{n} \log w'X_t\,. \tag{1}$$

The relative prices contained in $X_1, X_2, \dots, X_n$, except for the relative price 1 of the riskless asset, are nondegenerate random variables and so, in general, $\frac{1}{n}\sum_{t=1}^{n}\log w'X_t$ is a nondegenerate random variable, too. This means that for each element of the state space, $\omega \in \Omega$, and so for each realization of $X_1, X_2, \dots, X_n$, we maximize $\frac{1}{n}\sum_{t=1}^{n}\log w'X_t(\omega)$, which leads us to a particular realization, $w_n^*(\omega)$, of a BCRP w_n^*, which thus represents a random vector.

A BCRP can be considered an empirical version of the LOP. It is said to be the "best" constant re-balanced portfolio because w_n^* maximizes the final value after Period $n \in \mathbb{N}$, i.e., V_{wn}, over all constant re-balanced portfolios $w \in \mathcal{S}$. However, the maximization is done in *hindsight*, i.e., after all asset prices have been revealed to the investor, and thus the BCRP is unknown in advance.

4.1. Small-Sample Properties

4.1.1. Existence and Uniqueness

Let $n \in \mathbb{N}$ be the number of price observations. The following additional assumption is made for statistical reasons:

A4. The sample of price relatives, i.e., $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}$, has rank $N + 1$.

A4 can be considered an empirical version of **A3**, which implies that no risky asset is redundant. It requires that the number of observations exceeds the number of risky assets, i.e., $n > N$.

---

[4]Note that, after any local change, w must still belong to $\mathcal{S}$.

Theorem 2. *The BCRP exists and is unique. It is characterized by $w_n^* \in \mathcal{S}$ such that*

$$\frac{1}{n} \sum_{t=1}^{n} \frac{X_{it}}{w_n^{*\prime} X_t} \begin{cases} = 1, & w_{in}^* > 0 \\ \leq 1, & w_{in}^* = 0 \end{cases}.$$

Proof: Since $w \mapsto \frac{1}{n} \sum_{t=1}^{n} \log w' X_t$ is a concave objective function, Eq. 1 represents a convex optimization problem and, because the simplex is compact and convex, the BCRP exists. The rank of $\mathbf{X}$ is full so that $v, w \in \mathbb{R}^{N+1}$ must lead to different value processes unless $v = w$. That is, the given objective function is strictly concave, which implies that the BCRP is unique. The rest of the proof follows by the arguments that are used in the proof of Theorem 1. Q.E.D.

A simple numerical algorithm for computing the BCRP is developed by Cover (1984). We will come back to this point in Section 6.

4.1.2. Finite-Sample Bias

Let w_n be a portfolio that is based only on the price observations that have been made up to Day n. A standard assumption of portfolio theory is that w_n is stochastically independent of R_{n+1} or, equivalently, of X_{n+1} (Frahm, 2015). If w_n would depend on R_{n+1}, the decision of the investor at time n would be influenced by some asset returns at time $n + 1$ or, vice versa, his financial transactions would have an impact on forthcoming asset prices. In this case he would be able to predict the future price evolution on the basis of past asset prices. This is typically ruled out in finance theory and, especially, in portfolio theory. Put another way, we assume that the investor has no prediction power. This basic assumption will be elaborated also in Section 5.1.2.

For example, suppose that $X_1, X_2, \ldots, X_{n+1}$ are serially independent. Since w_n is a function of $X_1, X_2, \ldots, X_n$, the portfolio w_n does not dependent on X_{n+1}. However, the converse is not true. Consider some (measurable) real-valued function f of some random variable ξ. The fact that $f(\xi)$ is independent of another random variable ζ does not imply that ξ is independent of ζ. A trivial example is any constant function f. Another well-known and more sophisticated example is the case in which $\xi_1, \xi_2, \ldots, \xi_n$ are independent and identically normally distributed. Obviously, $\xi_{n+1} := \frac{1}{n} \sum_{t=1}^{n} \xi_t$ depends on $\xi_1, \xi_2, \ldots, \xi_n$, but it is known that $f(\xi_1, \xi_2, \ldots, \xi_n) = \sum_{t=1}^{n} (\xi_t - \xi_{n+1})^2$ is independent of ξ_{n+1}.[5] Thus, although ξ_{n+1} depends on $\xi_1, \xi_2, \ldots, \xi_n$ and $f(\xi_1, \xi_2, \ldots, \xi_n)$ is not constant, but a nondegenerate random variable, $f(\xi_1, \xi_2, \ldots, \xi_n)$ is still independent of ξ_{n+1}. We conclude that, although $X_1, X_2, \ldots, X_{n+1}$ may be serially dependent, a (random) portfolio based on $X_1, X_2, \ldots, X_n$ need not depend on X_{n+1}.

Hence, we can make the following additional assumptions:

A5. The BCRP w_n^* is stochastically independent of X_{n+1}.

A6. The BCRP does not coincide with the LOP, i.e., $\mathbb{P}(w_n^* = w^*) \neq 1$.

A6 just states that $w_n^* = w^*$ holds only with probability *lower* than 1. This assumption is trivial, since otherwise we would not have any estimation risk at all.

---

[5]This follows from Cochran's Theorem and represents a special result of Basu's Theorem.

Let X be any positive random vector that has the same distribution as the vectors $X_1, X_2, \ldots$ of price relatives and define the following quantities:

- $\varphi(w) := \mathbf{E}\big(\log w'X\big)$,

- $\varphi_n(w) := \mathbf{E}\big(\frac{1}{n}\sum_{t=1}^{n}\log w'X_t\big)$, and

- $\varphi_{n+1}(w) := \mathbf{E}\big(\log w'X_{n+1}\big)$.

Hence, by substituting w with w^* or w_n^*, respectively, we can see that

- $\varphi(w^*)$ is the expected log-return on the LOP,

- $\varphi_n(w_n^*)$ represents the expected in-sample average log-return on the BCRP, and

- $\varphi_{n+1}(w_n^*)$ denotes the expected out-of-sample log-return on the BCRP.

The investor cannot achieve $\varphi(w^*)$ because the LOP is unknown to him. Instead, he maximizes the average log-return $\frac{1}{n}\sum_{t=1}^{n}\log w_n^{*\prime}X_t$ in order to compute the BCRP. At the end of Day n he applies the BCRP and one day later he obtains the log-return $\log w_n^{*\prime}X_{n+1}$. For this reason, $\varphi_{n+1}(w_n^*)$ may be considered the basic performance measure for w_n^*.

The following theorem describes why the BCRP might lead to wrong conclusions in real-life situations, especially if the number of observations, n, is small.

Theorem 3. $\varphi_{n+1}(w_n^*) < \varphi(w^*) < \varphi_n(w_n^*)$

Proof: By definition, w^* is the element of $\mathcal{S}$ that maximizes the expected log-return. Moreover, due to **A5** and **A6**, and the fact that w^* is unique, we have that

$$\mathbf{E}\big(\log w'X_{n+1} \mid w_n^* = w\big) = \mathbf{E}\big(\log w'X_{n+1}\big) \leq \mathbf{E}\big(\log w^{*\prime}X_{n+1}\big)$$

with probability 1 but $\mathbf{E}\big(\log w'X_{n+1}\big) < \mathbf{E}\big(\log w^{*\prime}X_{n+1}\big)$ with positive probability. From the Law of Total Expectation and the stationarity of $\{X_t\}$ we conclude that

$$\varphi_{n+1}(w_n^*) = \mathbf{E}\Big(\mathbf{E}\big(\log w_n^{*\prime}X_{n+1} \mid w_n^*\big)\Big) < \mathbf{E}\big(\log w^{*\prime}X_{n+1}\big) = \mathbf{E}\big(\log w^{*\prime}X\big) = \varphi(w^*)\,.$$

Moreover, since w_n^* is unique and does not coincide with w^*, we have that

$$\mathbb{P}\left(\frac{1}{n}\sum_{t=1}^{n}\log w_n^{*\prime}X_t \geq \frac{1}{n}\sum_{t=1}^{n}\log w^{*\prime}X_t\right) = 1$$

and

$$\mathbb{P}\left(\frac{1}{n}\sum_{t=1}^{n}\log w_n^{*\prime}X_t > \frac{1}{n}\sum_{t=1}^{n}\log w^{*\prime}X_t\right) > 0.$$

This means that

$$\varphi_n(w_n^*) = \mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}\log w_n^{*\prime}X_t\right) > \mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}\log w^{*\prime}X_t\right) = \mathbf{E}\big(\log w^{*\prime}X\big) = \varphi(w^*)\,.$$

<div align="right">Q.E.D.</div>

Hence, the expected out-of-sample log-return on the BCRP, $\varphi_{n+1}(w_n^*)$, is always lower than the expected log-return on the LOP, i.e., $\varphi(w^*)$. Nonetheless, the investor typically *overestimates* not only $\varphi_{n+1}(w_n^*)$ but even $\varphi(w^*)$ when computing $\varphi_n(w_n^*)$ by maximizing $\frac{1}{n}\sum_{t=1}^n \log w'X_t$. This phenomenon is not limited to the BCRP. It is a general problem of portfolio optimization (see, e.g., Frahm, 2015, Frahm and Memmel, 2010, Kan and Zhou, 2007, Memmel, 2004).

4.2. Large-Sample Properties

4.2.1. Consistency

For the subsequent analysis it is convenient to define the function $x \mapsto f_w(x) := \log w'x$ for all $w \in \mathcal{S}$ and $x > 0$ as well as the functions

$$w \mapsto M(w) := \mathbf{E}\big(f_w(X)\big) \qquad \text{and} \qquad w \mapsto M_n(w) := \frac{1}{n}\sum_{t=1}^n f_w(X_t)$$

for all $n \in \mathbb{N}$. We make the following statistical assumption, which is often used in the theory of empirical processes (see, e.g., van der Vaart, 1998, Chapter 19):

A7. The family $\mathcal{F} = \big\{f_w\big\}_{w \in \mathcal{S}}$ is Glivenko-Cantelli, i.e.,

$$\sup_{w \in \mathcal{S}} |M_n(w) - M(w)| \to 0\,.$$

Hence, the Strong Law of Large Numbers shall hold true for the sequence $\big\{M_n(w)\big\}$ *uniformly* in $\mathcal{S}$. For example, according to van der Vaart (1998, p. 46), it is sufficient to guarantee that

(i) w stems from a compact set,

(ii) the elements of $\mathcal{F}$ are continuous for every $x > 0$, and

(iii) they are dominated by an integrable function,

provided $X_1, X_2, \ldots$ are serially independent.[6] The first two properties are clearly satisfied in our context. In order to see that the third property is satisfied, too, note that

$$-\mathbf{1}'(\log x)^- \leq w' \log x \leq \log w'x \leq \log \mathbf{1}'x$$

for all $x > 0$, where $(\log x)^-$ denotes the negative part of the vector $\log x$. Hence, the function

$$x \mapsto g(x) = \max\Big\{\mathbf{1}'(\log x)^-, \log \mathbf{1}'x\Big\}$$

dominates each f_w. Since $\mathbf{E}\big(|\log w'X_t|\big) < \infty$ for all $w \in \mathcal{S}$, we have that $\mathbf{E}\big(|\log X_{it}|\big) < \infty$ and thus $\mathbf{E}\big(\mathbf{1}'(\log X_t)^-\big) < \infty$. Moreover, note that $\mathbf{E}\left(\log w'X_t\right) < \infty$ for $w = \mathbf{1}/N$ and $\mathbf{1}'X_t > 1$

---

[6]See also Example 19.8 in van der Vaart (1998).

because $X_{0t} = 1$. Hence, we obtain

$$\infty > \log N + \mathbf{E}\left(\log \frac{\mathbf{1}'X_t}{N}\right) = \mathbf{E}\left(\log N + \log \frac{\mathbf{1}'X_t}{N}\right) = \mathbf{E}\left(\log \mathbf{1}'X_t\right).$$

The maximum of two nonnegative and integrable random variables is also integrable. Thus, we conclude that $\mathbf{E}(g(X_t)) < \infty$, i.e., the dominating function g is integrable.

At the beginning of Section 2 it has already been mentioned that each statement that refers to a random quantity is meant to be true with probability 1. The next theorem asserts that the BCRP is strongly consistent for the LOP. This means that w_n^* converges *almost surely* to w^*, which is simply denoted by "$w_n^* \to w^*$," i.e., without the additional remark "a.s.," for convenience.

Theorem 4. $w_n^* \to w^*$

Proof: The BCRP w_n^* represents an M-estimator, whose criterion functions are given by M and M_n. Let ε be any positive real number and $\mathcal{P}_\varepsilon := \{w \in \mathcal{S} : \|w - w^*\| = \varepsilon\}$.[7] Since M is strictly concave, there exists some $\delta > 0$ such that $M(w^*) - M(w) > \delta$ for all $w \in \mathcal{P}_\varepsilon$. Now, since $\mathcal{F}$ is Glivenko-Cantelli, we can find a sufficiently large number $m \in \mathbb{N}$ such that, for all natural numbers $n \geq m$, $|M_n(w^*) - M(w^*)| \leq \delta/2$ and $|M_n(w) - M(w)| \leq \delta/2$ for all $w \in \mathcal{P}_\varepsilon$. Thus, $M_n(w) < M_n(w^*)$ for all $w \in \mathcal{P}_\varepsilon$. Since M_n is strictly concave, too, we have that $\|w_n^* - w^*\| < \varepsilon$ for all $n \geq m$. This holds true for every $\varepsilon > 0$ and thus $w_n^* \to w^*$. Q.E.D.

The next theorem asserts that the expected out-of-sample log-return on the BCRP converges to the expected log-return on the LOP.

Theorem 5. $\varphi_{n+1}(w_n^*) \to \varphi(w^*)$

Proof: Theorem 4 and the Continuous Mapping Theorem reveal that $\log w_n^{*\prime}x \to \log w^{*\prime}x$ for all $x > 0$. Further, we already know that there exists an integrable function $x \mapsto g(x)$ such that $|f_w(x)| \leq g(x)$ for all $w \in \mathcal{S}$ and $x > 0$. Hence, by the Dominated Convergence Theorem, we obtain

$$\varphi_{n+1}(w_n^*) = \mathbf{E}\left(\log w_n^{*\prime}X_{n+1}\right) \to \mathbf{E}\left(\log w^{*\prime}X_{n+1}\right) = \mathbf{E}\left(\log w^{*\prime}X\right) = \varphi(w^*).$$

Q.E.D.

Finally, also the in-sample average log-return on the BCRP converges to the expected log-return on the LOP as the number of observations grows to infinity.

Theorem 6. $\frac{1}{n}\sum_{t=1}^{n} \log w_n^{*\prime}X_t \to \varphi(w^*)$

Proof: The statement is equivalent to $|M_n(w_n^*) - M(w^*)| \to 0$. Thus, it suffices to demonstrate that

$$\left|M_n(w_n^*) - M(w_n^*)\right| \to 0 \qquad \text{and} \qquad \left|M(w_n^*) - M(w^*)\right| \to 0.$$

The former is an immediate consequence of **A7**. Moreover, the Dominated Convergence Theorem tells us that $M(w_n) \to M(w)$ for every sequence $\{w_n\}$ with $w_n \in \mathcal{S}$ such that $w_n \to w \in \mathcal{S}$. This

---

[7]Here, "$\mathcal{P}_\varepsilon$" stands for "ε-periphery." Note that it contains only those w with distance ε to w^* that belong to $\mathcal{S}$.

means that M is continuous at each $w \in S$. Theorem 4 and the Continuous Mapping Theorem complete the proof. Q.E.D.

4.2.2. Asymptotic Distribution

In this section, the asymptotic distribution of $\sqrt{n}\left(w_n^* - w^*\right)$ is established. This can be done for all dimensions of w^* that are not bounded by S, i.e., $\mathbf{E}\left(X_{it}/w^{*\prime}X_t\right) = 1$. As already explained at the end of Section 3, each other component of w^* is bounded by the simplex. If $w_i^* = 0$ represents such a component, i.e., $\mathbf{E}\left(X_{it}/w^{*\prime}X_t\right) < 1$, it is well-known that

$$\sqrt{n}\left(w_{in}^* - w_i^*\right) = \sqrt{n}\, w_{in}^* \xrightarrow{P} 0,$$

i.e., w_{in}^* is superconsistent. However, not all components of the LOP can be affected by the given constraints on the portfolio weights. Indeed, we must have that $\mathbf{E}\left(X_{it}/w^{*\prime}X_t\right) = 1$ for at least one asset because otherwise the KKT conditions given by Theorem 1 cannot be satisfied. Thus, we can reduce the asset universe until there is no portfolio weight that is bounded by S. The riskless asset need not be part of the reduced asset universe. However, in order to avoid the trivial solution $w_n^* = 1$, there should be at least two remaining assets in the universe.

Hence, we assume that the given asset universe has been reduced such that $\mathbf{E}\left(X_t/w^{*\prime}X_t\right) = \mathbf{1}$. This guarantees that

$$(w - w^*)'\nabla M(w^*) = (w - w^*)'\mathbf{E}\left(\frac{X}{w^{*\prime}X}\right) = 0.$$

This means that the function M can be locally approximated at w^* by

$$M(w) = M(w^*) + \frac{1}{2}(w - w^*)'\nabla^2 M(w^*)(w - w^*) + o\left(\|w - w^*\|^2\right).$$

From the Monotone Convergence Theorem we conclude that the Hessian is given by

$$\nabla^2 M(w^*) = -\mathbf{E}\left(\frac{XX'}{(w^{*\prime}X)^2}\right).$$

The following assumption implies that the Hessian is finite. Finally, **A3** guarantees that $\nabla^2 M(w^*)$ is negative definite.

A8. The second moments of $X_t/(w^{*\prime}X_t)$ are finite.

Further, we have to make the following assumptions:

A9. The function f_w can be locally approximated at w^* by

$$f_w(X_t) = f_{w^*}(X_t) + (w - w^*)'\left(\frac{X_t}{w^{*\prime}X_t}\right) + \|w - w^*\|\, r(X_t; w),$$

where the process $\{r(X_t; w)\}$ is stochastically equicontinuous. This means that for every $\epsilon > 0$ and $\eta > 0$ there exists a neighborhood $\mathcal{U}$ of w^* in the simplex $\mathcal{S}$ such that

$$\limsup_{n \to \infty} \mathbb{P}^* \left(\sup_{w \in \mathcal{U}} \left| \sqrt{n} \left(\frac{1}{n} \sum_{t=1}^{n} r(X_t; w) - \mathbf{E}\big(r(X; w)\big) \right) \right| > \eta \right) < \epsilon,$$

where $\mathbb{P}^*$ is an outer measure associated with $\mathbb{P}$.

A10. We have that

$$\sqrt{n} \left(\frac{1}{n} \sum_{t=1}^{n} \frac{X_t}{w^{*\prime} X_t} - \mathbf{1} \right) \rightsquigarrow \mathcal{N}(\mathbf{0}, A).\text{[8]}$$

A9 is a basic regularity condition, which guarantees that the remainder $r(X_t, w)$ of the linear approximation becomes negligible as $n \to \infty$. To be more precise, it requires that

$$\sqrt{n}\big(M_n(w) - M(w)\big) \approx \sqrt{n}\big(M_n(w^*) - M(w^*)\big) + \sqrt{n}\,(w - w^*)' \left(\frac{1}{n} \sum_{t=1}^{n} \frac{X_t}{w^{*\prime} X_t} - \mathbf{1} \right)$$

if the sample size, n, is large and w is close to w^*.[9] Further, **A10** says that the process $\{X_t / w^{*\prime} X_t\}$ satisfies the Central Limit Theorem. In particular, if the elements of $\{X_t\}$ are serially independent, we obtain the asymptotic covariance matrix

$$A = \mathbf{Var}\left(\frac{X}{w^{*\prime} X} \right) = \mathbf{E}\left(\frac{XX'}{(w^{*\prime} X)^2} \right) - \mathbf{1}\mathbf{1}'.$$

Nonetheless, we could take also any form of serial dependence into account, provided the Central Limit Theorem expressed by **A10** is satisfied. There exist many strong mixing conditions that guarantee that this theorem holds true for the process $\{X_t / w^{*\prime} X_t\}$ (see, e.g., Bradley, 2005).

Suppose that $\Theta \subseteq \mathbb{R}^d$ is any parameter set and let $\theta \in \Theta$ be the "true" parameter. The tangent cone at θ is the set that we obtain after centering Θ at θ, blowing it up by some factor $\tau > 0$, and taking the set limit for $\tau \to \infty$ (Geyer, 1994, p. 1993). In order to study the asymptotic behavior of a sequence $\{\theta_n\}$ of global optimizers that converges to θ it is crucial to guarantee that the parameter set Θ is Chernoff regular (Geyer, 1994), viz.

$$\liminf_{\tau \to \infty} \tau(\Theta - \theta) = \limsup_{\tau \to \infty} \tau(\Theta - \theta) =: \lim_{\tau \to \infty} \tau(\Theta - \theta).$$

In our context, the parameter θ corresponds to $w^* \in \mathcal{S}$, which represents the global solution of the convex optimization problem expressed by Definition 1. The simplex $\mathcal{S}$ is Chernoff regular and so let $\mathcal{T}_{\mathcal{S}}(w^*) := \lim_{\tau \to \infty} \tau(\mathcal{S} - w^*)$ be the tangent cone of the simplex at w^*.[10]

---

[8]Here, "$\rightsquigarrow$" denotes convergence in distribution.

[9]Remember that $\mathbf{E}(X_t / w^{*\prime} X_t) = \mathbf{1}$ holds true by construction.

[10]We could imagine $w^* \in \mathcal{S}$ seeing through a microscope and increasing by and by the magnification. The visible part of $\mathcal{S}$ around w^* converges to $\mathcal{T}_{\mathcal{S}}(w^*)$, i.e., $\tau(\mathcal{S} - w^*) \to \mathcal{T}_{\mathcal{S}}(w^*)$, in the Painlevé-Kuratowski sense.

Consider any random vector $Y \sim \mathcal{N}(\mathbf{0}, A)$ and define the function

$$\zeta \mapsto \Psi_Y(\zeta) := \zeta'Y - \frac{1}{2}\zeta'\mathbf{E}\left(\frac{XX'}{(w^{*\prime}X)^2}\right)\zeta, \qquad \zeta \in \mathbb{R}^{N+1}.$$

The (unique) maximizer of Ψ_Y is denoted by

$$\zeta^* := \arg\max_{\zeta \in \mathcal{T}_{\mathcal{S}}(w^*)} \Psi_Y(\zeta).$$

The following theorem describes the asymptotic behavior of the BCRP.

Theorem 7. *We have that*

$$\sqrt{n}\left(w_n^* - w^*\right) \rightsquigarrow \arg\max_{\zeta \in \mathcal{T}_{\mathcal{S}}(w^*)} \zeta'\mathcal{N}(\mathbf{0}, A) - \frac{1}{2}\zeta'\mathbf{E}\left(\frac{XX'}{(w^{*\prime}X)^2}\right)\zeta.$$

Proof: The theorem asserts that $\sqrt{n}\left(w_n^* - w^*\right) \rightsquigarrow \zeta^*$, which is an immediate consequence of Theorem 4.4 in Geyer (1994).   Q.E.D.

Hence, if the sample size is large, $\sqrt{n}\left(w_n^* - w^*\right)$ behaves essentially like the solution of a relatively simple quadratic optimization problem. In the case in which the elements of $\{X_t\}$ are serially independent, we obtain

$$\sqrt{n}\left(w_n^* - w^*\right) \rightsquigarrow \arg\max_{\zeta \in \mathcal{T}_{\mathcal{S}}(w^*)} \zeta'\mathcal{N}\left(\mathbf{0}, \mathbf{Var}\left(\frac{X}{w^{*\prime}X}\right)\right) - \frac{1}{2}\zeta'\mathbf{E}\left(\frac{XX'}{(w^{*\prime}X)^2}\right)\zeta.$$

The following corollary establishes the long-run distribution of the log-return on the BCRP relative to the log-return on the LOP.

Corollary 1.
$$\log\frac{V_{w_n^*n}}{V_{w^*n}} \rightsquigarrow \zeta^{*\prime}\mathcal{N}(\mathbf{0}, A) - \frac{1}{2}\zeta^{*\prime}\mathbf{E}\left(\frac{XX'}{(w^{*\prime}X)^2}\right)\zeta^*.$$

Proof: Note that

$$M_n(w_n^*) = \frac{1}{n}\sum_{t=1}^{n}\log w_n^{*\prime}X_t \qquad \text{and} \qquad M_n(w^*) = \frac{1}{n}\sum_{t=1}^{n}\log w^{*\prime}X_t,$$

i.e.,

$$n\left(M_n(w_n^*) - M_n(w^*)\right) = \sum_{t=1}^{n}\log\frac{w_n^{*\prime}X_t}{w^{*\prime}X_t} = \log\frac{V_{w_n^*n}}{V_{w^*n}}.$$

The rest of the proof follows from Theorem 4.4 in Geyer (1994).   Q.E.D.

This completes our analysis of the BCRP. In the next section we focus on the MVE and derive its corresponding statistical properties.

5. The Mean-Variance Estimator

Consider some portfolio $w \in \mathcal{S}$ and let $\tilde{w} = (w_1, w_2, \ldots, w_N)$ be the "risky part" of that portfolio. The return on Asset i after Day t is given by $R_{it} = X_{it} - 1$ and so the return on w amounts to $R_{wt} = \tilde{w}' R_t$, where $R_t = (R_1, R_2, \ldots, R_N)$ denotes the vector of risky asset returns.[11] The assumptions **A1** to **A10** shall still hold true. Now, we make the following additional assumption:

B1. The second moments of R_t are finite.

Let R be any random vector that has the same distribution as $R_1, R_2, \ldots$. Define $\mu := \mathbf{E}(R)$ and $\Sigma := \mathbf{E}(RR')$. Note that the matrix Σ contains the second *noncentral* moments of the risky asset returns and thus it is not the covariance matrix of R. We already know that **A3** guarantees that there cannot be any $\tilde{w} \in \mathbb{R}^N$ with $\tilde{w} \neq \mathbf{0}$ such that $\tilde{w}' R = c \in \mathbb{R}$, i.e., Σ is positive definite.

Now, we may apply the quadratic approximation $\log(1 + r) \approx r - \frac{1}{2} r^2$ and come to the conclusion that

$$\mathbf{E}\big(\log(1 + R_{wt}) \big) \approx \mathbf{E}\left(\tilde{w}' R_t - \frac{1}{2} (\tilde{w}' R_t)^2 \right) = \tilde{w}' \mu - \frac{1}{2} \tilde{w}' \Sigma \tilde{w}. \tag{2}$$

This section is build upon the observation that this approximation is very good in most practical applications. Hence, instead of maximizing the expected log-return, we can simply maximize the objective function $w \mapsto \tilde{w}' \mu - \frac{1}{2} \tilde{w}' \Sigma \tilde{w}$.[12] In the following, this objective function is called "mean-variance" although Σ contains the second noncentral moments of R and so it does not coincide with the covariance matrix $\mathbf{Var}(R)$.

Definition 3. *A mean-variance optimal portfolio is a portfolio $w^\star \in \mathcal{S}$ that maximizes the mean-variance objective function, i.e.,*

$$w^\star \in \arg\max_{w \in \mathcal{S}} \tilde{w}' \mu - \frac{1}{2} \tilde{w}' \Sigma \tilde{w}.$$

Some important remarks may be appropriate at this point:

- The vector $w^\star$ is called mean-variance optimal, although Σ is not a covariance matrix. However, in most practical applications, Σ is close to $\mathbf{Var}(R)$ whenever R is a vector of daily asset returns.

- We focus on the feasible set $\mathcal{S}$ only because $w^\star$ serves as an approximation of the LOP. However, in general a mean-variance optimal portfolio need not be restricted to $\mathcal{S}$.

- Under general (but quite technical) regularity conditions, the MVOP can be considered an approximation of the GOP (Karatzas and Kardaras, 2007). Nonetheless, due to the reasons explained in Section 3, we should refrain from calling $w^\star$ "GOP."

Now, the Lagrange function of the optimization problem expressed by Definition 3 is

$$\mathcal{L}(w, \kappa, \lambda) = -\tilde{w}' \mu + \frac{1}{2} \tilde{w}' \Sigma \tilde{w} - \kappa' w + \lambda(\mathbf{1}' w - 1).$$

---

[11]Remember that the risk-free interest rate is supposed to be zero without loss of generality.
[12]Note that the domain of the objective function is $\mathcal{S}$ but its value is determined only by the "risky part" of w, i.e., $\tilde{w}$.

The following theorem is analogous to Theorem 2.

Theorem 8. *The MVOP exists and is unique. It is characterized by $w^\star \in \mathcal{S}$ such that the ith component of $\mu - \Sigma \tilde{w}^\star$ is*

$$\begin{cases} = \lambda, & w_i^\star > 0 \\ \leq \lambda, & w_i^\star = 0 \end{cases}, \qquad \lambda \geq 0.$$

Proof: The objective function

$$\tilde{w} \mapsto \tilde{w}'\mu - \frac{1}{2}\tilde{w}'\Sigma\tilde{w}$$

is strictly concave and the given set of constraints on the portfolio weights $w_1, w_2, \ldots, w_N$, i.e., $\tilde{w} \geq \mathbf{0}$ and $\mathbf{1}'\tilde{w} \leq 1$, is closed and convex. Hence, the "risky part" of $w^\star$, i.e., $\tilde{w}^\star$, exists and is unique, which means that $w^\star$ exists and is unique, too. Thus, we must have that

$$\begin{bmatrix} 0 \\ \mu - \Sigma\tilde{w}^\star \end{bmatrix} = \lambda\mathbf{1} - \kappa$$

with $w^\star \in \mathcal{S}$, $\lambda \in \mathbb{R}$, $\kappa = (\kappa_0, \kappa_1, \ldots, \kappa_N) \geq \mathbf{0}$, and $w_i^\star \kappa_i = 0$ for $i = 0, 1, \ldots, N$. It follows that $\lambda = \kappa_0 \geq 0$. Q.E.D.

The next corollary shows how to identify the components of $w^\star$ that are bounded by $\mathcal{S}$. This will be helpful later on.

Corollary 2. *The number λ in Theorem 8 is uniquely determined by $\lambda = \tilde{w}^{\star\prime}(\mu - \Sigma\tilde{w}^\star)$. Moreover, the portfolio weight*

- *$w_0^\star$ is bounded by $\mathcal{S}$ if and only if $\lambda > 0$, whereas*

- *$w_i^\star$ is bounded by $\mathcal{S}$ if and only if the ith component of $\mu - \Sigma\tilde{w}^\star$ is lower than λ.*

Proof: The proof of Theorem 8 reveals that $\tilde{w}^{\star\prime}(\mu - \Sigma\tilde{w}^\star) = \lambda = \kappa_0$. Since $w^\star$ is unique, the same holds true for λ. Moreover, $w_0^\star$ is bounded by $\mathcal{S}$ if and only if $\kappa_0 > 0$, i.e., $\lambda > 0$, whereas $w_i^\star$ is bounded by $\mathcal{S}$ if and only if $\kappa_i > 0$, i.e., the ith component of $\mu - \Sigma\tilde{w}^\star$ is below λ. Q.E.D.

In the following, let

$$\mu_n := \frac{1}{n}\sum_{t=1}^{n} R_t \qquad \text{and} \qquad \Sigma_n := \frac{1}{n}\sum_{t=1}^{n} R_t R_t'$$

be the moment estimators for μ and Σ. Now, we are ready to define the MVE for $w^\star$, which serves also as an estimator for the LOP w^*.

Definition 4. *A mean-variance estimator for $w^\star$ is a portfolio $w_n^\star \in \mathcal{S}$ that maximizes the in-sample mean-variance objective function, i.e.,*

$$w_n^\star \in \arg\max_{w \in \mathcal{S}} \tilde{w}'\mu_n - \frac{1}{2}\tilde{w}'\Sigma_n\tilde{w}.$$

5.1. Small-Sample Properties

5.1.1. Existence and Uniqueness

Let $\mathbf{R} = \begin{bmatrix} R_1 & R_2 & \cdots & R_n \end{bmatrix}$ be the sample of risky asset returns. **A4** implies that we cannot find any $\tilde{w} \in \mathbb{R}^N$ with $\tilde{w} \neq \mathbf{0}$ such that $\mathbf{R}'\tilde{w} = \mathbf{0}$. Hence, we have that

$$\tilde{w}'\Sigma_n\tilde{w} = \tilde{w}'\left(\frac{1}{n}\sum_{t=1}^n R_t R_t'\right)\tilde{w} = \frac{\tilde{w}'\mathbf{R}\mathbf{R}'\tilde{w}}{n} > 0$$

for all $\tilde{w} \in \mathbb{R}^N$ with $\tilde{w} \neq \mathbf{0}$, which means that Σ_n is positive definite.

The following corollary is a straightforward consequence of Theorem 8 and thus its proof can be skipped.

Corollary 3. *The MVE exists and is unique. It is characterized by $w_n^\star \in \mathcal{S}$ such that the ith component of $\mu_n - \Sigma_n\tilde{w}_n^\star$ is*

$$\begin{cases} = \lambda_n, & w_{in}^\star > 0 \\ \leq \lambda_n, & w_{in}^\star = 0 \end{cases}, \qquad \lambda_n \geq 0\,.$$

Numerical procedures for solving quadratic optimization problems exist in abundance and so it is easy to compute $w_n^\star$ even if the number of dimensions is high. Two points, which are discussed in more detail in Section 6, are worth emphasizing:

(i) The estimates w_{in}^* and $w_{in}^\star$ are indistinguishable in most real-life situations.[13] Put another way, the MVE leads to a very good approximation of the BCRP.

(ii) Cover's algorithm (1984) for w_n^* is slow compared to quadratic optimization algorithms for $w_n^\star$. In particular, this holds true in the high-dimensional case.

5.1.2. Finite-Sample Bias

Let w_n be any portfolio that is constructed on the basis of the asset returns $R_1, R_2, \ldots, R_n$. We know that the quantity $\tilde{w}_n' R_{n+1} - \frac{1}{2}\left(\tilde{w}_n' R_{n+1}\right)^2$ approximates the out-of-sample log-return on w_n and thus we call

$$\mathbf{E}\left(\tilde{w}_n' R_{n+1} - \frac{1}{2}\left(\tilde{w}_n' R_{n+1}\right)^2\right)$$

the expected out-of-sample performance of w_n. As already mentioned before, it is reasonable to presume that w_n is stochastically independent of R_{n+1}. Otherwise, the investment decision at time t would depend on some asset returns that occur one day later, which is usually considered implausible in finance theory. Thus, we obtain the conditional expectation

$$\begin{aligned}
\mathbf{E}\left(\tilde{w}_n' R_{n+1} - \frac{1}{2}\left(\tilde{w}_n' R_{n+1}\right)^2 \mid w_n\right) &= \tilde{w}_n'\underbrace{\mathbf{E}\left(R_{n+1} \mid w_n\right)}_{=\mu} - \frac{1}{2}\tilde{w}_n'\underbrace{\mathbf{E}\left(R_{n+1} R_{n+1}' \mid w_n\right)}_{=\Sigma}\tilde{w}_n \\
&= \tilde{w}_n'\mu - \frac{1}{2}\tilde{w}_n'\Sigma\tilde{w}_n,
\end{aligned}$$

---

[13]When using daily asset returns, the portfolio weights typically differ only from the fourth digit.

which can be viewed as the out-of-sample performance of w_n. Correspondingly, due to the Law of Total Expectation, its expected out-of-sample performance is

$$\phi_{n+1}(w_n) := \mathbf{E}\left(\tilde{w}_n' R_{n+1} - \frac{1}{2}\left(\tilde{w}_n' R_{n+1} \right)^2 \right) = \mathbf{E}\left(\tilde{w}_n' \mu - \frac{1}{2}\tilde{w}_n' \Sigma \tilde{w}_n \right).$$

The latter expectation is a basic performance measure in portfolio optimization (see, e.g., Frahm, 2015, Kan and Zhou, 2007, Markowitz and Usmen, 2003).[14] Hence, as already mentioned before, it is an implicit assumption of portfolio theory that w_n is stochastically independent of R_{n+1}.

If $w_n \equiv w$ is a fixed portfolio, we have that $\phi_{n+1}(w) = \tilde{w}'\mu - \frac{1}{2}\tilde{w}'\Sigma\tilde{w}$. In this case we may drop the prefix "expected out-of-sample" and just say that $\phi_{n+1}(w)$ is the performance of w. Further, then we can simply write $\phi(w)$ instead of $\phi_{n+1}(w)$. In particular,

$$\phi(w^\star) = \tilde{w}^{\star\prime}\mu - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma\tilde{w}^\star$$

represents the performance of the MVOP.

Hence, the following assumptions, which are analogous to **A5** and **A6**, are made:

B2. The MVE $w_n^\star$ is stochastically independent of R_{n+1}.

B3. The MVE does not coincide with the MVOP, i.e., $\mathbb{P}(w_n^\star = w^\star) \neq 1$.

Due to **B2** the expected out-of-sample performance of the MVE amounts to

$$\phi_{n+1}(w_n^\star) = \mathbf{E}\left(\tilde{w}_n^{\star\prime}\mu - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma\tilde{w}_n^\star \right).$$

Finally, $\tilde{w}_n'\mu_n - \frac{1}{2}\tilde{w}_n'\Sigma_n\tilde{w}_n$ represents the in-sample performance of the portfolio w_n and thus

$$\phi_n(w_n^\star) := \mathbf{E}\left(\tilde{w}_n^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma_n\tilde{w}_n^\star \right)$$

is the expected in-sample performance of the MVE.

The following theorem is similar to Theorem 3.

Theorem 9. $\phi_{n+1}(w_n^\star) < \phi(w^\star) < \phi_n(w_n^\star)$

Proof: By definition, $w^\star$ is the portfolio that maximizes the performance. Due to **B2** and **B3**, we conclude that

$$\phi_{n+1}(w_n^\star) = \mathbf{E}\left(\tilde{w}_n^{\star\prime}\mu - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma\tilde{w}_n^\star \right) < \tilde{w}^{\star\prime}\mu - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma\tilde{w}^\star = \phi(w^\star).$$

Moreover, since $w_n^\star$ is unique and does not coincide with $w^\star$, we have that

$$\mathbb{P}\left(\tilde{w}_n^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma_n\tilde{w}_n^\star \geq \tilde{w}^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma_n\tilde{w}^\star \right) = 1$$

---

[14] Some authors use the covariance matrix of R instead of $\Sigma = \mathbf{E}(RR')$.

and

$$\mathbb{P}\left(\tilde{w}_n^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma_n\tilde{w}_n^{\star} > \tilde{w}^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma_n\tilde{w}^{\star}\right) > 0\,,$$

which means that

$$\begin{aligned}
\phi_n(w_n^{\star}) &= \mathbf{E}\left(\tilde{w}_n^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma_n\tilde{w}_n^{\star}\right) > \mathbf{E}\left(\tilde{w}^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma_n\tilde{w}^{\star}\right) \\
&= \tilde{w}^{\star\prime}\mu - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma\,\tilde{w}^{\star} = \phi(w^{\star})\,.
\end{aligned}$$

Q.E.D.

Theorem 9 shows that we still suffer from the same problems that we have already found for the BCRP. This means that the in-sample performance of the MVE typically overestimates its expected out-of-sample performance and even the performance of the MVOP.

5.2. Large-Sample Properties

5.2.1. Consistency

The next assumption requires that $\{R_t\}$ and $\{R_t R_t'\}$ obey the Strong Law of Large Numbers. This holds true under very mild regularity conditions. If $R_1, R_2, \ldots$ are serially independent, **B1** is already sufficient. However, there exist much weaker mixing conditions, which guarantee that the Strong Law of Large Numbers is satisfied both for $\{R_t\}$ and for $\{R_t R_t'\}$. These mixing conditions are typically discussed in ergodic theory (see, e.g., Davidson, 1994).

B4. The estimators μ_n and Σ_n are strongly consistent for μ and Σ, i.e., $\mu_n \to \mu$ and $\Sigma_n \to \Sigma$.

Theorem 10. $w_n^{\star} \to w^{\star}$

Proof: Note that

$$w_n^{\star} = \arg\max_{w \in \mathcal{S}} \tilde{w}'\mu_n - \frac{1}{2}\tilde{w}'\Sigma_n\tilde{w}$$

represents a function of μ_n and Σ_n. Since $\mathcal{S}$ is convex, this function is continuous in μ_n and Σ_n. From **B4** and the Continuous Mapping Theorem it follows that $w_n^{\star} \to w^{\star}$.   Q.E.D.

The next theorem is analogous to Theorem 5.

Theorem 11. $\phi_{n+1}(w_n^{\star}) \to \phi(w^{\star})$

Proof: The objective function $w \mapsto \tilde{w}'\mu - \frac{1}{2}\tilde{w}'\Sigma\tilde{w}$ is continuous in $w \in \mathcal{S}$ and the set $\mathcal{S}$ is compact. From the Extreme Value Theorem we conclude that it has a minimum, a, and a maximum b. Hence, $w \mapsto \max\{|a|, |b|\}$ is a dominating function and it is clearly integrable. We already know that $w_n^{\star} \to w^{\star}$ and from the Dominated Convergence Theorem it follows that

$$\phi_{n+1}(w_n^{\star}) = \mathbf{E}\left(\tilde{w}_n^{\star\prime}\mu - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma\,\tilde{w}_n^{\star}\right) \to \tilde{w}^{\star\prime}\mu - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma\,\tilde{w}^{\star} = \phi(w^{\star})\,.$$

Q.E.D.

Moreover, analogous to Theorem 6, the Continuous Mapping Theorem immediately implies that

$$\tilde{w}_n^{\star\prime}\mu_n - \frac{1}{2}\tilde{w}_n^{\star\prime}\Sigma_n\tilde{w}_n^{\star} \to \tilde{w}^{\star\prime}\mu - \frac{1}{2}\tilde{w}^{\star\prime}\Sigma\,\tilde{w}^{\star} = \phi(w^{\star}),$$

i.e., the in-sample performance of the MVE converges to the performance of the MVOP.

5.2.2. Asymptotic Distribution

Now, the asymptotic distribution of $\sqrt{n}\,(w_n^{\star} - w^{\star})$ is derived. If some portfolio weight $w_i^{\star}$ is bounded by $\mathcal{S}$ it must be zero and the associated MVE is superconsistent, i.e., $\sqrt{n}\,w_{in}^{\star} \xrightarrow{\text{P}} 0$. Hence, in order to derive the asymptotic distribution of $\sqrt{n}\,(w_n^{\star} - w^{\star})$, we must guarantee that no component of the MVOP $w^{\star}$ is bounded by $\mathcal{S}$. According to Corollary 2, this holds true if and only if $\mu - \Sigma\,\tilde{w}^{\star} = \mathbf{0}$, i.e., $\tilde{w}^{\star} = \Sigma^{-1}\mu$. However, in practical situations it often happens that the weight of the riskless asset, $w_0^{\star}$, is bounded by $\mathcal{S}$, which means that the Lagrange multiplier λ in Theorem 8 is positive. In this case, we must abandon the riskless asset from our asset universe and focus on the risky assets. Then the MVOP is simply characterized by $\tilde{w}^{\star} \in \mathcal{S}$ such that the ith component of $\mu - \Sigma\,\tilde{w}^{\star}$ is

$$\begin{cases} = \lambda, & w_i^{\star} > 0 \\ \leq \lambda, & w_i^{\star} = 0 \end{cases}$$

with $\lambda > 0$. Thus, in the case in which the riskless asset has been removed, we assume that the remaining asset universe is such that $\mu - \Sigma\,\tilde{w}^{\star} = \lambda\mathbf{1}$ for *any* $\lambda > 0$.

Consider the family $\mathcal{F} = \{f_w\}_{w\in\mathcal{S}}$ with

$$r \mapsto f_w(r) = \tilde{w}'r - \frac{1}{2}\,(\tilde{w}'r)^2$$

for all $w \in \mathcal{S}$ and $r \in \mathbb{R}^N$. Further, define the functions

$$w \mapsto F(w) := \mathbf{E}\big(f_w(R)\big) = \tilde{w}'\mu - \frac{1}{2}\,\tilde{w}'\Sigma\,\tilde{w}$$

and

$$w \mapsto F_n(w) := \frac{1}{n}\sum_{t=1}^{n} f_w(R_t) = \tilde{w}'\mu_n - \frac{1}{2}\,\tilde{w}'\Sigma_n\tilde{w}\,.$$

It is obvious that the function F can be locally approximated at $w^{\star}$ by

$$F(w) = F(w^{\star}) - \frac{1}{2}\,(\tilde{w} - \tilde{w}^{\star})'\Sigma\,(\tilde{w} - \tilde{w}^{\star}),$$

where Σ is positive definite.

The next regularity conditions are analogous to **A9** and **A10**:

B5. The function f_w can be locally approximated at $w^{\star}$ by

$$f_w(R_t) = f_{w^{\star}}(R_t) + (\tilde{w} - \tilde{w}^{\star})'\big(R_t - R_t R_t'\tilde{w}^{\star}\big) + \|\tilde{w} - \tilde{w}^{\star}\|\,r(R_t;\tilde{w}),$$

where the process $\{r(R_t;\tilde{w})\}$ is stochastically equicontinuous.

B6. We have that

$$\sqrt{n}\left(\mu_n - \mu\right) - \sqrt{n}\left(\Sigma_n - \Sigma\right)\tilde{w}^\star \rightsquigarrow \mathcal{N}\left(\mathbf{0}, B\right).$$

Once again, **B5** guarantees that the remainder $r(X_t, w)$ of the linear approximation becomes negligible as $n \to \infty$. Further, **B6** requires the joint asymptotic normality of the given estimators for μ and Σ after the usual standardization. Since μ_n and Σ_n represent the moment estimators of μ and Σ, basically it states that $\{R_t - R_t R_t' \tilde{w}^\star\}$ should satisfy the Central Limit Theorem.[15]

The latter assumption indicates that we can decompose the estimation risk into two parts:

(i) $\sqrt{n}\left(\mu_n - \mu\right)$ represents the estimation risk that can be attributed to μ, whereas

(ii) $\sqrt{n}\left(\Sigma_n - \Sigma\right)\tilde{w}^\star$ stands for the estimation risk that is related to Σ.

Note that such a risk decomposition cannot be accomplished for the BCRP.

In some cases it is possible to calculate the asymptotic covariance matrix B in **B6**. For example, if $R_1, R_2, \ldots$ are serially independent and normally distributed, we have that

$$\begin{aligned}
B &= (1 - \tilde{w}^{\star\prime}\mu)^2\, \Gamma - (1 - \tilde{w}^{\star\prime}\mu)(\Gamma\,\tilde{w}^\star)\mu' - (1 - \tilde{w}^{\star\prime}\mu)\mu(\Gamma\,\tilde{w}^\star)' + \\
&\quad \left(\tilde{w}^{\star\prime}\Gamma\,\tilde{w}^\star\right)\left(\Gamma + \mu\mu'\right) + \left(\Gamma\,\tilde{w}^\star\right)\left(\Gamma\,\tilde{w}^\star\right)',
\end{aligned}$$

where $\Gamma = \Sigma - \mu\mu'$ denotes the covariance matrix of R.[16] More precisely, we can apply the decomposition $B = B_\mu + B_\Sigma$, where

$$B_\mu = \Gamma - \left[\left(\Gamma\,\tilde{w}^\star\right)\mu' + 2(\tilde{w}^{\star\prime}\mu)\Gamma + \mu\left(\Gamma\,\tilde{w}^\star\right)'\right]$$

quantifies the estimation risk that is associated with μ and

$$B_\Sigma = (\tilde{w}^{\star\prime}\mu)^2\,\Gamma + (\tilde{w}^{\star\prime}\mu)(\Gamma\,\tilde{w}^\star)\mu' + (\tilde{w}^{\star\prime}\mu)\mu(\Gamma\,\tilde{w}^\star)' + \left(\tilde{w}^{\star\prime}\Gamma\,\tilde{w}^\star\right)\left(\Gamma + \mu\mu'\right) + \left(\Gamma\,\tilde{w}^\star\right)\left(\Gamma\,\tilde{w}^\star\right)'$$

measures the estimation risk related to Σ. Similar results can be obtained if we assume that R has an elliptical distribution possessing heavy tails and tail dependence. Alternatively, we could apply a (block) bootstrap (see, e.g., Politis, 2003) in order to approximate B, or even B_μ and B_Σ, without making any parametric assumption.

Consider any random vector $Z \sim \mathcal{N}\left(\mathbf{0}, B\right)$. Now, we may define

$$\varsigma \mapsto \Phi_Z(\varsigma) := \tilde{\varsigma}'Z - \frac{1}{2}\,\tilde{\varsigma}'\Sigma\,\tilde{\varsigma}, \qquad \varsigma \in \mathbb{R}^{N+1},$$

with $\varsigma = (\varsigma_0, \varsigma_1, \ldots, \varsigma_N)$ and $\tilde{\varsigma} = (\varsigma_1, \varsigma_2, \ldots, \varsigma_N)$. The (unique) maximizer of Φ_Z is given by

$$\varsigma^\star = \arg\max_{\varsigma \in \mathcal{T}_S(w^\star)}\ \Phi_Z(\varsigma).$$

The following theorem clarifies the asymptotic behavior of the MVE. This result follows by the same arguments that were used for Theorem 7 and so the proof can be skipped.

---

[15]See also the explanations about the Central Limit Theorem regarding **A10**.
[16]The derivation of B can be found in the appendix.

Theorem 12. *We have that*

$$\sqrt{n}\left(w_n^\star - w^\star\right) \leadsto \underset{\varsigma \in \mathcal{T}_\mathcal{S}(w^\star)}{\arg\max} \; \varsigma' \mathcal{N}\left(\mathbf{0}, B\right) - \frac{1}{2}\,\varsigma' \Sigma\, \varsigma.$$

In the case in which the riskless asset has been removed from the asset universe, we may consider the (unique) maximizer

$$\tilde{\varsigma}^\star = \underset{\tilde{\varsigma} \in \mathcal{T}_\mathcal{S}(\tilde{w}^\star)}{\arg\max} \; \tilde{\varsigma}' Z - \frac{1}{2}\,\tilde{\varsigma}' \Sigma\, \tilde{\varsigma}$$

and then Theorem 12 reads

$$\sqrt{n}\left(\tilde{w}_n^\star - \tilde{w}^\star\right) \leadsto \underset{\tilde{\varsigma} \in \mathcal{T}_\mathcal{S}(\tilde{w}^\star)}{\arg\max} \; \tilde{\varsigma}' \mathcal{N}\left(\mathbf{0}, B\right) - \frac{1}{2}\,\tilde{\varsigma}' \Sigma\, \tilde{\varsigma}.$$

6. Some Practical Remarks

6.1. Computational Issues

Cover's (1984) algorithm for the BCRP is simple and works like this:

(i) Choose any initial portfolio $w^{(0)} \in \mathcal{S}$ and set $k \leftarrow 0$.

(ii) Update the portfolio according to

$$w^{(k+1)} = w^{(k)} \odot \frac{1}{n} \sum_{t=1}^{n} \frac{X_t}{w^{(k)\prime} X_t}$$

and set $k \leftarrow k + 1$.[17]

(iii) Repeat the second step until the largest component of the vector

$$\frac{1}{n} \sum_{t=1}^{n} \frac{X_t}{w^{(k)\prime} X_t}$$

falls below a critical threshold just above 1.

The computations made during this work are based on MATLAB. The critical threshold for the BCRP is $\exp 10^{-6}$. Further, the MOSEK optimization toolbox for MATLAB is used in order to compute the MVE, which proves to be very fast and reliable. It turns out that the BCRP, w_n^*, and the MVE, $w_n^\star$, are almost identical. However, computing $w_n^\star$ by quadratic optimization is much faster. In order to demonstrate these statements, we can simulate n independent and identically distributed vectors of daily asset returns $R_1, R_2, \ldots, R_n \sim \mathcal{N}\left(\mu, \Gamma\right)$ with

$$\mu = \frac{0.1}{250}\,\mathbf{1} \quad \text{and} \quad \Gamma = \frac{0.2^2}{250}\left(0.3\,\mathbf{1}\mathbf{1}' + 0.7\,\mathbf{I}_N\right). \tag{3}$$

---

[17]Here, "$\odot$" denotes the Hadamard, i.e., componentwise, matrix product.

Let us assume that the number of risky assets is $N = 100$ and the number of daily observations is $n = 250$. In this case, both w_0^* and $w_0^\star$ are bounded by $\mathcal{S}$, i.e., $w_0^* = w_0^\star = 0$. Thus, we abandon the riskless asset from the asset universe.

The numerical simulations are done 100 times. Each time Cover's algorithm for $\tilde{w}_n^*$ and the quadratic optimizer for $\tilde{w}_n^\star$ is applied. On average, Cover's algorithm needs 5.5914 seconds, whereas MOSEK takes only 0.0103 seconds.[18] The supremum norm of $\tilde{w}_n^* - \tilde{w}_n^\star$ is 0.0173. Although Cover's algorithm is much slower than the quadratic optimizer, the outcome of the latter turns out to be slightly better: The quadratic optimizer leads to an annualized average log-return of 0.4892, whereas Cover's algorithm yields only 0.4890 per year. That is, the quadratic optimizer comes even *closer* to the (true) BCRP than Cover's algorithm. In fact, the average log-returns produced by the quadratic optimizer are *always* better than those of Cover's algorithm. Hence, $\tilde{w}_n^\star$ dominates $\tilde{w}_n^*$ in a numerical sense. Moreover, Cover's algorithm is very slow in high dimensions, whereas the quadratic optimizer works well even for $N = 1000$ and $n = 2500$, in which case the computational time for $\tilde{w}_n^\star$ is still below 1 second.

There is another computational issue. For applying the asymptotic results derived in Section 4.2.2 we have to simulate the random vector $Y \sim \mathcal{N}(\mathbf{0}, A)$, where the covariance matrix A appears in **A10**. The problem is that A is singular. More precisely, we have that

$$\tilde{w}^{*\prime} A \, \tilde{w}^* = \tilde{w}^{*\prime} \mathbf{E}\left(\frac{XX'}{(\tilde{w}^{*\prime}X)^2} \right) \tilde{w}^* - \tilde{w}^{*\prime} \mathbf{1}\mathbf{1}'\tilde{w}^* = \mathbf{E}\left(\frac{(\tilde{w}^{*\prime}X)^2}{(\tilde{w}^{*\prime}X)^2} \right) - 1^2 = 0,$$

which means that A is not positive definite. Thus, we have to apply a matrix decomposition in order to simulate Y. This issue does not arise when applying the asymptotic results derived in Section 5.2.2, in which case we must simulate the random vector $Z \sim \mathcal{N}(\mathbf{0}, B)$. As already mentioned in Section 5.2.2, we can even provide a closed-form expression for B in many standard situations. The principal approach is demonstrated in the appendix.

To sum up, the quadratic approximation proposed at the beginning of Section 5 works very well and, in contrast to the BCRP, the MVE does not suffer from computational issues. For this reason, we focus on $\tilde{w}_n^\star$ in the following discussion.

6.2. Statistical Inference

Let us assume that the elements of $\{R_t\}$ are serially independent and normally distributed. To keep things as simple as possible, we may choose the parameterization in Eq. 3. Further, let the number of risky assets be $N = 2$ and the number of observations be $n = 250$.[19] Once again, we generate 100 samples and with each one we compute a realization of $\tilde{w}_n^\star$. On the upper left of Figure 1 we can see that most of the estimates are far away from $\tilde{w}^\star = (0.5, 0.5)$. The vast majority of the estimates are boundary solutions. More precisely, we have 50 estimates that equal $(0,1)$ and 41 that correspond to $(1,0)$. The given result does not improve, essentially, if we increase the number of observations to $n = 2500$ and it is still sobering even for 1 million observations. By contrast, if we assume that μ was known, the estimates turn out to be much

---

[18]The computations are done on a Windows Laptop with Intel Core i7-5500U CPU (2.4 GHz).
[19]The weight of the riskless asset is still bounded by $\mathcal{S}$ and thus $w_0^\star = 0$.
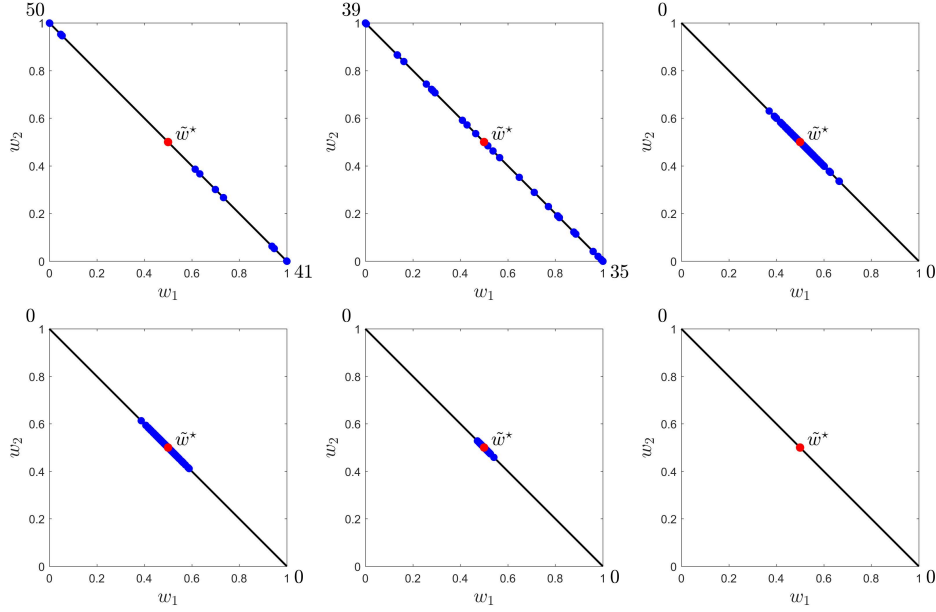
Figure 1: 100 realizations of $\tilde{w}_n^\star$ on the basis of 250, 2500, and 1 million daily observations (from left to right) if μ is unknown (upper part) and if it is known (lower part).

better (see the lower part of Figure 1). In particular, there is no more estimate at the boundary of the simplex, and in the case of $n = 10^6$ observations the estimates are almost identical with $\tilde{w}^\star$.

Are we able to replicate the finite-sample results by a large-sample approximation? For this purpose we could use Theorem 12 and the expressions for B_μ and B_Σ presented in Section 5.2.2. The corresponding realizations of the synthetic estimator $\tilde{w}_n^\infty := \tilde{w}^\star + \xi^\star/\sqrt{n}$ are depicted in Figure 2. The upper left of this figure indicates that there are 90 realizations *outside* the simplex. This is because the large-sample approximation is based on the maximizer $\xi^\star$, which belongs to the tangent cone of $\mathcal{S}$ at $\tilde{w}^\star$. Hence, the support of $\tilde{w}_n^\infty$ does not correspond to $\mathcal{S}$. Similarly, there are 82 realizations of $\tilde{w}_n^\infty$ missing in the simplex on the upper center. By contrast, the simplex on the upper right contains all 100 realizations of $\tilde{w}_n^\infty$. The picture changes essentially on the lower part of Figure 2, where it is assumed that μ is known. In this case, we cannot find any realization of $\tilde{w}_n^\infty$ outside $\mathcal{S}$. Moreover, the large-sample approximation satisfyingly reproduces the finite-sample results that are depicted on the lower part of Figure 1.

The problem is that the expected asset returns are unknown in real life. However, we can essentially improve the large-sample approximation by applying a finite-sample correction in order to guarantee that the realizations always belong to the simplex. We know from Theorem 12 that, if the sample size is large, $\sqrt{n}\,(\tilde{w}_n^\star - \tilde{w}^\star)$ behaves essentially like the maximizer, $\xi^\star$, of $\xi \mapsto \xi'Z - \frac{1}{2}\,\xi'\Sigma\,\xi$ over the tangent cone of $\mathcal{S}$ at $\tilde{w}^\star$. Hence, since the sample size is not large enough, we may substitute $\xi^\star$ with

$$\xi_n^\star := \underset{\xi \in \sqrt{n}\,(\mathcal{S}-\tilde{w}^\star)}{\arg\max}\ \xi'Z - \frac{1}{2}\,\xi'\Sigma\,\xi.$$

The constraint $\xi \in \sqrt{n}\,(\mathcal{S} - \tilde{w}^\star)$ can simply be implemented, numerically, by setting $\xi \geq -\sqrt{n}\,\tilde{w}^\star$ and $\mathbf{1}'\xi = 0$. The corrected version of $\tilde{w}_n^\infty$ reads $\tilde{w}_n^\diamond := \tilde{w}^\star + \xi_n^\star/\sqrt{n}$, which always belongs to
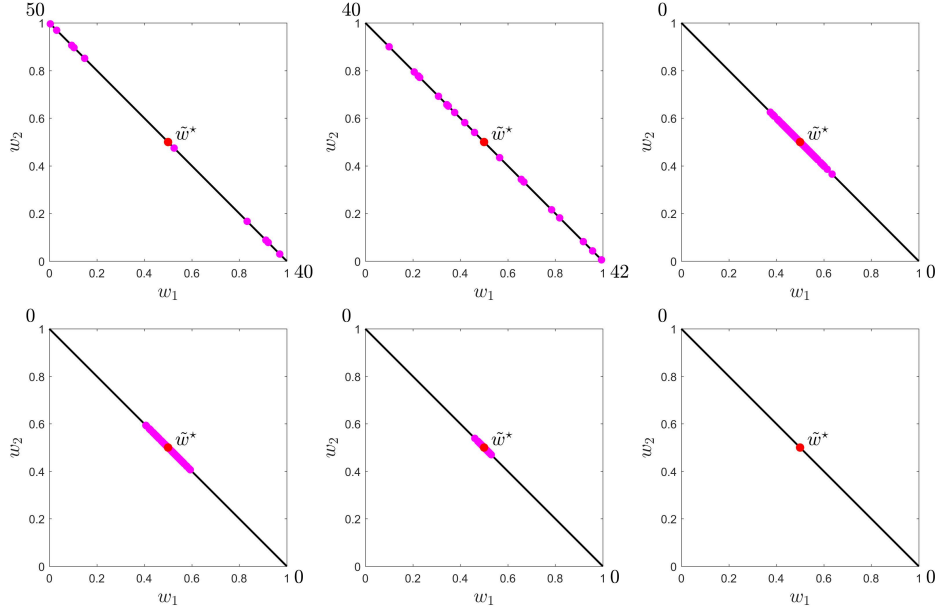
Figure 2: 100 realizations of $\tilde{w}_n^\infty$ on the basis of 250, 2500, and 1 million daily observations (from left to right) if μ is unknown (upper part) and if it is known (lower part).
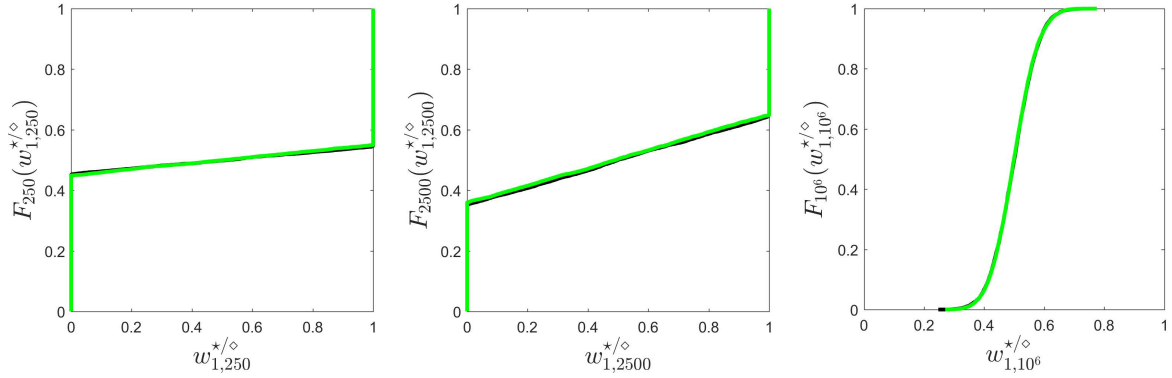

Figure 3: Empirical distribution functions of $w_{1n}^\star$ (black line) vs. $w_{1n}^\diamond$ (green line) for 250, 2500, and 1 million observations (from left to right).

the simplex.

In order to verify that the finite-sample correction works fine, we may compare the empirical distribution functions of 10000 realizations of $w_{1n}^\star$ and $w_{1n}^\diamond$, where μ is assumed to be unknown. We still have only $N = 2$ risky assets and the parameterization is the same as before (see Eq. 3). The results are given in Figure 3. Obviously, the finite-sample correction serves its purpose. Indeed, the corrected large-sample approximation is very accurate for all sample sizes.

Figure 3 reveals that most realizations of $\tilde{w}_n^\diamond$ are either $(0, 1)$ or $(1, 0)$ unless the sample size equals $n = 10^6$. The LOP corresponds to $\tilde{w}^\star = (0.5, 0.5)$ and thus it is precisely in between $(0, 1)$ and $(1, 0)$. It seems that estimating the LOP is a mission impossible in real-life situations—at least without any prior information about μ. Table 1 contains the probability that the realization of the MVE is a single-asset portfolio for different numbers of assets ($N = 5, 50, 100, 500, 1000$) and observations ($n = 250, 2500, 5000, 10000, 10^6$). The results are based on 1000 realizations

	n				
N	250	2500	5000	10000	10^6
5	0.817	0.527	0.337	0.216	0
50	0.698	0.264	0.135	0.055	0
100	0.652	0.224	0.121	0.046	0
500	—	0.196	0.081	0.030	0
1000	—	0.154	0.070	0.018	0

Table 1: Probability that $\tilde{w}_n^\star$ is a single-asset portfolio.

of $\tilde{w}_n^\diamond$ for each combination of N and n. Note that the LOP always corresponds to the equally weighted portfolio, i.e., $\tilde{w}^\star = \mathbf{1}/N$. The table shows that, in all practical applications, the MVE proposes a single-asset portfolio with high probability although the LOP is well-diversified. It is worth emphasizing that the results would not change essentially if we substitute the MVE with the BCRP, since these estimators for the LOP are almost identical.

Now, in principle, we are able to construct hypothesis tests and compute confidence regions. For example, we could try to apply a hypothesis test of the form $H_0 \colon \tilde{w}^\star = \tilde{w}_0^\star$ vs. $H_1 \colon \tilde{w}^\star \neq \tilde{w}_0^\star$ for any $\tilde{w}_0^\star \in \mathcal{S}$ even in the case of $N > 2$.[20] However, in the light of the previous results, we may doubt that any hypothesis test will ever lead to a rejection or that a confidence region will ever be sufficiently small in real-life situations. This conclusion might appear negative to the reader, but the author fears that this is the price we sometimes have to pay in science.

7. Conclusion

A quadratic approximation of log-returns works very well on a daily basis. Thus, in order to find the BCRP, we may focus on the MVE, which can easily be computed. The corresponding algorithm is very fast even if the number of dimensions is high and the results are even better compared to Cover's algorithm for the BCRP. However, in most practical applications, we typically overestimate the expected out-of-sample performance of the MVE and even the performance of the MVOP. The same holds true for the expected out-of-sample log-return on the BCRP and the expected log-return on the LOP.

Both the BCRP and the MVE exist and are unique under mild regularity conditions. Moreover, they are strongly consistent. Analogously, both their out-of-sample performance measures and their in-sample performances converge to the performance of the LOP or the MVOP, respectively, as the number of observations grows to infinity. The given estimators for the LOP are even $\sqrt{n}$-consistent. In principle, the asymptotic results derived in this work can be used for constructing hypothesis tests and for computing confidence regions, but for this purpose one should apply a finite-sample correction, which substantially improves the large-sample approximation.

However, it turns out that the impact of estimation risk concerning μ is tremendous in most

---

[20]Note that $\tilde{w}_0^\star$ is not the weight of the riskless asset but some portfolio of N risky assets.

real-life situations. Estimating the LOP without having any prediction power seems to be a futile undertaking. The estimators often lead to a single-asset portfolio even if the LOP corresponds to the equally weighted portfolio and thus is well-diversified. The given results confirm a general rule, which has become folklore during the last decades, namely that portfolio optimization typically fails on estimating expected asset returns.

A. The Asymptotic Covariance Matrix

Here, the asymptotic covariance matrix B, which occurs in Section 5.2.2, is derived. We assume that $R_1, R_2, \ldots$ are serially independent and normally distributed. Write

$$\sqrt{n}\left(\Sigma_n - \Sigma\right) = \sqrt{n}\left[\left(\Sigma_n - \mu\mu'\right) - \underbrace{\left(\Sigma - \mu\mu'\right)}_{=\Gamma}\right],$$

where Γ is the covariance matrix of R. The empirical covariance matrix of $R_1, R_2, \ldots, R_n$ is

$$\Gamma_n = \frac{1}{n}\sum_{t=1}^{n}\left(R_t - \mu_n\right)\left(R_t - \mu_n\right)' = \frac{1}{n}\sum_{t=1}^{n}R_t R_t' - \mu_n\mu_n' = \Sigma_n - \mu_n\mu_n'.$$

Thus, we obtain

$$\sqrt{n}\left(\Sigma_n - \Sigma\right) = \sqrt{n}\left(\Gamma_n - \Gamma\right) + \sqrt{n}\left(\mu_n\mu_n' - \mu\mu'\right).$$

Note that

$$\sqrt{n}\left(\mu_n\mu_n' - \mu\mu'\right) = \sqrt{n}\left(\mu_n - \mu\right)\left(\mu_n - \mu\right)' + \sqrt{n}\left(\mu_n\mu' + \mu\mu_n' - 2\mu\mu'\right),$$

where $\sqrt{n}\left(\mu_n - \mu\right)\left(\mu_n - \mu\right)'$ vanishes (in probability) as $n \to \infty$ and

$$\sqrt{n}\left(\mu_n\mu' + \mu\mu_n' - 2\mu\mu'\right) = \sqrt{n}\left(\mu_n - \mu\right)\mu' + \mu\sqrt{n}\left(\mu_n - \mu\right)'.$$

Thus, we conclude that

$$\sqrt{n}\left(\Sigma_n - \Sigma\right) = \sqrt{n}\left(\Gamma_n - \Gamma\right) + \sqrt{n}\left(\mu_n - \mu\right)\mu' + \mu\sqrt{n}\left(\mu_n - \mu\right)' + o_{\mathrm{p}}(1),$$

where $\sqrt{n}\left(\mu_n - \mu\right)$ and $\sqrt{n}\left(\Gamma_n - \Gamma\right)$ are asymptotically independent. Moreover, the given terms converge to a joint normal distribution. The asymptotic covariance matrix of $\sqrt{n}\,\mathrm{vec}\left(\Gamma_n - \Gamma\right)$ is $\left(\mathbf{I}_{N^2} + K_{N^2}\right)\left(\Gamma \otimes \Gamma\right)$, where the vec operator stacks the columns of a matrix on top of one another, $\mathbf{I}_{N^2}$ is the $N^2 \times N^2$ identity matrix, K_{N^2} is the $N^2 \times N^2$ commutation matrix, and "$\otimes$" denotes the Kronecker matrix product. According to Magnus and Neudecker (1979, Eq. 2.1) we have that

$$\sqrt{n}\left(\Gamma_n - \Gamma\right)\tilde{w}^{\star} = \left(\tilde{w}^{\star\prime} \otimes \mathbf{I}_N\right)\sqrt{n}\,\mathrm{vec}\left(\Gamma_n - \Gamma\right),$$

which means that

$$\left(\tilde{w}^{\star\prime} \otimes \mathbf{I}_N\right)\left(\mathbf{I}_{N^2} + K_{N^2}\right)\left(\Gamma \otimes \Gamma\right)\left(\tilde{w}^{\star} \otimes \mathbf{I}_N\right)$$

is the asymptotic covariance matrix of $\sqrt{n}\,(\Gamma_n - \Gamma)\tilde{w}^\star$. Due to Neudecker (1969, Eq. 2.2) we obtain

$$\left(\tilde{w}^{\star\prime} \otimes \mathbf{I}_N\right)\left(\Gamma \otimes \Gamma\right)\left(\tilde{w}^\star \otimes \mathbf{I}_N\right) = \left(\tilde{w}^{\star\prime}\Gamma \otimes \Gamma\right)\left(\tilde{w}^\star \otimes \mathbf{I}_N\right) = \left(\tilde{w}^{\star\prime}\Gamma\,\tilde{w}^\star\right)\Gamma$$

and from Theorem 3.1 in Magnus and Neudecker (1979) it follows that

$$
\begin{aligned}
\left(\tilde{w}^{\star\prime} \otimes \mathbf{I}_N\right)K_{N^2}\left(\Gamma \otimes \Gamma\right)\left(\tilde{w}^\star \otimes \mathbf{I}_N\right) &= \left(\mathbf{I}_N \otimes \tilde{w}^{\star\prime}\right)\left(\Gamma \otimes \Gamma\right)\left(\tilde{w}^\star \otimes \mathbf{I}_N\right) \\
&= \left(\Gamma \otimes \tilde{w}^{\star\prime}\Gamma\right)\left(\tilde{w}^\star \otimes \mathbf{I}_N\right) = \left(\Gamma\,\tilde{w}^\star\right)\left(\tilde{w}^{\star\prime}\Gamma\right).
\end{aligned}
$$

Hence, the asymptotic covariance matrix of $\sqrt{n}\,(\Gamma_n - \Gamma)\tilde{w}^\star$ is

$$\left(\tilde{w}^{\star\prime}\Gamma\,\tilde{w}^\star\right)\Gamma + \left(\Gamma\,\tilde{w}^\star\right)\left(\Gamma\,\tilde{w}^\star\right)'.$$

It remains to calculate the asymptotic covariance matrix of

$$\sqrt{n}\,(\mu_n - \mu)\mu'\tilde{w}^\star + \mu\sqrt{n}\,(\mu_n - \mu)'\tilde{w}^\star.$$

The asymptotic covariance matrix of $\sqrt{n}\,(\mu_n - \mu)$ is Γ, which leads to

$$(\tilde{w}^{\star\prime}\mu)^2\,\Gamma + (\tilde{w}^{\star\prime}\mu)(\Gamma\,\tilde{w}^\star)\mu' + (\tilde{w}^{\star\prime}\mu)\mu(\Gamma\,\tilde{w}^\star)' + \left(\tilde{w}^{\star\prime}\Gamma\,\tilde{w}^\star\right)\mu\mu'.$$

Thus, the asymptotic covariance matrix of $\sqrt{n}\,(\Sigma_n - \Sigma)$ is

$$B_\Sigma = (\tilde{w}^{\star\prime}\mu)^2\,\Gamma + (\tilde{w}^{\star\prime}\mu)(\Gamma\,\tilde{w}^\star)\mu' + (\tilde{w}^{\star\prime}\mu)\mu(\Gamma\,\tilde{w}^\star)' + \left(\tilde{w}^{\star\prime}\Gamma\,\tilde{w}^\star\right)\left(\Gamma + \mu\mu'\right) + \left(\Gamma\,\tilde{w}^\star\right)\left(\Gamma\,\tilde{w}^\star\right)',$$

which quantifies the estimation risk if the parameter μ was known to the investor. However, in real life the expected asset returns are unknown and so B equals the asymptotic covariance matrix of

$$\sqrt{n}\,(\mu_n - \mu) - \sqrt{n}\,(\Gamma_n - \Gamma)\tilde{w}^\star - \sqrt{n}\,(\mu_n - \mu)\mu'\tilde{w}^\star - \mu\sqrt{n}\,(\mu_n - \mu)'\tilde{w}^\star,$$

which can be rewritten as

$$\sqrt{n}\,(\mu_n - \mu)\left(1 - \tilde{w}^{\star\prime}\mu\right) - \mu\sqrt{n}\,(\mu_n - \mu)'\tilde{w}^\star - \sqrt{n}\,(\Gamma_n - \Gamma)\tilde{w}^\star.$$

By using the above arguments we conclude that

$$
\begin{aligned}
B &= (1 - \tilde{w}^{\star\prime}\mu)^2\,\Gamma - (1 - \tilde{w}^{\star\prime}\mu)(\Gamma\,\tilde{w}^\star)\mu' - (1 - \tilde{w}^{\star\prime}\mu)\mu(\Gamma\,\tilde{w}^\star)' + \\
&\quad \left(\tilde{w}^{\star\prime}\Gamma\,\tilde{w}^\star\right)\left(\Gamma + \mu\mu'\right) + \left(\Gamma\,\tilde{w}^\star\right)\left(\Gamma\,\tilde{w}^\star\right)'.
\end{aligned}
$$

Now, the reader can verify that the impact of estimating the expected asset returns is

$$B_\mu = \Gamma - \left[(\Gamma\,\tilde{w}^\star)\mu' + 2(\tilde{w}^{\star\prime}\mu)\Gamma + \mu\left(\Gamma\,\tilde{w}^\star\right)'\right].$$

References

Algoet, P., Cover, T. (1988): "Asymptotic optimality and asymptotic equipartition properties of log-optimum investment," *Annals of Probability* **16**, pp. 876–898.

Bell, R., Cover, T. (1980): "Competitive optimality of logarithmic investment," *Mathematics of Operations Research* **5**, pp. 161–166.

Bradley, R. (2005): "Basic properties of strong mixing conditions. A survey and some open questions," *Probability Surveys* **2**, pp. 107–144.

Breiman, L. (1961): "Optimal gambling systems for favorable games," in "Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability," pp. 63–68.

Christensen, M. (2005): "On the history of the growth optimal portfolio," Technical report, University of Southern Denmark.

Cover, T. (1984): "An algorithm for maximizing expected log investment return," *IEEE Transactions on Information Theory* **IT–30**, pp. 369–373.

Cover, T., Thomas, J. (1991): *Elements of Information Theory*, John Wiley & Sons.

Davidson, J. (1994): *Stochastic Limit Theory*, Oxford University Press.

Frahm, G. (2015): "A theoretical foundation of portfolio resampling," *Theory and Decision* **79**, pp. 107–132.

Frahm, G. (2016): "Pricing and valuation under the real-world measure," *International Journal of Theoretical and Applied Finance* **19**, DOI: 10.1142/S0219024916500060.

Frahm, G. (2018): "Arbitrage Pricing Theory in ergodic markets," *International Journal of Theoretical and Applied Finance* **21**, DOI: 10.1142/S021902491850036X.

Frahm, G., Memmel, C. (2010): "Dominating estimators for minimum variance portfolios," *Journal of Econometrics* **159**, pp. 289–302.

Geyer, C. (1994): "On the asymptotics of constraint M-estimation," *Annals of Statistics* **22**, pp. 1993–2010.

Hakansson, N. (1971): "Capital growth and the mean-variance approach to portfolio selection," *Journal of Financial and Quantitative Analysis* **6**, pp. 517–557.

Kan, R., Zhou, G. (2007): "Optimal portfolio choice with parameter uncertainty," *Journal of Financial and Quantitative Analysis* **42**, pp. 621–656.

Karatzas, I., Kardaras, C. (2007): "The numéraire portfolio in semimartingale financial models," *Finance and Stochastics* **11**, pp. 447–493.

Kelly, J. (1956): "A new interpretation of information rate," *The Bell System Technical Journal* **27**, pp. 379–423.

MacLean, L., Thorp, E., Ziemba, W. (editors) (2011): *The Kelly Capital Growth Investment Criterion*, World Scientific.

Magnus, J., Neudecker, H. (1979): "The commutation matrix: some properties and applications," *Annals of Statistics* **7**, pp. 381–394.

Markowitz, H., Usmen, N. (2003): "Resampled frontiers versus diffuse Bayes," *Journal of Investment Management* **1**, pp. 1–17.

Memmel, C. (2004): *Schätzrisiken in der Portfoliotheorie*, Ph.D. thesis, University of Cologne.

Merton, R., Samuelson, P. (1974): "Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods," *Journal of Financial Economics* **1**, pp. 67–94.

Neudecker, H. (1969): "Some theorems on matrix differentiation with special reference to Kronecker matrix products," *Journal of the American Statistical Association* **64**, pp. 953–963.

Platen, E., Heath, D. (2006): *A Benchmark Approach to Quantitative Finance*, Springer.

Politis, D. (2003): "The impact of bootstrap methods on time series analysis," *Statistical Science* **18**, pp. 219–230.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.