Consistency of robust portfolio estimators

Katrin Schöttle[*], Ralf Werner[†]

First draft: 1 December 2005
This version: 6 October 2006

Abstract

It is a matter of common knowledge that traditional Markowitz optimization based on sample means and covariances performs poorly in practice. For this reason, diverse attempts were made to improve performance of portfolio optimization. In this paper, we investigate three popular portfolio selection models built upon classical mean-variance theory. The first model is an extension of the traditional mean-variance optimization by introducing robust estimators. Second, the recently being en vogue robust counterpart approach is considered. The list of models is concluded by an extended version of Michaud's resampling approach. We show that for a very broad class of portfolio constraints these models can be seen as a generalization of the classical mean-variance setting: The optimal portfolios converge to the true optimal Markowitz portfolio if only the sample size is large enough.

Keywords: portfolio optimization, consistent estimators

AMS 2000 subject classification: 62H12, 62P05, 90C25, 90C31, 90C90

[*]Munich University of Technology, HVB-Institute for Mathematical Finance, Boltzmannstr. 3, 85748 Garching b. München, Germany, email: schoettle@ma.tum.de

[†]Hypo Real Estate Holding AG, Group Risk Control, Unsöldstr. 2, 80538 München, Germany, email: werner_ralf@gmx.net

1 Introduction

Traditionally, Markowitz portfolio optimization relies on historical data, or, to be more exact, on the sample estimators of the means and covariances. As such, optimal portfolios based on statistical estimators are in principle merely estimators for the true optimal portfolios. In this paper we show that several portfolio estimators used in practice are in fact consistent. Especially, we investigate the portfolio selection models used in Schöttle and Werner [26], i.e. Markowitz, robust and resampled portfolio estimators. In that recent empirical study the effectiveness of robustification for portfolio selection was demonstrated. The finite sample properties of the three different approaches to portfolio selection were investigated and it was shown that robustification is indeed able to add value to the portfolio selection process in asset management. The effect was made visible by examining stability of portfolio estimators both along the efficient frontier and over time.

In contrast to those empirical investigations based on finite samples, we concentrate here on the case when the sample size tends to infinity. As main result, we show in Theorems 3.7, 4.9 and 5.6 that each of the three portfolio estimators converges to the true optimal Markowitz portfolio if the sample size tends to infinity. In statistical terms this means that all considered portfolio estimators are *consistent*. This is based on two main facts: first, only consistent estimators for the first two moments are used, and second, the optimal portfolios depend continuously on the input data.

Concerning literature, there is only few work to directly relate to, as most publications concentrate on the finite sample case without considering the limit case of arbitrarily large samples. The seminal paper, Jobson and Korkie [13], dating back to 1980, provides characteristics and asymptotic properties of portfolio estimates. Jobson and Korkie derive asymptotic normality of the optimal portfolios under normality assumption on the return distribution. Their analysis relies on the fact that no constraints on the portfolios are considered and that an analytical formula for the optimal portfolio is available. To our knowledge the only other relevant work on asymptotic properties of portfolio estimators is Mori [20], who generalized the results of Jobson and Korkie to the case of general linear equality constraints. Concerning the introduction of robust estimators, there is recent work by Perret-Gentil and Victoria-Feser [23] who considered robust estimators in addition to the classical sample estimates. They show that for finite samples the stability of optimal portfolios can be significantly improved by using the robust estimators. A different approach towards robustification was investigated by Lauprête, Samarov and Welsch [16]. In contrast to this paper at hand, where the procedures of parameter estimation and portfolio optimization are strictly separated, Lauprête merged these two steps. For this merged setting consistency and asymptotic normality under elliptically distributed returns was proved. In their very recent working paper De Miguel and Nogales [5] proved robustification by looking at the influence function of the optimal portfolios in a setting very similar to Lauprête's. For more references on these one step procedures, we refer to this working paper [5]. For completeness, we want to mention the work by Kempf and Memmel [15] who concentrated on the estimation of the minimum variance portfolio by shrinkage estimators in a one step procedure. Alternative ways of robustification have also been considered by several authors, among those are for example Goldfarb and Iyengar [8], Tütüncü and Koenig [28], Ceria and Stubbs [4], Lutgens [17] as well as Schöttle and Werner [26]. All these approaches rely on applications of the robust counterpart methodology for standard optimization problems under uncertainty (see the seminal paper by Ben-Tal and Nemirovski [3]) with appropriately chosen uncertainty sets. Somewhat different in nature is the point of view taken by Okhrin and Schmid [22] who investigated distributional properties of optimal portfolio weights in specific settings. Their work can be seen supplementary to Jobson and Korkie, as they investigate the distribution of optimal portfolios in the finite sample case under otherwise similar conditions. Based on the same idea, i.e. the distribution of the optimal portfolio, the resampling approach has been introduced by Jorion [14] and Michaud [19], where the distributional aspects of the portfolios are considered to derive more robust portfolio estimators.

Accordingly, this paper makes three main contributions: First, Section 3 can be seen as a combination of the results by Jobson and Korkie with those of Perret-Gentil and Victoria-Feser. In that section we show that the assumptions by Jobson and Korkie can be significantly diminished: the normality assumption can be weakened to ellipticity and further, very general portfolio constraints can be imposed. Then consistency is not only given for the traditional sample estimates but as well for all consistent robust estimators considered by Perrit-Gentil and Victoria-Feser. Second, in Section 4, we introduce the notion of *consistent uncertainty sets*, close related to confidence sets. Using these consistent uncertainty sets, the robust counterpart approach is then also consistent with the traditional mean-variance. Finally, Section 5 contains the first results on the asymptotic properties of (an extended version of) Michaud's resampling procedure.

The paper is organized as follows: we start with a succinct review of some popular consistent estimators for the first and second moment of the return distribution. Subsequently, the three different portfolio selection problems are briefly described: the traditional Markowitz setting, robust portfolio optimization and resampled portfolios. For each of the three settings, consistency of the according portfolio estimator is verified.

2 Mathematical framework

In the following paragraphs we give a mathematical definition of the financial market under consideration, together with a detailed introduction of popular estimators. We will briefly review the main properties of these estimators which are needed in the next sections to establish the main consistency results.

2.1 Market setting

We consider a financial market with n risky assets and restrict ourselves to the single-period setting from time t_0 to t_1, as multi-period analysis is out of the scope of this paper. It is well-known, see Ingersoll [11], that in this setting the classical mean-variance theory is consistent with the concept of utility theory if the asset returns follow a multivariate elliptical distribution.

Assumption 2.1. *Let a suitable probability space $(\Omega, \mathcal{A}, \mathbf{P})$ be given, further let the asset return[1] $R \in \mathbb{R}^n$ be multivariate elliptically distributed with parameters μ, Σ and characteristic generator ψ, i.e. $R \sim \mathcal{E}(\mu, \Sigma, \psi)$. In addition, we assume that R possesses a density $\varphi_R = \varphi_{\mathcal{E}(\mu,\Sigma,\psi)}$ and that R has finite second moments.*

Remark 2.2. *According to Fang and Zhang [7], Theorem 2.5.4 we can assume that the characteristic generator ψ for R is differentiable. As we require R to have a density, it follows by Fang et al. [6], p. 46, that Σ has to be positive definite. For such an elliptically distributed return R it holds that*

$$\mathbf{E}[R] = \mu, \quad \mathbf{Cov}[R] = -2\psi'(0)\Sigma$$

and its density φ_R is given by

$$\varphi_R(r) = \varphi_{\mathcal{E}(\mu,\Sigma,\psi)}(r) = c_g|\Sigma|^{-\frac{1}{2}} g\Big((r-\mu)^T\Sigma^{-1}(r-\mu)\Big)$$

with density generator *g and appropriate normalizing constant c_g. Due to the scaling invariance (see Theorem 2.15(i) in Fang et al. [6]) we may assume for convenience that $\psi'(0) = -0.5$, i.e.*

$$\mathbf{Cov}[R] = \Sigma.$$

[1] We consider the linear one period return $R_i = \frac{S_{i,t_1} - S_{i,t_0}}{S_{i,t_0}}$.

Let us further assume that the elliptical distribution is chosen in such a way that its density generator g is continuous and decreasing. For more details on elliptical distributions, we refer to Fang and Zhang [7] or Fang et al. [6].

As the real distribution of R is not known, i.e. neither the characteristic generator ψ nor the parameters μ and Σ are given a priori, one has to rely on estimators to get information about these unknowns. As the generator ψ is not important for the concept of mean-variance optimization, but only μ and Σ are needed to derive the optimal portfolios, we solely concentrate on estimators for these two parameters. We assume that these estimators stem from a historical data sample:

Assumption 2.3. *In addition to Assumption 2.1, let S historical realizations $R_1, \dots, R_S$ of R be given. We assume that these historical samples R_s are i.i.d. with $R_s \sim \mathcal{E}(\mu, \Sigma, \psi), s = 1, \dots, S$.*

Assumptions 2.1 and 2.3 allow to model fat tails in the asset returns, e.g. by imposing a multivariate student-t distribution, but do not allow for any intertemporal dependencies as the historical realizations are assumed to be independently distributed.

2.2 Estimators for market parameters

In general, point estimators $Q_{p,S}$ for an unknown parameter p can be interpreted as measurable functions from the space of samples of length S to the appropriate parameter space, i.e. in our case $Q_{\mu,S} : (\mathbb{R}^n)^S \to \mathbb{R}^n$ and $Q_{\Sigma,S} : (\mathbb{R}^n)^S \to \mathbb{R}^{n \times n}$. More exact definitions of estimators may for example be found in Müller [21].

Because estimators are based on the random sample $R_1, \dots, R_S$, i.e. are a function thereof, they are random variables themselves. These random variables live on the same probability space $(\Omega, \mathcal{A}, \mathbf{P})$ as the original realizations R_s. In general, it is not possible to characterize the distribution of an arbitrary point estimator in analytical terms[2]. Although the distribution is not analytically available, the main properties of the estimators can still be derived. Among these main properties are biasedness, continuity, consistency and asymptotic normality. In our context, *consistency* is the most important characteristic. Consistency describes the behavior of a point estimator when the length S of the underlying sample tends toward infinity.

[2]For most elliptical distributions it is possible to explicitly state the joint density of the maximum likelihood estimators, see Fang and Zhang [7], Section 4.2.

Definition 2.4. *Let $Q_{p,S}$ denote a point estimator for the parameter p based on a sample of size S. This estimator $Q_{p,S}$ is called (strongly)* consistent, *if the sequence of point estimates converges almost surely to the true parameter p as the sample size tends to infinity, i.e. if*

$$\mathbf{P}(\lim_{S\to\infty} Q_{p,S} = p) = 1.$$

We say that the estimator $Q_{p,S}$ is asymptotically normally distributed with asymptotic covariance matrix K, if there is a random variable $Z \sim \mathcal{N}(0, K)$ such that

$$\sqrt{S}\big(Q_{p,S} - p\big) \to Z \quad \textit{for } S \to \infty$$

where convergence is understood in distribution.

Subsequently, we will introduce the classical estimator for the covariance matrix along with several popular estimators for the expected return. We will present results on consistency and asymptotic normality, where available.

Notation 2.5. *In the following, point estimators of μ and Σ based on a sample of length S will be abbreviated by $\hat{\mu}_S$ and $\hat{\Sigma}_S$.*

2.2.1 Sample / maximum likelihood estimators

Among the most popular estimators for the expected return and the covariance are the *sample estimators*. In the case of elliptical distributions it can be shown that the sample estimates coincide with the *maximum likelihood estimators*, if the density generator g is continuous and decreasing, see Fang and Zhang [7], Section 4.1. As we have assumed that this is the case, we will denote them as maximum likelihood estimators subsequently.

Definition 2.6. *The* maximum likelihood estimators *for μ and Σ based on a sample of length S are given by*

$$\hat{\mu}_S^{ML} := \frac{1}{S}\sum_{s=1}^{S} R_s,$$

$$\hat{\Sigma}_S^{ML} := \frac{1}{S}\sum_{s=1}^{S}(R_s - \hat{\mu}_S^{ML})(R_s - \hat{\mu}_S^{ML})^T.$$

In the special case of normally distributed returns, the joint distribution of the maximum likelihood estimators $\hat{\mu}_S^{ML}$ and $\hat{\Sigma}_S^{ML}$ for μ and Σ is analytically given as follows:

Proposition 2.7. *Under the assumption of* $R \sim \mathcal{N}(\mu, \Sigma)$*, i.e.* $R_s \sim \mathcal{N}(\mu, \Sigma)$ *i.i.d., the maximum likelihood estimators* $\hat{\mu}_S^{ML}$ *and* $\hat{\Sigma}_S^{ML}$ *are independently distributed as*

$$\hat{\mu}_S^{ML} \sim \mathcal{N}(\mu, \frac{1}{S}\Sigma), \quad \hat{\Sigma}_S^{ML} \sim \mathcal{W}(\frac{1}{S}\Sigma, S-1),$$

where $\mathcal{W}(Z, q)$ *denotes the Wishart distribution with scale matrix* $Z \in \mathbb{R}^{n \times n}$ *and* q *degrees of freedom.*

Proof. See Corollary 2 on page 135 in Fang and Zhang [7]. □

Notation 2.8. *We abbreviate the above distribution with* $\mathcal{NW}(\mu, \Sigma, S)$*, i.e.*

$$\mathcal{NW}(\mu, \Sigma, S) := \mathcal{N}(\mu, \frac{1}{S}\Sigma) \otimes \mathcal{W}(\frac{1}{S}\Sigma, S-1).$$

We will denote the density of this distribution by $\varphi_{\mathcal{NW}(\mu,\Sigma,S)}$*, i.e.*

$$\varphi_{\mathcal{NW}(\mu,\Sigma,S)}(r, C) = \varphi_{\mathcal{N}(\mu,\frac{1}{S}\Sigma)}(r) \cdot \varphi_{\mathcal{W}(\frac{1}{S}\Sigma,S-1)}(C)$$

where $\varphi_{\mathcal{N}(\mu,\frac{1}{S}\Sigma)}$ *denotes the density of the normal distribution and* $\varphi_{\mathcal{W}(\frac{1}{S}\Sigma,S-1)}$ *the density of the Wishart distribution, respectively.*

2.2.2 Quantile based estimators

In addition to the maximum likelihood estimators other well accepted estimators for the mean of an elliptical distribution are used in practice. One of the most common ones is the median estimator which can be seen as an estimator based on empirical quantiles. As elliptical distributions have symmetric marginal distributions, see Fang et al. [6], p. 37, the median equals the mean for these marginal distributions and thus is an appropriate estimator.

Definition 2.9. *Let* $y_1, \dots, y_S$ *denote an* $\mathbb{R}$*-valued random sample and let* $y_{(1)} \leq \dots \leq y_{(S)}$ *be the* ordered sample. *Then a point estimator for the theoretical* α*-quantile* q_α *(*$0 < \alpha < 1$*) of the according distribution is given by the* sample quantile

$$Q_{q_\alpha,S}^{QU} := y_{(\lfloor \alpha S \rfloor + 1)}.$$

The median estimator is now just a special case of the above quantile estimator:

Definition 2.10. *The* median estimator *or* sample median *for* μ *based on a sample of length* S *is defined componentwise by*

$$\hat{\mu}_{S,i}^{ME} := Q_{q_{0.5},S,i}^{QU} = R_{(\lfloor \frac{S}{2} \rfloor + 1),i}.$$

In the same manner, other symmetric quantile-based estimators can be used. A common alternative to the median estimator for symmetric distributions is based on its *quartiles*.

Definition 2.11. *The* quartile estimator *for μ based on a sample of length S is defined componentwise by*

$$\hat{\mu}^{QR}_{S,i} := \frac{1}{2} Q^{QU}_{q_{0.75},S,i} + \frac{1}{2} Q^{QU}_{q_{0.25},S,i} = \frac{1}{2} R_{(\lfloor 0.25 \cdot S \rfloor + 1),i} + \frac{1}{2} R_{(\lfloor 0.75 \cdot S \rfloor + 1),i}.$$

In general, when dealing with fat-tailed distributions, the median and the quartile estimator are more robust than the traditional maximum likelihood estimators, since their breakdown point[3] is larger than the one of the maximum likelihood estimators, see e.g. Huber [10], p. 54 and p. 60.

2.2.3 Further robust estimators

Two other commonly used robust estimators are given by the trimmed mean and the Huber estimator.

Definition 2.12. *Let $0 < \alpha < \frac{1}{2}$. The* α-trimmed mean estimator *for μ based on a sample of length S is defined componentwise by*

$$\hat{\mu}^{TM}_{S,i} := \frac{1}{S - 2\lfloor \alpha S \rfloor} \sum_{s=\lfloor \alpha S \rfloor + 1}^{S - \lfloor \alpha S \rfloor} R_{(s),i}.$$

The α-trimmed mean estimator thus ignores the $\lfloor \alpha S \rfloor$ smallest and largest values and calculates the average of the remaining sample.

Remark 2.13. *The maximum likelihood estimator, the median, quartile and trimmed mean estimator, they all can be expressed as linear combinations of an ordered sample $y_{(1)} \leq \cdots \leq y_{(S)}$, i.e. as*

$$\sum_{s=1}^{S} c_s y_{(s)}$$

with appropriately chosen c_s. Thus they can be subsumed within the general class of so-called L-estimates, *see Huber [10] for a general definition of L-estimates.*

[3] For a definition of the breakdown point see e.g. Huber [10].

A further class of estimators are the *M-estimates*, see also Huber [10]. M-estimates are maximum likelihood type estimators which come from minimizing some deviation measure. The following Huber estimator is an M-estimate, but also the maximum likelihood estimator and the median fall into that class.

Definition 2.14. *The* Huber estimator μ_S^{HU} *for* μ *based on a sample of length* S *is defined componentwise by*

$$\hat{\mu}_{S,i}^{HU} := \arg\min_{y\in\mathbb{R}} \sum_{s=1}^{S} \rho(R_{s,i} - y) \qquad \text{with } \rho(x) = \begin{cases} \frac{x^2}{2} & \text{if } |x| \leq k \\ k|x| - \frac{k^2}{2} & \text{if } |x| > k \end{cases}$$

for some[4] $k > 0$*, see Huber [9].*

In contrast to the maximum likelihood estimator $\hat{\mu}_S^{ML}$, the remaining return estimators are not linear estimators anymore. Nevertheless, linearity is not the most important property in the portfolio setting. It is of much higher interest if the estimators converge to the true parameters, given more and more historical data. This question concerning consistency will be positively answered in the next section.

2.3 Consistency of estimators

In this section, we will summarize the main properties of the above defined estimators.

Remark 2.15. *Due to the symmetry of the marginal distributions of* R*, it can be shown that all of the above defined L-estimators are unbiased, see Rinne [24], p. 474. Although this may seem to be a very nice property, it does not help portfolio optimization later on. As these estimators enter a portfolio optimization problem, they are mapped to portfolio estimators in a very non-linear way. This causes the resulting estimators to be biased anyway.*

A further property shared by all of the above defined L-estimates is their continuous dependence on the sample $R_1, \ldots, R_S$.

Proposition 2.16. *In the given framework, all of the above defined L-estimates are continuous mappings from* $(\mathbb{R}^n)^S \to \mathbb{R}^n$*.*

Proof. For each component, the mapping from the original sample to the ordered sample is obviously continuous. Further, each quantile is continuous in the ordered sample as it is merely a projection. □

[4]Usually, k is chosen as a suitable multiple of the standard deviation of the sample $R_{s,i}$. In our setting, we assume that k is fixed independent of the sample.

Proposition 2.17. *In the given framework, all estimators* $\hat{\mu}_S^{ML}$, $\hat{\mu}_S^{ME}$, $\hat{\mu}_S^{QR}$, $\hat{\mu}_S^{TM}$ *and* $\hat{\mu}_S^{HU}$ *for the mean* μ *are asymptotically normally distributed.*

Proof.

1. Due to the existence of continuous marginal densities, each finite collection of quantiles asymptotically follows a multivariate normal distribution, see Shorack and Wellner [27], Example 1 on page 639. Thus, both the median and the quartile estimator are asymptotically normally distributed.

2. The trimmed mean is asymptotically normal by Theorem 3.2 on page 60 in Huber [10], as all marginal distributions have continuous densities.

3. According to Huber [9], Section 4, the maximum likelihood estimator is asymptotically normal.

4. Following Huber [10], Theorem 2.4, p. 50 and Example 3.2, p. 135, asymptotic normality is given for the Huber estimator as the marginal densities are sufficiently smooth in our framework.

□

Based on the asymptotic normality, it is easy to derive consistency of all estimators.

Theorem 2.18. *In the given setting it holds that* $\hat{\mu}_S^{ML}$, $\hat{\mu}_S^{ME}$, $\hat{\mu}_S^{QR}$, $\hat{\mu}_S^{TM}$ *and* $\hat{\mu}_S^{HU}$ *are consistent estimators for* μ. *Further,* $\hat{\Sigma}_S^{ML}$ *is a consistent estimator for* Σ.

Proof. As consistency follows directly from asymptotic normality, it only remains to show consistency of the maximum likelihood estimator for Σ. This follows directly by Rinne [24], p. 454. □

3 Mean-variance optimization and portfolio estimators

In the next paragraphs we briefly introduce the traditional mean-variance portfolio selection problem. We show how each of the previous estimators gives rise to a corresponding portfolio estimator and we will prove that each of these consistent estimators leads to a consistent portfolio estimator.

3.1 Traditional mean-variance optimization

We start with the definition of the set of feasible portfolios. Note that the feasible portfolios do not depend on the historical data.

Definition 3.1. *Let* $X := \{x \in \mathbb{R}^n \,|\, x^T\mathbf{1} = 1, Ax \leq b\}$ *be the* set of feasible portfolios*. We assume that the data* $[A, b]$ *is given in such a way that the set* X *is non-empty and compact.*

As compactness of the set X is not a very strong restriction, it is usually fulfilled, for example in the case of long-only constraints, i.e. $A = -I, b = 0$.

If asset returns are distributed with mean r and covariance matrix C, the expected return of a portfolio $x \in X$ is given by $x^T r$ and its risk is $\sqrt{x^T C x}$. In traditional Markowitz optimization, an optimal trade-off – driven by a trade-off factor λ – between risk and expected return is sought for:

Definition 3.2. *Let* $0 \leq \lambda \leq 1$*. Then each optimal solution to the problem* $(MV_{r,C}^{(\lambda)})$ *is called* classical efficient portfolio *for the risk-return trade-off parameter* λ*, i.e.*

$$x^*(\lambda) := x^*(\lambda, r, C) := \arg\min_{x \in X} f_{cl}(x, \lambda, r, C) \qquad (MV_{r,C}^{(\lambda)})$$

with the mean-variance objective

$$f_{cl}(x, \lambda, r, C) := (1 - \lambda)\sqrt{x^T C x} - \lambda(x^T r).$$

Using the parameter λ we can trace the efficient portfolios $x^*(\lambda)$ on the *efficient frontier* from the minimum variance portfolio ($\lambda = 0$) to the maximum return portfolio ($\lambda = 1$).

Remark 3.3. *For the market parameters* (μ, Σ) *problem* $(MV_{\mu,\Sigma}^{(\lambda)})$ *is a convex optimization problem for each* $0 \leq \lambda \leq 1$ *and it possesses at least one optimal solution due to the compactness of* X*. In addition, if* $0 \leq \lambda < 1$ *the solution to* $(MV_{\mu,\Sigma}^{(\lambda)})$ *is unique. In the following, we assume that the maximum component of* μ *is unique, and in this case, uniqueness of the optimal portfolio is given for all* $0 \leq \lambda \leq 1$*.*

Note that problem $(MV_{r,C}^{(\lambda)})$ can be cast as a second-order cone problem and can be solved very efficiently as long as X is described by convex constraints.

Theorem 3.4. *Let* $(MV_{r,C}^{(\lambda)})$ *be uniquely solvable for all* (r, C) *in a neighborhood around* (r_0, C_0)*. Then the mapping* $(r, C) \mapsto x^*(\lambda, r, C)$ *is continuous in the neighborhood around* (r_0, C_0)*.*

Proof. According to Bank et al. [1], Theorem 4.2.1, the optimal set mapping is Hausdorff upper semicontinuous due to compactness of the feasibility set and due to continuity of the objective function. As the solution is assumed to be unique, upper semicontinuity of the set-valued map yields continuity in the usual sense. □

Remark 3.5. *In problem* $(MV_{r,C}^{(\lambda)})$ *the definition of the set* X *can be generalized to the case where it is continuously*[5] *depending on* r *and* C *as long as the set* X *remains non-empty, convex and compact, see Bank et al. [1], Theorem 4.2.1.*

3.2 Portfolio estimators

In the above derivation of the optimal mean-variance portfolios, it was assumed that the mean and variance of the asset returns are known exactly. Given that the true parameters μ and Σ are not known in our situation, estimators $\hat{\mu}_S$ and $\hat{\Sigma}_S$ have to be used instead. If these estimators are plugged into the optimization problem for r and C, i.e. $r = \hat{\mu}_S$, $C = \hat{\Sigma}_S$, the resulting portfolio $x^*(\lambda, \hat{\mu}_S, \hat{\Sigma}_S)$ can be interpreted as an estimator for the true optimal portfolio $x^*(\lambda, \mu, \Sigma)$.

Definition 3.6. *Let* $0 \leq \lambda \leq 1$ *and let* $\hat{\mu}_S$ *and* $\hat{\Sigma}_S$ *be point estimators for* μ *and* Σ*. Then the optimal solution* $x^*_{cl}(\lambda, \hat{\mu}_S, \hat{\Sigma}_S)$ *of the problem* $(MV_{\hat{\mu}_S,\hat{\Sigma}_S}^{(\lambda)})$ *is a portfolio estimator for the true efficient portfolio* $x^*(\lambda)$*.*

The subsequent theorem states that this is indeed a meaningful portfolio estimator if consistency of $x^*_{cl}(\lambda, \hat{\mu}_S, \hat{\Sigma}_S)$ is given.

Theorem 3.7. *Let* $0 \leq \lambda \leq 1$*. Then the portfolio estimator* $x^*_{cl}(\lambda, \hat{\mu}_S, \hat{\Sigma}_S)$ *is a consistent estimator for the true portfolio* $x^*(\lambda, \mu, \Sigma)$ *if the estimators* $\hat{\mu}_S$ *and* $\hat{\Sigma}_S$ *are consistent estimators for* μ *and* Σ*.*

Proof. According to Jacod and Protter [12], Theorem 17.5, almost sure convergence of random variables transfers to continuous functions thereof. Due to the uniqueness of the optimal solution, see Remark 3.3, the continuity is given by Theorem 3.4. □

4 Robustification of the traditional approach

As the classical efficient portfolios are known to be strongly dependent on the optimization parameters, approaches to robustify these solutions have

[5] Continuity is understood in the Hausdorff sense, see e.g. Bank et al. [1].

been developed. We want to apply and investigate the robust counterpart approach introduced by Ben-Tal and Nemirovski in 1998, see e.g. [3]. This approach does not assume or exploit any assumptions about the distribution of the unknown parameters, it rather represents a worst-case solution by optimizing over the worst outcome of the parameter from an entire *uncertainty set* $\mathcal{U}$ of possible realizations.

4.1 Robust counterparts of mean-variance problems

The robust counterpart program associated with the mean-variance problem formulation in Definition 3.2 is given in the following

Definition 4.1. *Let* $0 \leq \lambda \leq 1$. *Then each optimal solution to the problem* ($RC_{\mathcal{U}}^{(\lambda)}$) *is called* robust efficient portfolio *for the parameter* λ *and the uncertainty set* $\mathcal{U}$, *i.e.*

$$x^*_{rob}(\lambda, \mathcal{U}) = \arg\min_{x \in X} f_{rob}(x, \lambda, \mathcal{U}) \qquad (RC_{\mathcal{U}}^{(\lambda)})$$

with the robust mean-variance objective f_{rob}

$$f_{rob}(x, \lambda, \mathcal{U}) = \max_{(r,C) \in \mathcal{U}} f_{cl}(x, \lambda, r, C) = \max_{(r,C) \in \mathcal{U}} (1 - \lambda)\sqrt{x^T C x} - \lambda(x^T r).$$

The uncertainty set $\mathcal{U}$ depends on the parameters or estimates thereof, respectively, and needs to be fully specified. Traditional choices for uncertainty sets are confidence ellipsoids around some point estimates, see e.g. Meucci [18]. An alternative approach based on several point estimates has been investigated in Schöttle and Werner [26].

4.2 Choice of uncertainty sets

In this section the two types of uncertainty sets are investigated closer. We show that both sets are *consistent* uncertainty sets[6] in the following sense.

Definition 4.2. *An uncertainty set* $\mathcal{U} := \mathcal{U}_S$ *is called* consistent *for the parameters* (r, C) *if*

$$d_H\big(\mathcal{U}_S, \{(r, C)\}\big) \to 0 \quad \textit{almost surely for } S \to \infty,$$

where $d_H(A, B)$ *denotes the Hausdorff distance between two sets* A *and* B.

[6]The uncertainty sets are seen as set-valued random variables.

It is common practice to assume that the covariance matrix as given by the estimate $\hat{\Sigma}_S$ is rather exact and only uncertainty around $\hat{\mu}_S$ needs consideration. Therefore, from now on we will only consider uncertainty sets with fixed covariance matrix.

Example 4.3. *The traditional uncertainty set relies on the first two moments of the point estimates, see e.g. Meucci [18], (9.108).*

$$\mathcal{U}_S^1(r_0, C_0) := \left\{ r \mid (r - r_0)^T C_0^{-1} (r - r_0)^T \leq \frac{\kappa}{S} \right\} \times \{C_0\},$$

where κ is chosen to be an appropriate α-quantile of the chi-square distribution with n degrees of freedom, i.e. κ such that $\chi_n^2(\kappa) = \alpha$.

Proposition 4.4. *If $\hat{\mu}_S$ and $\hat{\Sigma}_S$ are consistent estimates for the parameters μ and Σ, then $\mathcal{U}_S^1(\hat{\mu}_S, \hat{\Sigma}_S)$ is a consistent uncertainty set for (μ, Σ).*

Proof. Trivial. □

It can be shown that in the case of the above uncertainty set the robust and the classical frontier coincide up to a certain risk (i.e. volatility) level.

Proposition 4.5. *Consider the classical program $(MV_{\hat{\mu}_S, \hat{\Sigma}_S}^{(\lambda)})$ and the robust counterpart $(RC_{\mathcal{U}}^{(\theta)})$ with the uncertainty set $\mathcal{U}_S^1(\hat{\mu}_S, \hat{\Sigma}_S)$. Then for each $\theta \in [0, 1]$ there exists $\lambda = \lambda(\theta) = \frac{\theta}{1+\theta\sqrt{\frac{\kappa}{S}}} \in \left[0, \frac{1}{1+\sqrt{\frac{\kappa}{S}}}\right]$ such that the optimal solution $x_{rob}^*(\theta, \mathcal{U}_S^1)$ of the robust problem equals the optimal solution $x_{cl}^*(\lambda, \hat{\mu}_S, \hat{\Sigma}_S)$ of the classical problem.*

Proof. We prove coincidence of the classical and the robust solution by showing equivalence of the corresponding optimization problems. Let us start with the classical problem given by

$$\min_{x \in X} \quad (1 - \lambda)\sqrt{x^T \hat{\Sigma}_S x} - \lambda(x^T \hat{\mu}_S).$$

In the robust version, we can solve the inner maximization problem analytically and thus we can reformulate the robust problem with parameter θ as follows:

$$\begin{aligned}
&\min_{x \in X} \quad \max_{(\mu, \Sigma) \in \mathcal{U}_S^1(\hat{\mu}_S, \hat{\Sigma}_S)} (1 - \theta)\sqrt{x^T \Sigma x} - \theta(x^T \mu) \\
=&\min_{x \in X} \quad (1 - \theta)\sqrt{x^T \hat{\Sigma}_S x} - \theta(x^T \hat{\mu}_S) + \theta\sqrt{\frac{\kappa}{S}}\sqrt{x^T \hat{\Sigma}_S x} \\
=&\min_{x \in X} \quad \left(1 - \theta + \theta\sqrt{\frac{\kappa}{S}}\right)\sqrt{x^T \hat{\Sigma}_S x} - \theta(x^T \hat{\mu}_S).
\end{aligned}$$

Now plug in $\lambda(\theta)$ as proposed in the claim

$$\lambda := \frac{\theta}{1+\theta\sqrt{\frac{\kappa}{S}}}$$

and the classical problem reformulates to

$$\begin{aligned}
&\min_{x\in X} \quad (1-\lambda)\sqrt{x^T\hat{\Sigma}_S x} - \lambda(x^T\hat{\mu}_S) \\
&= \min_{x\in X} \quad \left(1 - \frac{\theta}{1+\theta\sqrt{\frac{\kappa}{S}}}\right)\sqrt{x^T\hat{\Sigma}_S x} - \frac{\theta}{1+\theta\sqrt{\frac{\kappa}{S}}}(x^T\hat{\mu}_S) \\
&= \min_{x\in X} \quad \frac{1}{1+\theta\sqrt{\frac{\kappa}{S}}}\left[\left(1-\theta+\theta\sqrt{\frac{\kappa}{S}}\right)\sqrt{x^T\hat{\Sigma}_S x} - \theta(x^T\hat{\mu}_S)\right]
\end{aligned}$$

which is obviously equivalent to the robust formulation. □

An alternative approach, leading to a different robust frontier, was presented in Schöttle and Werner [26].

Example 4.6. *Consider the following set M_S of consistent estimators for μ:*

$$M_S := \{\hat{\mu}_S^{ML}, \hat{\mu}_S^{ME}, \hat{\mu}_S^{QR}, \hat{\mu}_S^{TM}, \hat{\mu}_S^{HU}\}$$

i.e. the set of the five considered estimators for the mean return vector. An obvious choice for a consistent uncertainty set would be to use the convex hull of these estimators:

$$\mathcal{U}_S := conv(M_S) \times \{\hat{\Sigma}_S\}.$$

As shown in Schöttle and Werner [25], robustification is more successful if ellipsoidal uncertainy sets are employed rather than polyhedral ones. Thus, instead of taking the polyhedral convex hull of these estimators the following ellipsoidal set was introduced:

$$\mathcal{U}_S^2 = \{u \,|\, (u-\bar{\mu}_S)^T\bar{\Sigma}_S^{-1}(u-\bar{\mu}_S) \le \delta_S^2\} \times \{\hat{\Sigma}_S\} \tag{4.1}$$

$$\begin{aligned}
\textit{with} \quad \bar{\mu}_S &= \frac{1}{|M_S|}\sum_{m\in M_S} m \\
\bar{\sigma}^2{}_{S,ii} &= \frac{1}{|M_S|-1}\sum_{m\in M_S}(m_i - \bar{\mu}_{S,i})^2 \\
\bar{\Sigma}_S &= \mathrm{diag}(\bar{\sigma}^2_{S,11}, \ldots, \bar{\sigma}^2_{S,nn}) \\
\delta_S^2 &= \max_{m\in M_S}(m-\bar{\mu}_S)^T\bar{\Sigma}_S^{-1}(m-\bar{\mu}_S).
\end{aligned}$$

Verbally expressed, the uncertainty set for the return is given by the smallest ellipsoid centered at the common average of the selected estimators and shaped by their standard deviations.

Figure 1 illustrates the difference between the two uncertainty sets $\mathcal{U}_S^1$ and $\mathcal{U}_S^2$.

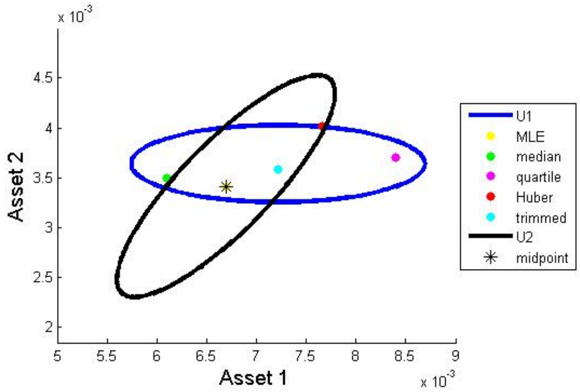

Figure 1: Illustration of the uncertainty sets $\mathcal{U}_S^1$ ($\alpha = 10\%$) and $\mathcal{U}_S^2$ in the two-dimensional case.

Proposition 4.7. *If M_S only contains consistent estimators for μ and if $\hat{\Sigma}_S$ is a consistent estimator for Σ, then $\mathcal{U}_S^2$ as defined in (4.1) is a consistent uncertainty set for (μ, Σ).*

Proof. It suffices to show that for $S \to \infty$

$$\{u \,|\, (u - \bar{\mu}_S)^T \bar{\Sigma}_S^{-1} (u - \bar{\mu}_S) \leq \delta_S^2\} \to \{\mu\}.$$

Using the square root matrix of $\bar{\Sigma}_S$ this set can be equivalently written as

$$\{u \,|\, u = \bar{\mu}_S + \delta_S \bar{\Sigma}_S^{\frac{1}{2}} z, \;\; \|z\|_2 \leq 1\}.$$

Obviously, we have $\bar{\mu}_S \to \mu$. Furthermore, δ_S^2 is bounded:

$$\delta_S^2 = \max_{m \in M_S} (m - \bar{\mu}_S)^T \bar{\Sigma}_S^{-1} (m - \bar{r}_S) = \max_{m \in M_S} \sum_{i=1}^{n} (m_i - \bar{\mu}_{S,i})^2 \cdot \frac{1}{\bar{\sigma}_{S,ii}^2} \leq n|M_S|$$

where the last inequality is due to the definition of $\bar{\sigma}_{S,ii}^2$. Then the statement follows as $\bar{\Sigma}_S \to 0$ due to consistency. □

4.3 Robustified portfolio estimators

Analogous to Section 3.2 the robust portfolio estimator is defined by plugging in the appropriate uncertainty set $\mathcal{U}_S$.

Definition 4.8. *Let* $0 \leq \lambda \leq 1$*. Then the optimal solution* $x^*_{rob}(\lambda, \mathcal{U}_S)$ *to the robust problem* ($RC^{(\lambda)}_{\mathcal{U}_S}$) *is a portfolio estimator for the true efficient portfolio* $x^*(\lambda)$*.*

Given the explicit specification of the uncertainty set in Equation 4.1, the inner maximization problem in the robust objective function can be solved and thus f_{rob} simplifies

$$\begin{aligned} f_{rob}(x, \lambda, \mathcal{U}_S^2) &= \max_{(r,C)\in\mathcal{U}_S^2} (1-\lambda)\sqrt{x^T C x} - \lambda(x^T r) \\ &= (1-\lambda)\sqrt{x^T \hat{\Sigma}_S x} - \lambda x^T \bar{\mu}_S + \delta_S \lambda \sqrt{x^T \bar{\Sigma}_S x}. \end{aligned}$$

From the above formulation, it becomes clear that the robust portfolio estimator is again consistent. This can be proved in a very general form in the next theorem.

Theorem 4.9. *Let* $0 \leq \lambda \leq 1$ *and let* $\mathcal{U}_S$ *be consistent for* (μ, Σ)*. Then the robust portfolio estimator* $x^*_{rob}(\lambda, \mathcal{U}_S)$ *is a consistent estimator for the true optimal portfolio* $x^*(\lambda)$*.*

Proof. The proof is similar in idea to the proof of Theorem 3.4 and is based on the same concepts. According to the following Lemma 4.10, f_{rob} is continuous in $\mathcal{U}$. Thus, as f_{rob} is also continuous in x and the feasibility set X is constant we can apply Theorem 4.2.1 in Bank et al. [1]. This yields upper semicontinuity of the set valued map of optimal solutions. It can easily be shown that $x^*_{rob}(\lambda, \mathcal{U})$ is again unique, and hence continuity of $x^*_{rob}(\lambda, \mathcal{U})$ in $\mathcal{U}$ is obtained. As $\mathcal{U}_S$ is reduced in the limit to the true parameters, the statement follows. □

Lemma 4.10. *The mapping* $\mathcal{U} \mapsto f_{rob}(x, \lambda, \mathcal{U})$ *is continuous in* $\mathcal{U}$ *for all* $x \in X$ *and* $\lambda \in [0, 1]$*.*

Proof. Let $\mathcal{U}_n \to \mathcal{U}$ in the Hausdorff sense, i.e. $\lim_{n\to\infty} d_H(\mathcal{U}_n, \mathcal{U}) = 0$. We proceed in two steps.
Step 1: Let (r^*, C^*) be a maximizing element in $\mathcal{U}$ which maximizes f_{cl} for x and λ, i.e.

$$f_{cl}(x, \lambda, r^*, C^*) = \max_{(r,C)\in\mathcal{U}} f_{cl}(x, \lambda, r, C).$$

Due to convergence of $\mathcal{U}_n$ to $\mathcal{U}$ in the Hausdorff sense there exists $(r_n, C_n) \in \mathcal{U}_n$ with $(r_n, C_n) \to (r^*, C^*)$. As $f_{cl}(x, \lambda, r, C)$ is continuous in (r, C), we have

$$f_{rob}(x, \lambda, \mathcal{U}) = f_{cl}(x, \lambda, r^*, C^*) = \lim_{n\to\infty} f_{cl}(x, \lambda, r_n, C_n) \leq \lim_{n\to\infty} f_{rob}(x, \lambda, \mathcal{U}_n).$$

Step 2: Let $(r_n^*, C_n^*) \in \mathcal{U}_n$ be the maximizing elements for $\mathcal{U}_n$, i.e.

$$f_{cl}(x, \lambda, r_n^*, C_n^*) = \max_{(r,C)\in\mathcal{U}_n} f_{cl}(x, \lambda, r, C).$$

Because of Hausdorff convergence of $\mathcal{U}_n$ to $\mathcal{U}$ there exist $(r_n, C_n) \in \mathcal{U}$ with $||r_n - r_n^*||_2 \to 0$ and $||C_n - C_n^*||_2 \to 0$. Then any cluster point $(r^*, C^*) \in \mathcal{U}$ is also an accumulation point for the sequence (r_n^*, C_n^*). Without loss of generality assume that this cluster point is unique. Then it holds

$$\lim_{n\to\infty} f_{rob}(x, \lambda, \mathcal{U}_n) = \lim_{n\to\infty} f_{cl}(x, \lambda, r_n^*, C_n^*) = f_{cl}(x, \lambda, r^*, C^*) \leq f_{rob}(x, \lambda, \mathcal{U}).$$

Combining both inequalities proves the statement. □

Remark 4.11. *In the above theorem, the strict convexity of f_{cl} in x for $0 \leq \lambda < 1$ carries over to f_{rob} and thus guarantees uniqueness of the solution. In the case $\lambda = 1$, uniqueness follows from the uniqueness assumption in Remark 3.3.*

5 Estimators via resampled portfolios

Different from the robust counterpart approach the *resampling technique* presented in the next paragraphs accounts for uncertainty in the parameters by considering the distributional aspect of the parameters. After presenting the resampling approach in more detail, we again introduce the according portfolio estimator and show its consistency.

5.1 Resampling methodology

The resampling technique as a bootstrapping technique was introduced by Michaud [19] in 1998 and is based on concepts dating back to Jorion [14]. The core idea is not only to consider point estimators $(\hat{\mu}, \hat{\Sigma})$ for (μ, Σ) but to incorporate additional information. Thus, the random nature of the estimators $(\hat{\mu}, \hat{\Sigma})$ is taken into account and information about their distribution is exploited to obtain more robust portfolio estimators. As only the estimators $(\hat{\mu}, \hat{\Sigma})$ are known, some assumption about their distribution has to be made. This is usally done via assumptions on the return distribution, which then carry over to the distribution of the estimators. In the following, we will describe a very general resampling procedure, which includes the original resampling procedure of Michaud as a special case.

Definition 5.1. *The resampling procedure to obtain a* resampled portfolio *based on a sample of length S for resampling parameters $\hat{r}$ and $\hat{C}$ is the following.*

1. *Fix an elliptical distribution of resampled asset returns, assuming a suitable characteristic generator* $\hat{\psi}$.

2. *Then, the* resampling distribution, *i.e. the distribution of the resampled asset return* $\hat{R}$, *is given by*

$$\hat{R} \sim \mathcal{E}(\hat{r}, \hat{C}, \hat{\psi}).$$

3. *Based on* S *independent realizations from the resampling distribution, bootstrapped estimators* $(\tilde{r}, \tilde{C})$ *for* $(\hat{r}, \hat{C})$ *can be computed.*

   It is required that the estimation methods for the bootstrapped estimators satisfy certain restrictions. First, the estimators have to be continuous in the sample, and second, the estimators must be consistent. Thus, all presented L-estimates (see Proposition 2.16 and Theorem 2.18) and especially the sample covariance estimator fall within this category.

4. *From the assumed resampling distribution of* $\hat{R}$ *the corresponding bootstrapped distribution* $\mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi})$ *of the bootstrapped estimators* $(\tilde{r}, \tilde{C})$ *can be obtained, i.e.*

$$(\tilde{r}, \tilde{C}) \sim \mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi}).$$

   This distribution is usually not analytically available, but can be approximated by Monte Carlo techniques.

5. *Plugging the bootstrapped estimators* $(\tilde{r}, \tilde{C})$ *into the classical portfolio formulation yields the bootstrapped portfolios* $x^*_{cl}(\lambda, \tilde{r}, \tilde{C})$. *Their distribution can be derived from the distribution of the bootstrapped estimators.*

6. *Based on the distribution of the bootstrapped portfolios, one resulting* resampled portfolio *has to be determined. Michaud [19] suggests to use the average of the bootstrapped portfolios as resampled portfolio:*

$$x^*_{re}(\lambda, \hat{r}, \hat{C}, \hat{\psi}, S) := \int_{\mathbb{R}^n \times \mathbb{R}^{n \times n}} x^*_{cl}(\lambda, r, C)\varphi_{\mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi})}(r, C) dr dC. \quad (5.1)$$

Remark 5.2. *In case the characteristic generator is fixed in such a way that the distribution of the resampled asset return* $\hat{R}$ *is multivariate normal and the maximum likelihood estimators are used, then the bootstrapped distribution* $\mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi})$ *is given by Proposition 2.7, i.e.*

$$\mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi}) = \mathcal{NW}(\hat{r}, \hat{C}, S).$$

Remark 5.3. *Formula (5.1) for the resampled portfolio can be interpreted as calculating the* expected bootstrapped portfolio *with respect to the distribution of the bootstrapped estimators:*

$$x^*_{re}(\lambda, \hat{r}, \hat{C}, \hat{\psi}, S) = \int\limits_{\mathbb{R}^n \times \mathbb{R}^{n \times n}} x^*_{cl}(\lambda, r, C)\varphi_{\mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi})}(r, C) dr dC$$

$$= \mathbf{E}[x^*_{cl}(\lambda, r, C)] \quad \textit{with } (r, C) \sim \mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi}).$$

At first glance it may seem that this definition of the resampled portfolio relies heavily on the choice of the generator, but it can be shown that in the limit this generator does not play a crucial role. Of course, for small sample sizes S the choice of the generator function has an impact on the resulting resampled portfolio.

5.2 Resampled portfolio estimators

As already mentioned, the choice of the marginal generator is not important. The central role is instead played by the appropriate choice of the estimator, which has to be consistent and continuous.

Proposition 5.4. *Let* $0 \leq \lambda \leq 1$ *and let sequences of parameters* $(\hat{r}_k)_k$ *and* $(\hat{C}_k)_k$ *be given with* $\hat{r}_k \to \hat{r}$ *and* $\hat{C}_k \to \hat{C}$ *for* $k \to \infty$*. Then it holds for the resampled portfolio – defined in (5.1) – that*

$$\lim_{k \to \infty} x^*_{re}(\lambda, \hat{r}_k, \hat{C}_k, \hat{\psi}, S) = x^*_{re}(\lambda, \hat{r}, \hat{C}, \hat{\psi}, S).$$

Proof.
Step 1: As the generator $\hat{\psi}$ is continuous, it is easy to see that the characteristic functions converge pointwise. According to Breiman [2], Corollary 8.30, this is exactly convergence in distribution: $\lim_{k \to \infty} \mathcal{E}(\hat{r}_k, \hat{C}_k, \hat{\psi}) = \mathcal{E}(\hat{r}, \hat{C}, \hat{\psi})$.
Step 2: In the definition of the resampling procedure it was required that the estimator is continuous. Then according to Breiman [2], Proposition 8.19, the estimators also convergence in distribution: $\lim_{k \to \infty} \mathcal{B}_S(\hat{r}_k, \hat{C}_k, \hat{\psi}) = \mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi})$.
Step 3: Due to the compactness assumption on X and due to continuity of the classical portfolio estimator x^*_{cl} with respect to the parameters (see Theorem 3.4), we can apply Breiman [2], Proposition 8.12 to show the statement of the proposition. □

Proposition 5.5. *Let* $0 \leq \lambda \leq 1$ *and let sequences of parameters* $(\hat{r}_k)_k$ *and* $(\hat{C}_k)_k$ *be given with* $\hat{r}_k \to \hat{r}$ *and* $\hat{C}_k \to \hat{C}$ *for* $k \to \infty$*. Then for* $k \to \infty$ *and* $S \to \infty$ *it holds for the resampled portfolio that*

$$\lim_{\substack{k \to \infty \\ S \to \infty}} x^*_{re}(\lambda, \hat{r}_k, \hat{C}_k, \hat{\psi}, S) = x^*_{cl}(\lambda, \hat{r}, \hat{C})$$

independent of the generator $\hat{\psi}$.

Proof.
Step 1: Obviously: $\lim_{\substack{k\to\infty \\ S\to\infty}} x^*_{re}(\lambda, \hat{r}_k, \hat{C}_k, \hat{\psi}, S) = \lim_{S\to\infty} \big(\lim_{k\to\infty} x^*_{re}(\lambda, \hat{r}_k, \hat{C}_k, \hat{\psi}, S) \big)$.
Hence, it remains to show $\lim_{S\to\infty} x^*_{re}(\lambda, \hat{r}, \hat{C}, \hat{\psi}, S) = x^*_{cl}(\lambda, \hat{r}, \hat{C})$.
Step 2: Let $R_1, \ldots, R_S \sim \mathcal{E}(\hat{r}, \hat{C}, \hat{\psi})$ i.i.d. As the estimators were chosen to be consistent, the estimators $(\tilde{r}_S, \tilde{C}_S)$ converge almost surely to $(\hat{r}, \hat{C})$ as $S \to \infty$. As almost sure convergence implies convergence in distribution, we thus have $\lim_{S\to\infty} \mathcal{B}_S(\hat{r}, \hat{C}, \hat{\psi}) = \delta(\hat{r}, \hat{C})$, i.e. convergence to the Dirac measure with mass in $(\hat{r}, \hat{C})$.
Step 3: In analogy to Step 3 in the proof of the previous proposition, the statement follows due to continuity and boundedness of x^*_{cl} from Breiman [2], Proposition 8.12. □

Theorem 5.6. *Let $0 \le \lambda \le 1$ and let the estimators $\hat{\mu}_S$ and $\hat{\Sigma}_S$ be consistent estimators for μ and Σ. Then the portfolio estimator $x^*_{re}(\lambda, \hat{\mu}_S, \hat{\Sigma}_S, \hat{\psi}, S)$ is a consistent estimator for the true portfolio $x^*(\lambda)$.*

Proof. Follows directly from the previous proposition. □

Remark 5.7. *Note that the estimators inside the resampling procedure have to fulfill slightly stronger restrictions – continuous dependence on the sample – than the estimators used for the estimation of the resampling parameters.*

6 Summary

In this paper we have shown consistency of three different portfolio estimators, provided that consistent estimators for first and second moments are used. In detail, consistency was verified for the traditional Markowitz setting under very weak conditions which guarantee continuity of the optimal solution with respect to the input parameters. In the same way, consistency for robust portfolio estimators was derived for appropriate *consistent* choices of uncertainty sets. Under slightly more restrictive assumptions – continuity of the estimators – it was also possible to establish consistency for the (generalized) resampled procedures in the sense of Michaud.

As we have seen in Section 2, all estimators are not only consistent, but even asymptotically normally distributed. In a next step we want to investigate conditions under which this property carries over to the portfolio estimator, i.e. under which conditions the optimal portfolios, especially the robustified ones, depend differentiably on the input parameters.

Acknowledgements

We appreciate the supporting comments by two anonymous referees on the original paper [26]. Their comments motivated the more detailed investigation of consistency presented in this paper.

References

[1] Bank, B. et al., Non-Linear Parametric Optimization, *Birkhäuser, Basel*, 1983.

[2] Breiman, L., Probability, *SIAM, Philadelphia*, 1992.

[3] Ben-Tal, A. and Nemirovski, A., Robust convex optimization, *Mathematics of Operations Research*, **23**, **4**, 1998.

[4] Ceria, S. and Stubbs, R. A., Incorporating estimation errors into portfolio selection: Robust portfolio construction, *Journal of Asset Management*, **7**, **2** 2006.

[5] DeMiguel, V. and Nogales, F., Portfolio Selection with Robust Estimates of Risk, *SSRN working paper*, 2006 [online]. Available from: http://papers.ssrn.com

[6] Fang, K.-T. and Kotz, S. and Ng, K.-W., Generalized multivariate analysis, *Science Press, Beijing*, 1990.

[7] Fang, K.-T. and Zhang, Y.-T., Symmetric Multivariate and Related Distributions, *Chapman and Hall, London*, 1990.

[8] Goldfarb, D. and Iyengar, G., Robust Portfolio Selection Problems, *Mathematics of Operations Research* **28**, **1**, 2003.

[9] Huber, P. J., Robust Estimation of a Location Parameter, *The Annals of Mathematical Statistics*, **35**, **1**, 1964.

[10] Huber, P. J., Robust Statistics, *John Wiley & Sons, New York*, 1981.

[11] Ingersoll, J. E., Theory of Financial Decision Making, *Rowman & Littlefield, Maryland*, 1987.

[12] Jacod, J. and Protter, P., Probability Essentials, Second Edition, *Springer, Berlin*, 2003.

[13] Jobson, J. D. and Korkie, B., Estimation for Markowitz efficient portfolios, *J. Amer. Statist. Assoc.*, **75**, **371**, 1980.

[14] Jorion, P., Portfolio Optimization in Practice, *Financial Analysts Journal*, **48**, **1**, 1992.

[15] Kempf, A. and Memmel, C., On the Estimation of the Global Minimum Variance Portfolio, *SSRN working paper*, 2003 [online]. Available from: http://papers.ssrn.com

[16] Lauprête, G. J. and Samarov, A. M. and Welsch, R. E., Robust Portfolio Optimization, *Metrika*, **55**, **2**, 2002.

[17] Lutgens, F., Robust Portfolio Optimization, *PhD Thesis, Maastricht University*, 2004.

[18] Meucci, A., Risk and Asset Allocation, *Springer, Berlin*, 2005.

[19] Michaud, R. O., Efficient Asset Management, *Harvard Business School Press, Boston*, 1998.

[20] Mori, H., Finite sample properties of estimators for the optimal portfolio weight, *Journal of the Japan Statistical Society*, **34**, **1**, 2004.

[21] Müller, P. H. (Hrsg.), Wahrscheinlichkeitsrechnung und Mathematische Statistik. Lexikon der Stochastik. 5., bearbeitete und wesentlich erweiterte Auflage, *Akademie Verlag, Berlin*, 1991.

[22] Okhrin, Y. and Schmid, W., Distributional properties of portfolio weights, *Journal of Econometrics*, **134**, **1**, 2006.

[23] Perret-Gentil, C. and Victoria-Feser, M.-P., Robust Mean-Variance Portfolio Selection, *FAME Research Paper no 140*, 2004.

[24] Rinne, H., Taschenbuch der Statistik, 3., vollständig überarbeitete und erweiterte Auflage, *Harri Deutsch GmbH, Frankfurt*, 2003.

[25] Schöttle, K. and Werner, R., Costs and benefits of robust optimization, *Working Paper, TU München*, 2005 [online]. Available from: http://www.mathfinance.ma.tum.de

[26] Schöttle, K. and Werner, R., Towards reliable efficient frontiers, *Journal of Asset Management*, **7**, **2**, 2006.

[27] Shorack, G. R. and Wellner, J. A., Empirical Processes with Applications to Statistics, *John Wiley & Sons, New York*, 1986.

[28] Tütüncü, R. H. and Koenig, M., Robust Asset Allocation, *Annals of Operations Research*, **132**, **1-4**, 2004.