COMISEF WORKING PAPERS SERIES

WPS-041 23/07/2010

Robust Portfolio Optimization with a Hybrid Heuristic Algorithm

B. Fastrich
P. Winker

Robust Portfolio Optimization with a Hybrid Heuristic Algorithm

Björn Fastrich, Peter Winker

Department of Economics, Justus-Liebig University Giessen

{Bjoern.Fastrich, Peter.Winker}@wirtschaft.uni-giessen.de

September 24, 2009

Abstract

Estimation errors in both the expected returns and the covariance matrix hamper the constructing of reliable portfolios within the Markowitz framework. Robust techniques that incorporate the uncertainty about the unknown parameters are suggested in the literature. We propose a modification as well as an extension of such a technique and compare both with another robust approach. In order to eliminate oversimplifications of Markowitz' portfolio theory, we generalize the optimization framework to better emulate a more realistic investment environment. Because the adjusted optimization problem is no longer solvable with standard algorithms, we employ a hybrid heuristic to tackle this problem. Our empirical analysis is conducted with a moving time window for returns of the German stock index DAX100. The results of all three robust approaches yield more stable portfolio compositions than those of the original Markowitz framework. Moreover, the out-of-sample risk of the robust approaches is lower and less volatile while their returns are not necessarily smaller.

Keywords: Hybrid heuristic algorithm, Markowitz, Robust optimization, Uncertainty sets.

1 Introduction

An investor's primary objective is to optimally allocate his financial resources among a given choice of assets. This allocation is traditionally modeled following Markowitz (1952), the first to explicitly quantify the important trade-off between risk and return within the process of portfolio selection. The model assumes a market of K assets with multivariate normally distributed expected returns, given in the $(K \times 1)-$vector $\boldsymbol{\mu}$, and a $(K \times K)-$covariance matrix of returns $\boldsymbol{\Sigma}$. Efficient portfolios can be constructed if the weights w_i of the assets $i = 1, ..., K$, are chosen such that the following problem is solved:

$$\max_{\boldsymbol{w}} \ (1 - \lambda)\boldsymbol{w}'\boldsymbol{\mu} - \lambda\sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}} \tag{1}$$

subject to

$$\boldsymbol{w}'\boldsymbol{l}_K = 1$$

where $\boldsymbol{l}_K = (1, ..., 1)'$ and $\boldsymbol{w} = (w_1, ..., w_K)'$. The weighting parameter $\lambda \in [0, 1]$ can be interpreted as a risk aversion parameter, since it takes into account the trade-off between risk (measured by $\sigma_p = \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}}$) and return of the portfolios. Repeatedly solving (1) for several values $\lambda \in [0, 1]$ yields the efficient frontier. This framework is called the Mean-Variance-Optimization (MVO) approach.

In order to eliminate certain oversimplifications of Markowitz' portfolio theory, the framework can be adjusted by side conditions to better emulate a more realistic investment environment. Hence, in this paper, the asset weights are restricted to be only within a lower ($w_l \geq 0$) and an upper ($w_u \leq 1$) bound, meaning that short-selling is prohibited. Furthermore, following, e.g., Maringer (2005), the limited divisibility of assets is considered as well as the fact that investors must pay transaction costs. The former ensures the minimum fraction of the investor's capital endowment V being invested in asset i with price P_i is P_i/V, corresponding to one piece of asset i. These constraints lead to discrete weights $w_i = n_i P_i / V$, where $n_i \in \mathbb{N}_0^+$. The transaction costs are modeled as a composite of a fixed payment c_f per asset traded plus a fraction c_v of the volume $n_i P_i$, i.e. $c_v n_i P_i$. As a result of the transaction costs and the integer constraints, it is likely that the investor's capital endowment cannot be entirely invested into assets; the remainder R will be held in cash. Moreover it appears to be unrealistic to hold as many different assets as possible in order to diversify the risk of the portfolio. Portfolios with a very large number of different assets may become impractical to handle for administrative reasons (Maringer 2005). In order to take this issue into account, a limit $K_{max} < K$ on the number of different assets held in a portfolio p is introduced (Chang *et al.* 2000). A subset $\mathfrak{I}_p \subset \{1, ..., K\}$

2

containing $K_{max}^{ef} \leq K_{max}$ assets' indices i is introduced to represent a specific portfolio's constituents. The subscript I denotes vectors and matrices corresponding to the asset indices in $\mathfrak{I}_p$.[1] Taking these constraints into account the optimization problem becomes:

$$\max_{n} \ (1 - \lambda)\left[\frac{\left(\sum_{i \in \mathfrak{I}} n_i P_i (1 + \mu_i) - C_i\right) + R_p}{V} - 1\right] - \lambda\left(\sqrt{\boldsymbol{w}_I' \boldsymbol{\Sigma}_I \boldsymbol{w}_I}\right) \quad (2)$$

subject to

$w_i = \frac{n_i P_i}{V} \quad \forall\, i \in \mathfrak{I}_p$	discrete portfolio weights
$n_i \in \mathbb{N}_0^+ \quad \forall\, i \in \mathfrak{I}_p$	integer constraints
$C_i = c_f + c_p n_i P_i \quad \forall\, i \in \mathfrak{I}_p$	transaction costs
$R_p = V - \sum_{i \in \mathfrak{I}_p}(n_i P_i + C_i)$	residual cash holdings
$\boldsymbol{w}_I' \boldsymbol{l}_{K_I^{ef}} + \frac{\left(\sum_{i \in \mathfrak{I}} C_i\right) + R_p}{V} = 1$	budget constraint
$w_l \leq w_i \leq w_u \quad \forall\, i \in \mathfrak{I}_p$	no short-selling
$\#\mathfrak{I}_p = K_{max}^{ef} \leq K_{max}$	cardinality constraint

Due to its complexity, a discrete search space, and multiple optima, problem (2) will be optimized with a modified hybrid heuristic algorithm (HHA) introduced by Maringer and Kellerer (2003). Heuristic algorithms have been employed in the field of portfolio optimization since Dueck and Winker (1992). An overview of heuristic optimization techniques is provided by Gilli and Winker (2009) and in finance by Maringer and Winker (2007).

In order to optimize either problem with real world data, estimators $\hat{\mu}_i$ and $\hat{\boldsymbol{\Sigma}}$ for the true (but unknown) parameters μ_i and $\boldsymbol{\Sigma}$ must be used. Estimation errors will inevitably occur that coincide with a high sensitivity to changes in the input parameters of MVO-portfolios. Michaud (1989) argues that Markowitz' portfolio theory is *error-maximizing*, because those assets in a portfolio that get overweighted (underweighted) have the largest (smallest) expected return-variance-ratios, and that these exact assets also exhibit the highest probability of large (small) estimation errors. There is a vast literature addressing this problem ranging from restricting portfolio weights (e.g. Frost and Savarino 1988) to Bayesian approaches (e.g. Black and Litterman 1991) and resampling methods (e.g. Michaud 1998). More recently proposed methods can be categorized as *robust estimation* and *robust optimization* approaches.

---

[1] I represents an individual, i.e. a portfolio. This nomenclature will become clear later.

Robust estimation focuses on robust statistical procedures to estimate $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, which are less sensitive to outliers and might result in more robust portfolios. While Maronna *et al.* (2006) provide an overview of robust estimators, the contributions by, e.g., Cavadini *et al.* (2001), Lauprete *et al.* (2002), Perret-Gentil and Victoria-Feser (2004), Welsch and Zhou (2007), Genton and Ronchetti (2008), DeMiguel and Nogales (2009), and Winker *et al.* (2009) belong to this class of robust estimation approaches.

In contrast to robust estimation, the idea of *robust optimization* is to explicitly incorporate the uncertainty about the parameters into the optimization process. Therefore, instead of considering a single point estimate, uncertainty sets are used that contain a certain selection of point estimates. With the objective of constructing a portfolio that exhibits good characteristics for many possible scenarios, the portfolio is optimized under the worst-case-scenario for the given uncertainty set. The articles by, e.g., Goldfarb and Iyengar (2003), Tütüncü and Koenig (2004), Ceria and Stubbs (2006), Bertsimas and Pachamanova (2008), and Zymler *et al.* (2009) apply this idea in different ways. Fabozzi *et al.* (2007) provide a comprehensive overview. In this paper, we enrich this choice of robust optimization techniques with an approach that upgrades the technique of Ceria and Stubbs (2006) by an uncertainty set for the portfolio risk. Moreover, we propose refined proceedings for the computation of uncertainty sets. Our results are compared empirically to those obtained by to the approaches of Tütüncü and Koenig (2004) and Ceria and Stubbs (2006). To solve optimization problem (2) for these robust approaches the aforementioned HHA is employed.

The rest of the paper is structured as follows: Chapter 2 describes the optimization algorithm as well as the implemented robust techniques. Chapter 3 presents empirical results for daily German stock returns before Chapter 4 concludes.

2 The implemented techniques

As previously mentioned, the methods to construct uncertainty sets vary over different robust approaches. However, as this paper employs the optimizer HHA, it is explained in some detail first. Afterwards the robust techniques and their combination with the HHA are addressed.

2.1 The Optimizer: Hybrid Heuristic Algorithm

Maringer and Kellerer (2003) introduced a novel heuristic in a successful application to a portfolio selection problem. The authors' intent was to

overcome certain shortcomings of local search algorithms such as Simulated Annealing (SA) (Kirkpatrick *et al.* 1983). While the shortcomings are the dependence on the *single* search agent's starting point and the rapidly increasing peril of getting stuck in local optima when the problem complexity increases, these algorithms also exhibit the advantage of a relatively precise search within a predefined (local) neighborhood. Population based methods, such as genetic algorithms, work with *multiple* search agents that constitute a greater global search potential and consequently a greater capacity to cope with rather complex problems. In order to include the advantages of both local and population based methods, the proposed hybrid algorithm of Maringer and Kellerer (2003) has (i) more than one starting point and (ii) an entire population of communicating search agents. However, a precise search is preserved through (iii) embedding an SA-algorithm.

After initializing a population of *Pop* search agents I_p, $p = 1, ..., Pop$, that represent portfolios and are termed individuals, the three phases *Modification, Evaluation* and *Replacement* are repeated over a predefined number of generations g. In the *Modification Phase* an SA-algorithm is conducted independently for all individuals and the *Evaluation Phase* ranks these individuals according to their objective function values. In the *Replacement Phase* the worst individuals are replaced either by so-called *Clones*, i.e., exact copies of the current populations' best individuals, or by so-called *Averaged Idols*, i.e., individuals that combine characteristics which have been proven to be successful in other individuals. In contrast to, e.g., genetic algorithms, experience does not only get passed on by two parents, but by an entire group of successful individuals. Moreover, in combination with the SA-algorithm in the *Modification Phase*, this arrangement of the phases helps to find a successful combination of assets, i.e., a core structure, in earlier generations, before it then contributes to assigning proper portfolio weights to this core structures' assets.

The hybrid heuristic algorithm (HHA) that we employ in this paper is based on the algorithm by Maringer and Kellerer (2003). Our advancements modify primarily two aspects.[2] Firstly, as the embedded local search strategy we use a Threshold Accepting (TA) algorithm (Dueck and Scheuer 1990). Therefore, impairments in the objective function are accepted *deterministically* instead of stochastically whenever they do not exceed a predefined threshold T_t, $t = 1, ..., thresh$, that decreases as t grows and the algorithm matures over the generations g respectively. Similar to the TA in Gilli and Këllezi (2002) the step size $U_t \in [U_{min}, U_{max}]$, that is applied in the *Modification Phase* and that defines the neighborhood of an individual I_p^g, is also

---

[2]A detailed description of the HHA can be found in Appendix 1.

5

deterministic. Secondly, the *Replacement Phase* is written in such a way that the worst individuals only get replaced if the opponents' objective function values are not worse by more than the current threshold value T_t.[3] This additional application of the TA-acceptance rule only becomes relevant if an unsuccessful individual is ought to be replaced by an *Averaged Idol*.[4]

The threshold sequence $\{T_t\}$ has a strong influence on the performance of the HHA, since it determines the tolerance towards impairments. Hence, the numerical values should consider the "typical" difference in the objective function value that is caused by a search step, which is, in turn, dependent on the step size U_t. Following Fang and Winker (1997) a data driven choice of the threshold sequence is applied. This is done by initially carrying out $1,000$ search steps (for each step size) in the fashion of the *Modification Phase* with the only difference being that all new solutions are accepted without reservation. Next, an empirical distribution for each step size is obtained by the absolute differences between consecutively computed objective function values, denoted $f(|\Delta F|_t)$. Then, the threshold values are defined as quantiles of the empirical distributions that result from different step sizes. In order to gradually decrease the tolerance towards impairments and promote a greedy search, decreasing quantiles are taken and the least value is set to $T_{thresh} = 0$.

2.2   Robustification: Tütüncü and Koenig approach

Tütüncü and Koenig (2004) (TK) attempt to capture the uncertainty regarding the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in their uncertainty sets S_{μ}^{TK} and S_{Σ}^{TK} by carrying out the following three steps. Firstly, the historical data sample is bootstrapped, e.g., $1,000$ times, with a moving block bootstrap-procedure (MBB)[5] (Efron and Tibshirani 1993). Secondly, the means $\hat{\boldsymbol{\mu}}_s$ and covariance matrices $\hat{\boldsymbol{\Sigma}}_s$ of these bootstrap samples are computed. Thus, $1,001$ point estimators are gained. Thirdly, based on the $1,001$ mean vectors, S_{μ}^{TK} is defined in such a way that it includes independently for each asset i a choice of the middle $(1-\alpha)100$ percent of all $\hat{\mu}_{i,s}$. In the same component

---

[3] For the problem at hand, empirical tests indicate that the additional application of the TA-acceptance rule in the *Replacement Phase* leads to superior results compared to both versions with a certain replacement and versions that only replace individuals in the case of an improvement (results mot reported, but available on request).

[4] The alternative to an *Averaged Idol*, i.e a *Clone* of one of the most successful individuals, definitely exhibits a better objective function value and replaces the unsuccessful solution with certainty.

[5] Although the TK-approach ignores certain dependencies between the assets (as will be explained), for simplicity reasons, this MBB-procedure is used in all bootstrap applications due to its capability of capturing possible (auto)correlations within the historical return data.

wise manner, a choice of the middle $(1-\alpha)100$ percent of the $1,001$ drawings available for each component in $\hat{\boldsymbol{\Sigma}}_s$ defines S_Σ^{TK}.

With the objective of constructing a portfolio that exhibits good characteristics for many possible scenarios of the point estimators, the portfolio is optimized under the uncertainty sets' worst-case-scenarios. Due to the no-short-selling constraint in (2), the worst-case-expected return $\hat{\boldsymbol{\mu}}_W^{TK}$ possible in S_μ^{TK} is given by the $(\alpha/2)100$ percent quantile independently for each asset i. It is important to notice that this procedure ignores all correlations among the K expected returns. This is problematic, since exactly these correlations will often avoid a simultaneous occurrence of the worst-case-situation for all assets. Even more problematic is the component wise construction of the worst-case-covariance matrix $\hat{\boldsymbol{\Sigma}}_W^{TK}$. Because short-selling is forbidden, the worst-case for each (co)variance is given by the largest value in S_Σ^{TK}, i.e. by the $(1-\alpha/2)100$ percent quantile of the corresponding positions' entries in the $1,001$ covariance matrices $\hat{\boldsymbol{\Sigma}}_s$. Due to the picking of single components from different covariance matrices, there is no assurance that $\hat{\boldsymbol{\Sigma}}_W^{TK}$ is positive definite.

The constructed parameters $\hat{\boldsymbol{\mu}}_W^{TK}$ and $\hat{\boldsymbol{\Sigma}}_W^{TK}$ are then used as the inputs for problem (2), which is optimized with the HHA. The outcome is a stochastic approximation of the global maximum represented by the elitist's (robust) portfolio I_*^{TK}, i.e. the best (robust) portfolio found.

2.3 Robustification: Ceria and Stubbs approach

Ceria and Stubbs (2006) (CS) only consider the uncertainties regarding the expected returns and neglect the uncertainties regarding the covariance matrix. Their reasoning for this procedure is the finding from Chopra and Ziemba (1993) that cash-equivalent losses due to errors in estimates of expected returns are an order of magnitude greater than those for errors in estimates of variances or covariances. S_μ^{CS} is constructed as a K-dimensional ellipsoid that defines a region which envelopes the *joint deviation* of the estimator $\hat{\boldsymbol{\mu}}$ from its true value $\boldsymbol{\mu}$ with a given confidence level $1-\alpha$:

$$(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \kappa_{(1-\alpha),K}^2 \tag{3}$$

In expression (3), $\boldsymbol{\Omega}$ represents the $(K \times K)$-covariance matrix of the expected returns and $\kappa_{(1-\alpha),K}^2$ the inverse cumulative distribution function value of a chi-squared distribution with K degrees of freedom and level of significance α. The worst-case-scenario in the CS-approach is defined by the maximum *joint deviation*, i.e. by the maximum deviation of the true return from its estimator that theoretically can occur within ellipsoid (3). Thus, it is given

7

by solving the Lagrangian:

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_W^{CS} &= arg \max_{\boldsymbol{\mu}\ \theta} L(\boldsymbol{\mu}, \theta) \\
&= arg \max_{\boldsymbol{\mu}\ \theta} \left\{ \boldsymbol{w}'\hat{\boldsymbol{\mu}} - \boldsymbol{w}'\boldsymbol{\mu} - \frac{\theta}{2}\left[(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) - \kappa_{(1-\alpha),K}^2 \right] \right\} \\
&= \hat{\boldsymbol{\mu}} - \sqrt{\frac{\kappa_{(1-\alpha),K}^2}{\boldsymbol{w}'\boldsymbol{\Omega}\boldsymbol{w}}}\boldsymbol{\Omega}\boldsymbol{w}
\end{aligned}
\tag{4}
$$

Equation (4) shows that the $(K \times 1)$-mean vector $\hat{\boldsymbol{\mu}}$ is component wise penalized in such a way that the larger an asset's portfolio weight $\boldsymbol{w}$ is, the greater the asset's penalty becomes. Due to the penalty's positive dependence on $\boldsymbol{w}$ it (partly) compensates the *error-maximizing* characteristic of MVO-portfolios. Moreover, considering the expected returns' correlations through $\boldsymbol{\Omega}$ when constructing the worst-case-scenario is reasonable.[6] Of course, $\boldsymbol{\Omega}$ is not known and has to be estimated with the given data sample. For a few types of return estimators $\hat{\boldsymbol{\mu}}$, Stubbs and Vance (2005) give suggestions. In the case of stationary returns and historical means for $\hat{\boldsymbol{\mu}}$ one can use $\hat{\boldsymbol{\Omega}}^{CS} = T^{-1}\hat{\boldsymbol{\Sigma}}$.[7] Assigning a probability $1 - \alpha$ with which the ellipsoid envelopes the true expected return-vector is only valid if the return distribution is elliptical (see, e.g., Fang *et al.*, 1990) so that $\kappa_{(1-\alpha),K}^2$ actually has its attributed meaning.

Due to the constraint that limits the number of different assets held to K_{max}, equation (4) has to be adjusted to the dimensions of the considered individual I_p, $p = 1, ..., Pop$. As before, this is denoted with an index I:

$$
\hat{\boldsymbol{\mu}}_{I,W}^{CS} = \hat{\boldsymbol{\mu}}_{0,I} - \sqrt{\frac{\kappa_{(1-\alpha),K_{max}^{ef}}^2}{\boldsymbol{w}_I'\hat{\boldsymbol{\Omega}}_I^{CS}\boldsymbol{w}_I}}\hat{\boldsymbol{\Omega}}_I^{CS}\boldsymbol{w}_I
\tag{5}
$$

During optimization of problem (2) with the HHA, the penalized return (5) is applied for each of the *Pop* individuals whenever their fitness is computed. The elitist I_*^{CS} defines the robustly optimized portfolio of the CS-approach.

2.4 Robustification: An Extension of the CS-approach

We enrich the robust optimization techniques with an approach that upgrades the technique of Ceria and Stubbs (2006) by an uncertainty set for

---

[6]Assets with larger (co)variances will c.p. be penalized stronger and vice versa. Therefore, unlike in the TK-approach, an independent simultaneous occurrence of the worst-case-situation for all assets will be prevented.

[7] T denotes the amount of return observations used to estimate $\hat{\boldsymbol{\Sigma}}$.

the covariance matrix $\mathbf{\Sigma}$. Our *extended* model is denoted as ECS-approach. Furthermore, while S_μ^{ECS} is also constructed with an ellipsoid similar to (3), its components are computed in a refined way. Firstly, as is shown in *Algorithm 1*, the estimator for the matrix $\mathbf{\Omega}$ is generated using the MBB-technique as follows: the historical return sample is bootstrapped $5,000$ times to allow for the computation of an identical number of $\hat{\boldsymbol{\mu}}$-vectors (1: to 3:). The gained sequences are then used to calculate $\hat{\mathbf{\Omega}}^{ECS}$ (4:) out of which only those components are written into $\hat{\mathbf{\Omega}}_I^{ECS}$ that correspond to the assets held by an individual I (5:).

Algorithm 1 ECS-Covariance Matrix for Expected Returns.

1: Generate $5,000$ return samples s with the MBB-technique
2: Compute the expected return vectors $\hat{\boldsymbol{\mu}}_s$, $s = 1, ..., 5,000$
3: Build K expected return sequences $\{\mu_{1,s}\}, ..., \{\mu_{K,s}\}$, each of length $5,000$
4: Estimate a covariance matrix $\hat{\mathbf{\Omega}}^{ECS}$ from these sequences
5: $\hat{\mathbf{\Omega}}_I^{ECS}$ consists of only those components that correspond to the assets held

---

Secondly, as is shown in *Algorithm 2*, the ellipsoid's size is also determined using the MBB-technique. To this end the historical return sample (with mean-vector $\hat{\boldsymbol{\mu}}_0$) is bootstrapped $10,000$ times to allow for the computation of an identical number of mean-vectors $\boldsymbol{\mu}_s$, $s = 1, ..., 10,000$ (1: to 2:).[8] Out of each $(K \times 1)$-vector $\boldsymbol{\mu}_s$, K_{max} assets are randomly picked and written into $\boldsymbol{\mu}_{s,I}$ (4:). These $\boldsymbol{\mu}_{s,I}$ can be interpreted as the true expected return vectors

---
Algorithm 2 Size of the ECS-Ellipsoid for Expected Returns.

1: Generate $10,000$ return samples s with the MBB-technique
2: Compute mean vectors $\boldsymbol{\mu}_s$, $s = 1, ..., 10,000$
3: **for** $s = 1$ to $10,000$ **do**
4:     Choose randomly K_{max} assets from the $(K \times 1) - vector$ $\boldsymbol{\mu}_s$ that define $\boldsymbol{\mu}_{s,I}$
5:     Insert $\boldsymbol{\mu}_{s,I}$ with the corresponding assets' $\hat{\mathbf{\Omega}}_I^{ECS}$ and $\hat{\boldsymbol{\mu}}_{0,I}$ into ellipsoid (3)
6:     Obtain the joint deviation τ_s
7: **end for**
8: Determine the $(1 - \alpha)100$ percent quantile $f_{1-\alpha}(\tau)$

---

and must be inserted together with their estimators $\hat{\boldsymbol{\mu}}_{0,I}$ as well as with $\hat{\mathbf{\Omega}}_I^{ECS}$ into an ellipsoid such as (3) to obtain $10,000$ *joint* deviations τ_s (5: to 6:). The $(1 - \alpha)100$ percent quantile of the generated distribution function $f(\tau)$

---
[8] *Algorithm 2* generates a larger number of bootstrap samples compared to *Algorithm 1*, due to the latter algorithm's memory requirements of the procedures in line (4:).

is then used to replace $\kappa^2_{(1-\alpha),K^{ef}_{max}}$. Thus, the worst-case-return is given by:

$$\hat{\boldsymbol{\mu}}_{I,W}^{ECS} = \hat{\boldsymbol{\mu}}_{0,I} - \sqrt{\frac{f_{1-\alpha}(\tau)}{\boldsymbol{w}_I'\hat{\boldsymbol{\Omega}}_I^{ECS}\boldsymbol{w}_I}}\hat{\boldsymbol{\Omega}}_I^{ECS}\boldsymbol{w}_I \tag{6}$$

The usage of $f_{1-\alpha}(\tau)$ rather than $\kappa^2_{(1-\alpha),K^{ef}_{max}}$ has the advantage of not having to make assumptions regarding the asset return distribution. However, randomly choosing K_{max} out of K assets as well as the application of the MBB-technique for generating both $\hat{\boldsymbol{\Omega}}^{ECS}$ and $\boldsymbol{\mu}_s$ lead to stochastic quantiles $f_{1-\alpha}(\tau)$.[9]

The crucial extension in the ECS-approach is the construction of an uncertainty set for $\boldsymbol{\Sigma}$. Analogous to the returns, the portfolio risk's worst-case is defined by the maximum *joint* deviation of the portfolio's true variance $\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$ from its estimator $\boldsymbol{w}'\hat{\boldsymbol{\Sigma}}\boldsymbol{w}$ that theoretically can occur within an ellipsoid. Again this ellipsoid is considered through a constraint in a Lagrangian. Hence, the portfolio's worst-case covariance matrix is given by:[10]

$$\hat{\boldsymbol{\Sigma}}_{I,W}^{ECS} = arg \max_{\boldsymbol{\Sigma}_I \; \theta} \left\{ \boldsymbol{w}_I'\boldsymbol{\Sigma}_I\boldsymbol{w}_I - \boldsymbol{w}_I'\hat{\boldsymbol{\Sigma}}_I\boldsymbol{w}_I \right.$$
$$\left. - \frac{\theta}{2}\left[(\boldsymbol{\Sigma}_I - \hat{\boldsymbol{\Sigma}}_I)'\boldsymbol{\Theta}_I^{-1}(\boldsymbol{\Sigma}_I - \hat{\boldsymbol{\Sigma}}_I) - \Phi \right] \right\} \tag{7}$$

By using the definitions $\boldsymbol{\eta}_I = vec(\boldsymbol{\Sigma}_I)$, $\boldsymbol{W}_I = 2(\boldsymbol{w}_I\boldsymbol{w}_I') - dg(\boldsymbol{w}_I\boldsymbol{w}_I')$ as well as $\boldsymbol{\omega}_I = vec(\boldsymbol{W}_I)$ expression (7) can be rewritten as:[11]

$$\hat{\boldsymbol{\Sigma}}_{I,W}^{ECS} = vec^{-1}\left(arg \max_{\boldsymbol{\eta} \; \theta} \left\{ \boldsymbol{\omega}_I'\boldsymbol{\eta}_I - \boldsymbol{\omega}_I'\hat{\boldsymbol{\eta}}_I \right.\right.$$
$$\left.\left. - \frac{\theta}{2}\left[(\boldsymbol{\eta}_I - \hat{\boldsymbol{\eta}}_I)'\boldsymbol{\Theta}_I^{-1}(\boldsymbol{\eta}_I - \hat{\boldsymbol{\eta}}_I) - \Phi \right] \right\} \right)$$
$$= vec^{-1}\left(arg \max_{\boldsymbol{\eta} \; \theta} L(\boldsymbol{\eta}_I, \theta) \right)$$
$$= \hat{\boldsymbol{\Sigma}}_{I,0} + vec^{-1}\left(\sqrt{\frac{\Phi}{\boldsymbol{\omega}_I'\boldsymbol{\Theta}_I\boldsymbol{\omega}_I}}\boldsymbol{\Theta}_I\boldsymbol{\omega}_I \right). \tag{8}$$

---

[9]Empirical tests show that the variation in the distribution functions and their quantiles $f_{1-\alpha}(\tau)$ is negligible when they are based on $10,000$ bootstrapped values τ_s.

[10]The expression is already considering that an individual can only hold $K_{max} < K$ assets, which is, as usual, denoted by the index I.

[11]Here, the $vec(\cdot)$-operator is columnwise stacking the components within the lower triangle of a matrix to a vector. A $(K \times K)$-matrix $\boldsymbol{A}$ results in the $([K^2 + K]/2 \times 1)$-vector $vec(\boldsymbol{A})$. The vec^{-1}-operator reverses this operation and restores the original matrix $\boldsymbol{A}$. The $dg(\cdot)$ operator sets all elements off the main diagonal to a value of zero.

In equations (7) and (8), $\boldsymbol{\Theta}_I$ represents the covariance matrix of the components in vector $\boldsymbol{\eta}_I$, i.e., the covariance matrix of the $[(K_{max}^{ef})^2 + K_{max}^{ef}]/2$ asset returns' (co)variances. The estimator $\hat{\boldsymbol{\Theta}}$ is generated by applying the MBB-technique analogously to generating $\hat{\boldsymbol{\Omega}}^{ECS}$ (see *Algorithm 1*). Also, the ellipsoid's size Φ has to be determined via MBB, since the distribution of $\hat{\boldsymbol{\eta}}_I$ is unknown. To this end, as depicted in *Algorithm 3*, the historical return sample (with the stacked covariance matrix $\hat{\boldsymbol{\eta}}_0$) is bootstrapped $10,000$ times and the same number of stacked covariance matrices $\boldsymbol{\eta}_s$, $s = 1, ..., 10,000$, is computed (1: to 2:). Out of each $\boldsymbol{\eta}_s$-vector those components are written into $\boldsymbol{\eta}_{I,s}$ that correspond to K_{max} randomly chosen assets (4:). Together with

---

Algorithm 3 Size of the ECS-Ellipsoid for Covariance Matrices.

1: Generate $10,000$ return samples s with the MBB-technique
2: Compute the stacked covariance matrices $\boldsymbol{\eta}_s$, $s = 1, ..., 10,000$
3: **for** $s = 1$ to $10,000$ **do**
4:     Define $\boldsymbol{\eta}_{s,I}$ by only those components from $\boldsymbol{\eta}_s$ that correspond to K_{max} randomly chosen assets
5:     Insert $\boldsymbol{\mu}_{s,I}$ with the corresponding assets' $\hat{\boldsymbol{\Theta}}_I^{ECS}$ and $\hat{\boldsymbol{\eta}}_{0,I}$ into an ellipsoid as in (7)
6:     Obtain the joint deviation ϕ_s
7: **end for**
8: Determine the $(1 - \alpha)100$ percent quantile $f_{1-\alpha}(\tau)$

---

their corresponding $\hat{\boldsymbol{\Theta}}_I$ these $10,000$ $\boldsymbol{\eta}_{I,s}$ as well as $\hat{\boldsymbol{\eta}}_{I,0}$ are inserted into the ellipsoid to compute $10,000$ *joint* deviations ϕ_s (5: to 6:). The $(1 - \alpha)100$ percent quantile of the generated distribution function $f(\phi)$ is then used as the ellipsoid's size, i.e., $\Phi = f_{1-\alpha}(\phi)$ (8:).[12] The penalized covariance matrix is given by:

$$\hat{\boldsymbol{\Sigma}}_{I,W}^{ECS} = \hat{\boldsymbol{\Sigma}}_{I,0} + vec^{-1}\left(\sqrt{\frac{f_{1-\alpha}(\phi)}{\boldsymbol{\omega}_I' \hat{\boldsymbol{\Theta}}_I \boldsymbol{\omega}_I}} dg(\hat{\boldsymbol{\Theta}}_I)\boldsymbol{\omega}_I\right) \tag{9}$$

Unlike in equations (5) and (6), in equation (9) it cannot be assumed that the weighted sum of the components in a covariance matrix' row is positive.[13] In order to ensure $\hat{\boldsymbol{\Sigma}}_{I,W}^{ECS}$ to be a penalized covariance matrix, only the main diagonal components of $\hat{\boldsymbol{\Theta}}_I$, i.e., the variances of the returns' (co) variances, are used. The interpretation of equation (9) is as follows: the greater a (co)variance's variance, the more the historically estimated (co)variance will

---

[12] Empirical tests indicated that the quantity of bootstrap samples was sufficient to ensure a small enough variation.

[13] In $\hat{\boldsymbol{\Omega}}_I^{EC}$ and $\hat{\boldsymbol{\Omega}}_I^{ECS}$ the variances can be assumed to be greater than the (absolute) covariances, since in the empirical application daily stock returns were used.

be penalized. Also, due to a positive dependence of the penalization on the weights, the *error-maximization* is (partly) compensated. Both the penalized return (6) and the covariance matrix (9) are applied for each of the *Pop* individuals whenever their fitness is computed. The elitist I_*^{ECS} defines the robustly optimized portfolio in the ECS-approach.

3   Computational results

For the empirical analysis, firstly, several parameter values in problem (2) have to be set. Therefore, an investor is assumed whose utility function corresponds to $\lambda = 0.6$. She has an endowment of $V = 1,000,000$ Euro to be invested into a maximum of $K_{max} = 7$ stocks that are constituents of the German index DAX100 on March 16th, 2006.[14] Stocks can be purchased for their previous closing price with costs $c_f = 10$ Euro and $c_v = 0.005$. The parameters $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\Sigma}}_0$ are based on 250 daily log-returns. The results are transformed into monthly values, since the investor is assumed to rebalance the portfolio once a month. The value α determines the most extreme parameter values that are still included in the uncertainty sets. The smaller α is, the greater an uncertainty set will be, and thus the greater the worst-case estimation errors will be. Hence, α can be interpreted as a parameter that captures the investor's tolerance for estimation errors. We assume the investors utility function to correspond to $\alpha = 0.05$.

Secondly, the HHA has to be parameterized. The minimum step size is set to $U_{min} = 0.0004$ corresponding to 400 Euro. The maximum step size is set to $U_{max} = 0.3$ to ensure the trial of enough asset combinations for the case $K_{max} = 7$. Through extensive empirical testing the following parameterizations have been found to result in a sufficiently reliable stochastic outcome: $Pop = 100$, $thresh = 30$, $iter = 15$, and $steps = 8$.[15] Hence 360,100 objective function values are computed.

Thirdly, to evaluate a portfolio's performance with respect to its robustness, a moving time window procedure is implemented. After the portfolio is optimized and held for 21 subsequent out-of-sample trading days, the window of 250 trading days is moved forward by 21 days. Then a new optimization is run and the resulting portfolio is again held for 21 out-of-sample trading days before the window is moved again and so forth.[16] The sample spans

---

[14]Two firms were removed from the sample due to missing data.

[15]9,000 portfolios were optimized testing a large spectrum of possible parameter settings (results not reported, but available on request). The list of parameters also contains: $\pi = 15$, $\epsilon = 10$, $p_c = 0.7$, and $p_r = 1$ (see Appendix 1).

[16]Whenever a window is moved so that the samples' observations are updated, new

from March 17th, 2005 to January 21st, 2008, allowing for the construction of 23 portfolios.[17]

Table 1: Out-of-sample performance

	objective function value				portfolio return				portfolio risk			
win	MVO	TK	CS	ECS	MVO	TK	CS	ECS	MVO	TK	CS	ECS
1	-3.04	-3.30	-3.14	-2.62	-0.25	-4.60	-0.75	-2.07	4.89	2.44	4.73	2.99
2	-4.05	-3.95	-4.29	-4.05	4.00	-3.09	2.67	-1.23	9.42	4.53	8.93	5.93
3	-14.48	-6.70	-12.51	-8.01	-14.28	-6.65	-11.49	-8.30	14.62	6.73	13.19	7.82
4	-3.39	-2.25	-2.71	-2.25	-2.76	-0.89	-1.11	-0.56	3.80	3.15	3.81	3.38
5	-1.19	-0.96	-1.19	-0.94	2.70	2.07	2.65	1.96	3.78	2.98	3.76	2.87
6	-1.79	-1.50	-1.74	-1.48	-2.15	-1.48	-1.98	-1.47	1.56	1.52	1.58	1.50
7	-0.63	-0.53	-0.61	-0.40	2.01	1.56	2.07	1.91	2.40	1.92	2.40	1.94
8	-0.64	-0.61	-0.64	-0.35	1.65	1.40	1.63	2.93	2.17	1.95	2.15	2.54
9	-0.60	-0.62	-0.23	-1.18	2.49	1.37	2.65	0.29	2.66	1.94	2.14	2.16
10	-2.20	-0.57	-2.23	-1.35	0.75	2.16	0.65	1.42	4.16	2.38	4.11	3.19
11	-1.25	-1.19	-1.24	-0.55	1.14	0.91	0.86	2.46	2.85	2.59	2.64	2.56
12	-4.08	-2.90	-3.70	-2.82	-3.66	-2.85	-3.43	-2.40	4.36	2.93	3.87	3.10
13	-0.54	-0.44	0.05	0.00	3.43	3.42	4.21	5.07	3.19	3.01	2.72	3.37
14	-2.83	-2.86	-3.43	-2.07	1.57	1.46	3.10	-1.91	5.76	2.73	7.78	2.18
15	-2.05	-1.73	-2.04	-2.07	-0.44	-0.18	-0.92	-0.82	3.12	2.76	2.79	2.90
16	-1.27	-1.20	-1.21	-0.75	2.72	0.17	2.48	2.06	3.93	2.11	3.68	2.62
17	-8.79	-6.24	-8.60	-6.84	-10.82	-9.69	-11.55	-8.76	7.44	3.95	6.64	5.56
18	-3.66	-1.54	-1.52	-2.93	0.01	1.67	2.64	0.27	6.10	3.69	4.30	5.07
19	1.12	1.09	0.93	0.89	6.52	5.51	5.91	6.06	2.47	1.86	2.39	2.55
20	-1.68	-2.86	-2.34	-2.71	4.97	-2.47	0.64	0.07	6.11	3.12	4.32	4.56
21	-8.72	-4.70	-7.43	-5.84	-10.74	-3.80	-9.41	-6.74	7.38	5.30	6.11	5.23
22	-4.85	-2.63	-4.45	-2.93	-5.65	-2.09	-4.84	-2.89	4.32	2.99	4.18	2.96
$\bar{x}$	-3.21	-2.19	-2.92	-2.33	-0.76	-0.73	-0.61	-0.57	4.84	3.16	4.46	3.50
$\bar{\mu}_p^{ef}$	-	-	-	-	-0.61	-0.44	-0.41	-0.40	-	-	-	-
σ_x	3.42	1.88	3.05	2.17	5.26	3.41	4.73	3.70	2.87	1.32	2.64	1.51

Table 1 summarizes for all approaches the actual objective function value, portfolio return, and portfolio risk, measured as the return volatility. In each column $\bar{x}$ represents the mean and σ_x the standard deviation. Since problem (2) is optimized for each time window win as if there would not already exist a portfolio, the reduction of transaction costs due to holding similar portfolios in subsequent periods is ignored. The return that approximately considers the reduced costs is given by $\bar{\mu}_p^{ef}$.

Table 1 shows the out-of-sample performance of the MVO and all robust approaches, in which the former serves as a benchmark. It can be seen that the mean risks are lower and the mean returns are higher for all three robust approaches compared to the MVO-approach. In addition, smaller variations around these better mean values can be observed. Among the robust approaches, on average TK- and ECS-portfolios exhibit the lowest risks, whereas ECS- and CS-portfolios exhibit the highest returns. The fact that also the highest return is on a very low absolute level is mainly caused by using historical means as estimators and by the conservative choice of the transaction costs.[18] However, since our aim is to examine portfolios

threshold sequences, parameters, and distributions (when applicable) must be computed for each approach.

[17]All used return series were shown to be stationary according to the ADF-test as well as the KPSS-test.

[18]The construction of portfolios based on estimators with such a limited forecast quality

regarding their robustness-properties, a low absolute level of returns that all portfolios equally "suffer" from might be of minor importance. The robust approaches' advantages become most apparent when investigating actual bad-case-scenarios, e.g., periods 3, 17, and 21. While, e.g., in period 3 the TK-approach achieves a return that is twice as high and a risk that is half as high as that of the MVO-approach, in actual good-case-scenarios, e.g., period 19, the return is only a little smaller while the risk is still lower. A similar tendency can be observed for all robust approaches.[19]

Table 2: MVO-approach

win	fluctuation of composition				portfolio weights of stock No.					forecast errors		
	TraS	ToR	IdS	RoF	7	30	34	48	95	e_F	e_μ	e_σ
1	-	-	-	-	-	-	-	-	-	-3.20	-7.86	0.09
2	37,477	0.31	5(7)	1.08	-	-	-	-	-	-4.43	-4.86	4.15
3	32,739	0.37	5(7)	2.73	22.85	-	-	-	-	-14.22	-20.80	9.84
4	75,427	0.74	2(7)	19.91	57.20	-	-	-	-	-2.33	-6.40	-0.39
5	31,570	0.28	4(7)	2.27	50.43	-	5.18	8.26	-	0.02	-0.10	-0.09
6	13,724	0.14	6(7)	1.75	54.80	-	3.17	7.25	-	-0.75	-5.62	-2.50
7	25,894	0.23	5(7)	4.13	46.98	9.79	7.10	10.67	-	0.47	-0.81	-1.32
8	19,794	0.26	4(7)	0.40	46.50	9.62	7.61	10.23	-	0.13	-1.47	-1.19
9	12,205	0.16	6(7)	2.97	47.59	16.53	10.11	7.39	-	0.20	-0.65	-0.77
10	11,896	0.13	5(7)	1.18	46.61	14.71	10.29	6.41	-	-1.38	-2.12	0.89
11	14,631	0.26	4(7)	3.61	42.34	20.87	11.53	9.19	-	-0.38	-1.73	-0.52
12	6,473	0.10	6(7)	1.87	43.00	16.20	12.06	10.56	-	-3.04	-5.87	1.16
13	16,142	0.29	4(7)	6.31	56.41	11.42	5.63	-	-	0.69	2.19	0.31
14	19,031	0.19	4(7)	1.41	56.60	8.69	-	-	-	-1.73	0.28	3.07
15	21,259	0.43	4(7)	7.73	37.65	14.11	-	-	10.32	-1.09	-2.68	0.03
16	22,217	0.62	3(7)	3.02	-	-	-	-	13.60	-0.52	-0.80	0.34
17	16,868	0.42	4(7)	2.84	-	-	-	-	18.08	-8.46	-16.23	3.28
18	30,358	0.66	3(7)	9.55	-	-	-	-	36.63	-2.79	-4.87	1.40
19	28,048	0.45	3(7)	6.99	-	-	-	-	32.72	1.99	2.90	-1.39
20	10,475	0.26	5(7)	4.81	-	-	-	-	41.05	-0.91	0.16	1.63
21	6,169	0.39	5(7)	3.24	-	-	-	-	38.36	-7.85	-16.15	2.31
22	16,121	0.39	4(7)	3.92	-	-	-	-	32.46	-3.60	-9.53	-0.35
23	16,404	0.43	4(7)	3.64	-	-	-	-	26.41	-	-	-
	a)22.042	0.34	4.32	4.33	b)34.36	12.18	8.89	4.27	30.73	f)-2.42	-4.68	0.91
					c)9.06	3.78	2.92	1.55	10.62	g)-3.54	-6.03	h)2.19
					d)31			e)19.97		i)5.05	8.44	j)3.36

Table 2 summarizes the results of the MVO-approach. The values within the framed line $a)$ show mean values; lines $b)$ and $c)$ show the full range of fluctuation and the standard deviation (each in percentage points) of the assets' portfolio weights. All displayed stocks were held for at least six subsequent periods. The number of stocks held for a minimum of two subsequent periods is given by value $d)$, whereas $e)$ is the averaged ratio of the weight's standard deviation resulting from the sequences $d)$ and the corresponding weight's mean (in percent); lines $f)$, $g)$ and $i)$ show mean forecast errors, mean negative (in $h)$ positive) forecast errors, and root mean negative (in $j)$ positive) squared forecast errors.

To gain further meaningful insight, Table 2 exhibits the portfolio compositions as well as the forecast errors. There, $TraS$ (traded stocks) shows the traded volume measured in pieces of stocks and ToR (turnover ratio) is

apparently contributes only little to good out of sample properties. Instead, this procedure solely contributes to a construction of portfolios that exhibit good in-sample risk-return characteristics.

[19]This can also be gathered from Figure 1 in Appendix 2.

the priced traded volume relative to the twofold capital endowment, both in period win compared to $win - 1$; the number of identical stocks in these subsequent periods is given by IdS (K_{max}^{ef} is shown in parenthesis) and RoF is their averaged range of fluctuation. It is obvious that the MVO-approach exhibits an overall high fluctuation in the portfolio composition. On average 22,042 pieces of stocks are traded to rebalance the portfolio in each period. It is, however, noticeable that four out of the seven stocks are simultaneously held in periods $win \in [7; 12]$, for which reason this might be considered as a more stable period, as can be seen by all variables; e.g., the average turnover ratio is only about half as high and $TraS$ exhibits an average value of only 12,985 in that period.

The last three columns of Table 2 show the differences between actually realized and predicted objective function values (multiplied by 100), portfolio returns, and portfolio volatilities. An examination of the portfolio returns' forecast errors shows that the actual (out-of-samples) performance is worse than its expectation in 18 out of 22 periods. The return was on average overestimated by 4.68 percentage points, as is shown by the mean forecast error. A mean overestimation of 6.03 percentage points (as shown by the mean negative forecast error) as well as the root mean negative squared forecast error of 8.44 provide more detail about the extent of the return overestimation. The corresponding results for the portfolio risk are, as expected, better. Nevertheless, the mean forecast error also points in the unfavorable direction, i.e., the portfolio risk is on average underestimated by almost one percentage point. Due to $\lambda = 0.6$ the objective function's mean forecast error is between that of the return and the risk. However, since both underestimating risk and overestimating return increases the objective function's forecast error, on average no error compensation takes place. The average actual return (risk) of -0.76 (4.84) percent is by the magnitude of the displayed errors lower (higher) than its expected average value of 3.92 (3.93) percent.

The results of the TK-approach are shown in Table 3. Although the return predictions improved compared to the MVO-approach, there is still on average, despite optimizing for the theoretical worst-case-scenario, an overestimation of 0.23 percentage points present. In contrast, the risk's mean forecast error of -0.48 percentage points shows that on average the risk is no longer underestimated. The surprising observation that the expected risk of the MVO-portfolios in some periods is greater than that of the TK-portfolios (see Figure 1 in Appendix 2) can be explained as follows: compared to the expected (worst-case) returns that are mostly close to zero, the expected (co)variances are high. In addition, by weighting the risk heavier through $\lambda = 0.6$, the objective function value is greatly determined by the portfolio risk. Thus, an implicit movement towards the minimum-variance-portfolio

(MVP) takes place. As can be seen in periods 14, 19, and 20, the portfolio is only diversified among six stocks, which must be seen critically and is originally caused by many negative returns in the worst-case-scenario. However, the aforementioned mechanism contributes to portfolio diversification so that not solely stocks with positive returns are picked due to the implicit risk-overweighing.[20] Yet whenever the return's norm is too large, this mechanism fails such as in periods 14, 19, and 20. The movement toward the

Table 3: TK-approach

win	fluctuation of composition				portfolio weights of stock No.					forecast errors		
	$TraS$	ToR	IdS	RoF	7	12	30	48	64	e_F	e_μ	e_σ
1	-	-	-	-	-	-	-	-	-	-0.94	-5.92	-1.40
2	17,684	0.35	4(7)	2.43	-	-	-	-	-	-2.56	-4.24	0.72
3	21,942	0.46	4(7)	4.36	23.78	-	-	-	-	-6.12	-6.74	3.15
4	28,768	0.56	4(7)	14.58	54.86	-	10.30	4.21	-	0.53	-0.74	-1.33
5	14,296	0.17	6(7)	3.11	44.50	-	9.42	7.31	-	1.86	3.16	-1.19
6	6,832	0.11	7(7)	3.04	54.00	-	9.25	8.34	-	1.29	-1.02	-2.84
7	6,200	0.12	6(7)	2.65	49.92	-	10.12	13.37	-	2.29	1.97	-2.32
8	14,223	0.17	5(7)	1.43	51.63	-	12.42	13.44	8.32	2.00	1.52	-2.03
9	5,479	0.09	7(7)	2.47	49.25	-	17.88	10.34	11.41	1.12	1.77	-1.98
10	5,545	0.10	6(7)	1.86	49.47	-	16.49	10.31	6.42	0.98	2.55	-1.48
11	13,674	0.18	5(7)	1.66	48.58	-	18.79	8.34	5.08	1.74	1.52	-1.08
12	5,601	0.09	7(7)	2.69	54.10	-	12.43	10.21	4.21	-0.67	-1.72	-0.49
13	9,757	0.20	6(7)	4.28	67.41	3.24	9.35	-	2.03	2.20	4.74	-0.18
14	5,708	0.12	5(6)	2.54	74.75	2.16	6.08	-	-	0.07	2.27	2.46
15	16,954	0.36	5(7)	6.92	51.22	11.40	5.16	-	5.28	-0.19	0.14	-0.57
16	5,359	0.10	5(7)	2.03	50.73	11.22	-	-	5.31	1.07	0.86	-0.65
17	6,149	0.15	4(7)	1.11	52.90	10.32	-	-	4.23	-5.55	-9.79	1.15
18	13,288	0.42	4(7)	4.46	43.66	16.31	-	-	2.34	-0.81	2.78	0.42
19	4,208	0.16	5(6)	6.04	32.12	13.80	-	-	7.61	2.88	6.13	-1.76
20	2,388	0.10	5(6)	3.17	41.03	11.54	-	-	4.34	-0.70	-1.88	-0.40
21	5,484	0.16	5(7)	3.18	40.99	8.26	-	-	4.31	-3.73	-2.34	1.93
22	9,248	0.34	6(7)	6.91	28.72	15.44	-	-	8.13	-0.41	-0.02	-0.72
23	6,384	0.19	6(7)	3.43	31.79	15.61	-	-	-	-	-	-
	[a)]10.237	0.21	5.32	3.83	[b)]50.97	6.02	13.63	9.23	9.38	[f)]-0.16	-0.23	-0.48
					[c)]11.69	2.05	4.16	2.75	2.43	[g)]-2.17	-3.44	[h)]1.64
					[d)]20			[e)]28.72		[i)]3.02	4.56	[j)]1.90

Table 3 summarizes the results of the TK-approach for $\alpha = 0.05$ with the usual key data (explained for Table 2). The weights of those stocks are listed that were held for at least eight subsequent periods.

MVP also explains the differences in the portfolio compositions between the MVO- and the TK-approach.[21] Indicated by all variables, the TK-portfolios' compositions are more stable; e.g., the turnover rate is about one third lower and the traded stocks with an average of 10,237 are about half as many as in the MVO-approach. With an average IdS of 5.32, on average one stock less is exchanged per period. In three periods no stocks get exchanged in the TK-portfolio. Stocks are generally held for longer periods, which is most apparent for stock No.7 but it can be seen also by the amount of stocks that

---

[20]On average, four stocks in the portfolio exhibit negative expected returns.

[21]Setting $\lambda = 0.8$ only in the MVO-approach creates a portfolio composition that is more similar to that of the TK-approach.

are held for at least two periods (20 versus 31 in the MVO-approach).[22]

The CS-approach (see Table 4) exhibits only minor improvements of the mean forecast error compared to the MVO-approach. On average the actual portfolio returns are still 3.15 percentage points lower than predicted. But considering a less conservative penalization of the portfolio returns, this is not surprising. Even though the CS-approach does not penalize the covariance matrix, the portfolio risk's forecast error measures are slightly better than the benchmark. Nevertheless, the necessity of an uncertainty set for the covariance matrix is apparent, e.g., in period 14, where the CS-portfolio risk exceeds that of the MVO-approach by more than two percentage points (see Table 1). As can be seen by the stocks held, the corresponding periods,

Table 4: CS-approach

	fluctuation of composition				portfolio weights of stock No.					forecast errors		
win	TraS	ToR	IdS	RoF	7	30	34	48	95	e_F	e_μ	e_σ
1	-	-	-	-	-	-	-	-	-	-2.85	-7.00	0.08
2	39,609	0.44	4(7)	1.09	-	-	-	-	-	-4.19	-4.12	4.24
3	35,850	0.46	5(7)	2.69	31.45	-	-	-	-	-11.88	-16.19	9.00
4	44,638	0.65	2(7)	10.35	49.01	-	-	9.48	-	-1.28	-3.22	-0.01
5	16,977	0.18	5(7)	2.00	47.73	-	5.08	10.35	-	0.40	0.91	-0.05
6	13,264	0.14	6(7)	1.76	52.71	-	3.32	9.40	-	-0.30	-4.37	-2.41
7	24,786	0.21	5(7)	3.68	46.14	10.38	6.96	11.48	-	0.84	0.16	-1.30
8	20,949	0.26	4(7)	0.43	46.11	9.54	6.92	10.68	-	0.44	-0.67	-1.18
9	9,589	0.09	7(7)	2.70	47.54	15.68	8.81	7.77	-	0.93	0.61	-1.15
10	8,492	0.05	6(7)	0.75	47.05	14.42	8.94	7.34	-	-1.07	-1.38	0.87
11	18,995	0.27	4(7)	3.19	41.64	18.51	10.23	9.31	-	-0.06	-0.95	-0.54
12	8,183	0.10	6(7)	2.10	43.63	13.40	9.89	10.25	-	-2.30	-4.49	0.85
13	13,606	0.24	5(7)	5.01	57.31	9.36	5.15	-	-	1.54	3.97	0.08
14	11,134	0.15	6(7)	2.89	54.03	7.22	2.04	-	-	-2.18	2.43	5.25
15	21,312	0.34	5(7)	8.01	42.09	11.35	-	-	7.72	-0.85	-2.17	-0.03
16	25,426	0.72	2(7)	4.70	-	-	-	-	11.58	-0.14	-0.02	0.22
17	14,163	0.32	5(7)	2.42	-	-	-	-	14.89	-7.88	-15.55	2.77
18	17,422	0.50	4(7)	7.77	-	-	-	-	25.96	-0.20	0.55	0.69
19	9,016	0.15	6(7)	1.93	-	-	-	-	27.29	2.17	3.73	-1.13
20	10,699	0.29	5(7)	1.63	-	-	-	-	28.64	-1.18	-1.51	0.96
21	13,268	0.48	4(7)	3.98	-	-	-	-	32.60	-6.14	-12.96	1.60
22	10,781	0.38	4(7)	1.92	-	-	-	-	28.06	-2.86	-7.14	0.00
23	8,903	0.37	5(7)	2.96	-	-	-	-	21.82	-	-	-
	[a)]18.048	0.31	4.77	3.36	[b)] 25.85	11.29	8.19	5.50	24.88	[f)]-1.78	-3.15	0.86
					[c)]6.20	3.37	2.65	1.61	8.16	[g)]-2.84	-5.45	[h)]2.05
					[d)]28			[e)]18.97		[i)]4.26	7.52	[j)]3.28

Table 4 summarizes the results of the CS-approach for $\alpha = 0.05$ with the usual key data (explained for Table 2). The weights of those stocks are listed that were held for at least eight subsequent periods.

and their weights' standard deviations, the portfolio is similar to that of the MVO-approach, even if it is slightly more stable. The turnover ratio is about ten percent, $TraS$ about 20 percent, and the range of fluctuation about one percentage point lower.

---

[22]It is to be noticed here, that these 20 stocks are held longer and that this relatively small number is not the result of more stocks being held for only one period.

Table 5 shows the ECS-portfolios, which, with a value of only -0.02, produced the lowest mean forecast error for the risk of all approaches. However, this result of exhibiting on average almost no forecast error is put into perspective when also considering the other forecast error measures for the risk. Also the portfolio return is predicted more accurately than that of both the MVO-and the CS-portfolios.

Table 5: ECS-approach

	fluctuation of composition				portfolio weights of stock No.					forecast errors		
win	TraS	ToR	IdS	RoF	7	13	16	30	95	e_F	e_μ	e_σ
1	-	-	-	-	-	-	10.26	-	-	-0.94	-2.69	-0.22
2	33,056	0.62	2(7)	6.18	-	-	-	-	-	-2.56	-3.26	2.09
3	31,413	0.57	3(7)	5.95	19.03	-	18.04	-	-	-6.12	-8.43	4.58
4	41,195	0.53	3(7)	8.10	32.74	8.93	18.77	13.91	-	0.53	0.78	-0.37
5	21,730	0.18	6(7)	2.34	29.46	10.44	17.86	9.05	-	1.86	3.63	-0.68
6	5,659	0.07	7(7)	2.07	31.24	12.52	14.55	9.49	-	1.29	-0.04	-2.18
7	8,701	0.14	6(7)	1.63	32.02	13.72	12.97	10.99	-	2.29	3.24	-1.66
8	31,635	0.31	4(7)	2.37	39.20	13.89	12.17	12.32	-	2.00	3.43	-1.06
9	26,295	0.24	5(7)	2.27	36.23	10.87	9.01	11.48	-	1.12	1.19	-1.07
10	2,434	0.04	7(7)	1.03	36.93	12.41	7.34	11.11	-	0.98	2.35	-0.07
11	7,181	0.13	6(7)	1.60	37.00	14.45	8.30	12.45	-	1.74	3.32	-0.69
12	27,550	0.34	4(7)	4.56	49.00	10.64	6.52	11.82	-	-0.67	-2.13	-0.31
13	13,111	0.24	4(7)	4.20	56.60	8.43	-	8.41	6.18	2.20	5.47	-0.02
14	11,572	0.23	4(7)	1.32	53.46	8.87	0.07	9.07	7.23	0.07	-1.59	-1.17
15	14,743	0.38	4(7)	5.13	37.94	-	-	9.68	9.25	-0.19	-0.92	-0.30
16	16,347	0.34	4(7)	4.08	28.20	0.12	-	-	11.98	1.07	1.93	-0.50
17	11,175	0.32	5(7)	3.99	28.95	-	-	-	10.48	-5.55	-10.36	2.33
18	24,820	0.71	3(7)	19.21	5.51	-	-	-	32.55	-0.81	-0.66	0.90
19	17,822	0.37	4(7)	6.46	15.88	-	-	-	25.98	2.88	5.57	-1.09
20	2,538	0.09	7(7)	2.59	6.96	-	-	-	27.38	-0.70	-0.58	0.78
21	12,269	0.33	5(7)	4.70	16.93	-	-	-	24.28	-3.73	-7.64	1.12
22	27,910	0.51	4(7)	4.88	-	-	0.17	-	15.13	-0.41	-2.14	-0.75
23	17,153	0.52	4(7)	2.13	0.24	0.10	-	-	14.37	-	-	-
	a)18.469	0.33	4.59	4.40	b) 51.09	6.02	12.25	5.50	26.37	f)-0.16	-0.43	-0.02
					c)13.60	2.05	4.42	1.61	8.70	g)-2.17	-3.37	h)1.97
					d)19		e)22.74			i)3.02	4.73	j)2.36

Table 5 summarizes the results of the ECS-approach for $\alpha = 0.05$ with the usual key data (explained for Table 2). The weights of those stocks are listed that were held for at least nine subsequent periods.

Even though most portfolio compositions' key data seem to indicate only minor improvements in the stability of the ECS-portfolios compared to MVO-portfolios, this observation, with its impact on transaction costs, is primarily caused by relatively many stocks held for only one period (not shown). Only 19 stocks were held for a minimum of two subsequent periods which is the most stable result of all approaches. Furthermore, similar to the TK-approach, in three periods (6,10, and 20) no stocks get exchanged and holding periods of the displayed stocks are relatively long. Except for period 22, stock No. 7 is held in the same periods, but with less extreme weights and a lower standard deviation as in the TK-approach. The importance of stock No. 13, that did not appear in another approach, is presumably caused by its small (co)variances' variances, which is not (satisfactorily) considered in the other approaches. Therefore stock No. 13 is penalized less heavily and is conse-

18

quently more likely to be picked than stocks that exhibit smaller historical (co)variances.[23]

To sum up, the robustification techniques lead to an improvement compared to the MVO-approach. This improvement is possible due to a reduced risk that is not necessarily accompanied by the corresponding lower returns. Among the robust approaches the TK- and the ECS-approach clearly outperform the CS-approach.

4   Conclusion

In this work different robust optimization techniques are empirically tested within a complex optimization problem that emulates a realistic investment environment. The employed hybrid heuristic algorithm is well capable of tackling the complexity of the resulting optimization problems. We find that the explicit incorporation of the uncertainty about the true (but unknown) parameters into the optimization process leads overall to superior results over the MVO-approach. The portfolio compositions are shown to be more stable and consequently lead to a reduction of the transaction costs. On average the out-of-sample portfolio risk is lower and accompanied by smaller deviations, but not necessarily lower returns. This is possible since robust portfolios exhibit an improved performance in bad-case-scenarios without necessarily a worse performance in good-case-scenarios.

Although for the used data sample the TK-approach performs slightly better than the ECS-approach, due to the formers' shortcomings of not considering expected returns' correlations, a possibly singular "covariance matrix" of returns, and a reduced diversification, it is not considered as the superior approach. The CS-approach that only uses an uncertainty set for the expected returns seems to only adjust the MVO-portfolios weights rather than it constructs a whole new composition. Therefore the CS-approach exhibits the disadvantage of limited effectiveness. Furthermore, the employed estimator for the expected returns' covariance matrix needs to rely on distributional assumptions. If this covariance matrix is generated by the MBB-technique rather than by linear transformation, on average around 8,000 out of 9,604 components are larger. This indicates that the distributional assumptions are not met with the consequence of a penalization that is too small.

The ECS-approach seems to pool some desired characteristics: (i) it offers uncertainty sets for both the expected return and the covariance matrix in

---

[23]This is not true to the same extend for stock No. 16, since it was also part of the MVO-, the TK-, and the CS-portfolios for some short holding sequences (not shown).

an intuitive way. The greater the components deviations are, the greater the uncertainty and with that the penalization will be. The penalization (ii) takes correlations into account and (iii) does not rely on distributional assumptions. The ECS-approach remains also applicable when different (expected return) estimators are employed. This might become a major interest if portfolios are based on, e.g., factor models as employed by Gilli and Roko (2008).

The application of different (expected return) estimators is a possible direction for future research. Beside factor models, also popular techniques from time series analysis seem to be promsising when incorporated in the robust portfolio optimization setup. Obviously, future research should also take further data sets into account to generalize the results found in this work.

Appendix 1 - The HHA in more Detail

The pseudo code for the HHA is given in *Algorithm 4*. In generation $g = 0$, the algorithm generates and evaluates a population of *Pop* random solutions I_p^0, $p = 1, ..., Pop$, that satisfy the constraints (2: to 3:). The solutions I_p^0 are referred to as search agents or individuals and represent portfolios that are not only component wise (stepwise) altered but are also subject to evolutionary procedures. For each step size U_t and threshold value T_t respectively, where $t = 1, ..., thresh$, a number of *iter* generations evolve. The step size U_t, which linearly decreases from U_{max} to U_{min} by ΔU when t increases and the algorithm matures (5:) (but is held constant for *iter* generations for fixed

---

Algorithm 4 Hybrid Heuristic Algorithm.

1: Initialize *Pop*, *thresh*, *iter*, U_{min}, U_{max} and $\Delta U = (U_{max} - U_{min})/thresh$
2: Generate a valid initial ($g = 0$) population of random solutions $\{I_p^0\}$, $p = 1, ..., Pop$
3: Evaluate $F(I_p^0) \forall p$ and determine elitist I_*^0
4: **for** $t = 1$ to *thresh* **do**
5:     Determine the step size $U_t = U_{max} - \Delta U \cdot (t - 1)$
6:     **for** $l = 1$ to *iter* **do**
7:         $g = (t - 1) \cdot iter + l$
8:         $\rightarrow$ **Modification Phase** *(Algorithm 5)*
9:         $\rightarrow$ **Evaluation Phase** *(Algorithm 6)*
10:       $\rightarrow$ **Replacement Phase** *(Algorithm 7)*
11:     **end for**
12: **end for**
13: terminate algorithm and report current elitist I_*

---

t), can be interpreted as a fraction of the total capital endowment V that is subject to trades between two evolutionary steps. More precisely, these

trades, which are conducted within the *Modification Phase*, adjust the individuals in a component wise manner as it is known from TA. Each generation $g = (t-1)\cdot iter + l$, $g = 0, ..., thresh\cdot iter$, undergoes two more phases, namely a so-called *Evaluation Phase* and a so-called *Replacement Phase*, which rank and recombine the individuals as it is known from evolutionary algorithms.

In the *Modification Phase*, shown in more detail in *Algorithm 5*, the agents develop independently from each other over *steps* search steps. In each of these steps, an amount of $\Delta n_{i,p}$ pieces of a randomly picked asset $i \in \mathfrak{I}_p^g$, amounting to the step size's cash equivalent, is sold (3:). Of course, $\Delta n_{i,p} = g(\cdot)$ is a (discrete) function of various parameters of the problem. If the disposition leads to a drop of the asset's weight below the lower bound, or if it leads to a complete clearance of asset $i \in \mathfrak{I}_p^g$, two proceedings exist: (i) with probability $p_{replace}$ asset i gets substituted by a random asset $k \notin \mathfrak{I}_p^g$, or (ii) with probability $1 - p_{replace}$ asset i is kept with quantity $n_{i,p} = w_{i,p} = 0$ and a random other portfolio constituent $j \in \mathfrak{I}_p^g$, $j \neq i$, is bought (4: to 6:). In contrast, if the disposition does not violate the lower bound on the weights, i is kept in the reduced (but still positive) quantity and, again, a random other portfolio constituent $j \in \mathfrak{I}_p^g$, $j \neq i$, is bought (7: to 9:). Under consideration of all constraints, the assets j and k respectively are bought from the cash that was generated by selling asset i. Hence, an altered version of portfolio I_p^g denoted by $I_p'^g$ is generated. This new solution has to be evaluated (10:) and compared to the current solution I_p^g. Whenever the objective function

---

Algorithm 5 Modification Phase (HHA).

1: Initialize p_r $\{T_t\}$,; import $\{I_p^g\}$, V, U_t
2: **for** $s = 1$ to *steps* (parallel $\forall\, p$) **do**
3:     Sell $\Delta n_{i,p} = g(U_t, V, P_{i,p}...)$ pieces of a random assets $i \in \mathfrak{I}$
4:     **if** $n_{i,p} = 0 \wedge w_{i,p} < w_l$ **then**
5:         With $p_{replace}$: buy random asset $k \notin \mathfrak{I}_p^g$ versus $i \in \mathfrak{I}_p^g$
6:         With $1 - p_{replace}$: buy a random asset $j \in \mathfrak{I}_p^g$, $j \neq i$, $i \in \mathfrak{I}_p^g \wedge n_{i,p} = w_{i,p} = 0$
7:     **else**
8:         Buy random asset $j \in \mathfrak{I}_p^g$, $j \neq i$
9:     **end if**
10:     Evaluate $F(I_p'^g)$
11:     **if** $F(I_p'^g) \geq [F(I_p^g) - T_t]$ **then**
12:         $I_p^g = I_p'^g$ & update I_*^g if necessary
13:     **end if**
14: **end for**

---

value (the fitness) is greater than that of the current solution I_p^g, or whenever the impairment is not greater than the threshold value T_t, the solution $I_p'^g$ is accepted as the new current solution I_p^g (11: to 12:). This possibility to accept impairments in the objective function value ultimately enables the

agents to escape local extrema. Hence, the threshold value can be interpreted as the tolerance towards impairments; its value decreases with an increasing t to a final value of $T_{thresh} = 0$. If an accepted solution's objective function value is also smaller than that of the best found portfolio, the so-called elitist I_*^g, is updated (12:).

After the individuals developed independently from each other, the *Evaluation Phase*, which is presented in *Algorithm 6*, applies. In that, the *Pop* individuals get sorted in an ascending order according to their objective function values (2:). A selection of promising tendencies that will be reinforced as well as the selection of unpromising tendencies that might be excluded, i.e., the evolutionary component of the HHA, relies on this order as follows. The $\pi < 0.5Pop$ best individuals, the so-called prodigies $I_{<x>}^g \in \{I_p^g\}$, $x = 1, ..., \pi$, are defined to be 'idols' for the remaining population (3:).[24] This group of idols, denoted by Π^g, is enlarged by the current generation's elitist I_*^g. Hence, the set of idols is defined by $\Pi^g = \{I_{<x>}^g, I_*^g\}$ (4:). Based on their ranks, the prodigies' portfolios are assigned linearly decreasing amplifying factors $a_{<x>}^g$, ranging from $\pi + 1$ down to 1. The elitist's amplifying factor is chosen to be ϵ (5:). Corresponding to Π^g the π worst individuals, subsequently called underdogs, are pooled in set $\Gamma^g = \{I_{<Pop-(x-1)>}^g\} \subset \{I_p^g\}$ (6:).

Algorithm 6 Evaluation Phase (HHA).

1: Initialize π, ϵ; import $\{I_p^g\}$, *Pop*
2: Rank individuals according to their fitness
3: Define prodigies $I_{<x>}^g \in \{I_p^g\}$, $x = 1, ..., \pi$
4: Enlarge the set of idols by the elitist, i.e., $\Pi^g = \{I_{<x>}^g, I_*^g\}$
5: Assign the prodigies linearly decreasing amplifying factors $a_{<x>}^g$, ranging from $\pi + 1$ down to 1; the elitist's factor is ϵ
6: Define underdogs $I_{<Pop-(x-1)>}^g \in \{I_p^g\}$, merged in the set Γ^g

---

The last phase in the life of a population's generation is the *Replacement Phase*, shown in *Algorithm 7*, in which the set of idols, in cooperation with the amplifying factors, is used to (possibly) replace the π underdogs $I_{<Pop-(x-1)>}^g$.

An underdog is replaced with probability p_{clone} by an exact copy, i.e., by a *Clone*, of a prodigy. Therefore, each prodigy within Π^g gets assigned a selection probability $p(I_{<x>}^g)$, that increases with the prodigy's fitness, i.e.,[25]

$$p(I_{<x>}^g) = \frac{f[F(I_{<x>}^g)]a_{<x>}^g}{\sum_{x=1}^{\pi} f[F(I_{<x>}^g)]a_{<x>}^g}, \quad \forall\, x \tag{10}$$

---

[24]The index $<z>$ represents the zth-best position in this sorted order.
[25]In minimization problems the function $f[\cdot]$ transforms low objective function values into high fitness values. It can be neglected in maximization problems.

Algorithm 7 Replacement Phase (HHA).

1: Import $\{I_p^g\}$, Π^g, Γ^g, π, $\{a_{<x>}^g\}$, ϵ
2: **if** Replacement by *Clone* (with p_{clone}) (parallel $\forall\, p \in \Gamma^g$) **then**
3:     Compute selection probabilities according to equation (10)
4:     Randomly choose a prodigy $I_{picked}^g \in \Pi^g$ based on $p(I_{<x>}^g)$
5:     $I_{<Pop-(x-1)>}^g = I_{picked}^g$
6: **else**
7:     Compute averaged weights according to equation (11)
8:     Randomly choose K_{max} assets $i \in \mathfrak{I}_\Pi^g$ based on $\bar{w}_i^g$ and normalize
9:     Make I_{idol}^g a valid solution and evaluate $F(I_{idol}^g)$
10:     **if** $F(I_{idol}^g) \geq [F(I_{<Pop+(x-1)>}^g) - T_t]$ **then**
11:         $I_{<Pop-(x-1)>}^g = I_{idol}^g$; update I_*^g if necessary
12:     **end if**
13: **end if**

The resulting probability distribution is then used to randomly choose a prodigy to replace the underdog (2: to 5:). Starting from a promising point in the search space, during the *Modification Phase*, a *Clone* will ultimately develop a different portfolio structure than its twin.

In contrast, an underdog is replaced by a so-called *Averaged Idol* (instead of by a *Clone*) with probability $1 - p_{clone}$ (6:). An *Averaged Idol* is a solution that is created based on the idol's pool of successful asset combinations and portfolio weights. The different indices of assets that are held by the group of idols Π^g are collected in a set $\mathfrak{I}_\Pi^g$. For each of these assets $i \in \mathfrak{I}_\Pi^g$ an averaged weight $\bar{w}_i^g$ is computed as follows:

$$\bar{w}_i^g = \frac{\displaystyle\sum_{n=\{x:x=1,\dots,\pi\,|\,i\in\mathfrak{I}_{<x>}^g\}} w_{i,n}^g a_{<n>}^g + b\epsilon w_{i,*}^g}{\displaystyle\sum_{i\in\mathfrak{I}_\Pi^g}\left(\sum_{n=\{x:x=1,\dots,\pi\,|\,i\in\mathfrak{I}_{<x>}^g\}} w_{i,n}^g a_{<n>}^g + b\epsilon w_{i,*}^g\right)}, \quad \forall\, i \in \mathfrak{I}_\Pi^g \qquad (11)$$

where

$$b = \begin{cases} 1 & \text{if } i \in \mathfrak{I}_*^g \\ 0 & \text{else} \end{cases}$$

In equation (11) an averaged weight of asset $i \in \mathfrak{I}_\Pi^g$ is computed, firstly, by building the sum (over all idols holding this asset) of the products of this assets' weights and the idols' amplifying factors. Secondly, this sum is build for all assets $i \in \mathfrak{I}_\Pi^g$ and summed up over all $\#\mathfrak{I}_\Pi^g$ assets to, thirdly, apply a normalization (7:). The number of different assets held by the idols, $\#\mathfrak{I}_\Pi^g$, varies over the generations and will usually be larger than K_{max}. Consequently, a decision of which assets to include in the *Averaged Idol* must be taken. This

23

selection is executed randomly with $\bar{w}_i^g$ as the selection probabilities (8:). Hence, those assets are more likely to be selected that appear in portfolios of idols more often and/or with larger portfolio weights.[26] The *Averaged Idol* is made a valid solution and gets evaluated (9:). It will replace the corresponding underdog, if its objective function value (the fitness) is greater than that of the underdog, or if the impairment is not greater than the threshold value T_t (10: to 12:). Hence, the acceptance rule from the *Modification Phase* is applied again.[27] The idea of *Averaged Idols* is to exploit not only the experience of two parents, like it is common in genetic algorithms, but those of a whole group of successful predecessors as is defined with Π^g.

After the *Replacement Phase* a new generation $g = (t-1) \cdot iter + l$ evolves, whose individuals will again, firstly, independently develop in the *Modification Phase*, before they, secondly, get ranked in the *Evaluation Phase*, before, thirdly, another *Replacement Phase* follows and so forth. After having reduced the threshold and the step size to their minimum values ($T_{thresh} = 0$, $U_{thresh} = U_{min}$) and after having computed $Pop + Pop \cdot thresh \cdot iter \cdot steps$ objective function values in $thresh \cdot iter$ generations, the algorithm terminates and reports the current elitist.

---

[26]In order to avoid multiple selections of assets into one Averaged Idol, an already picked asset is excluded from the list of (remaining) options. The "free probability" gets distributed over the remaining choices according to their weights.

[27]The application of a TA in the *Modification Phase* as well as the additional application of a TA-acceptance rule in the *Replacement Phase* are modifications of the algorithm of Maringer and Kellerer (2003).
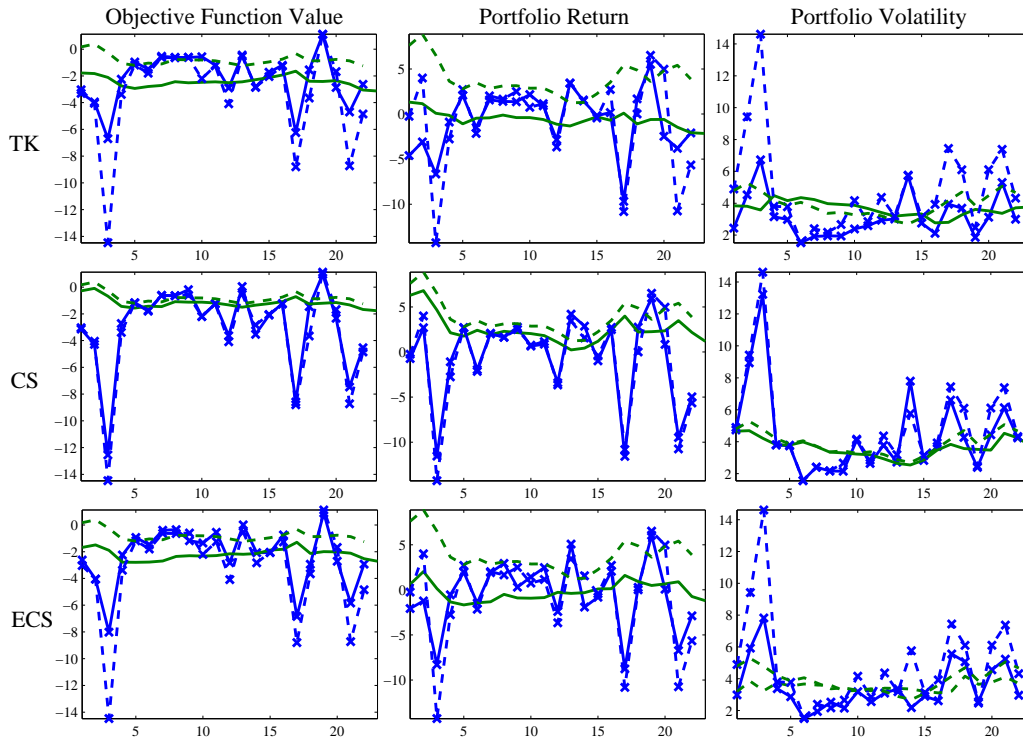
Appendix 2 - Graphical Representation

Figure 1: robust optimization approaches

All nine graphs in Figure 1 map the time window *win* on the x-coordinate, while the y-coordinate maps the objective function value (left column of graphs), the portfolio return (middle column of graphs), and the portfolio volatility deviation (right column of graphs). The series with markers show actually realized (out-of-sample) values, while the non-marked series show the corresponding expected values. In all graphs the dashed lines represent the MVO-approach.

References

Bertsimas, D. and D. Pachamanova (2008). Robust multiperiod portfolio management in the presence of transaction costs. *Computers & Operations Research* **35(1)**, 3–17.

Black, F.R. and R. Litterman (1991). Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income* **1(2)**, 7–18.

Cavadini, F., A. Sbuelz and F. Trojani (2001). A simplified way of incorporating model risk, estimation risk and robustness in mean-variance portfolio management. Working paper. Tilburg University.

Ceria, S. and R.A. Stubbs (2006). Incorporation estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management* **7(2)**, 109–127.

Chang, T.-J., N. Meade, J.E. Beasly and Y.M. Sharaiha (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research* **27**, 1271–1302.

Chopra, V. and W.T. Ziemba (1993). The effects of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* **19(2)**, 6–11.

DeMiguel, V. and F.J. Nogales (2009). Portfolio selection with robust estimation. *Operations Research* **57(3)**, 560–577.

Dueck, G. and P. Winker (1992). New concepts and algorithms for portfolio choice. *Applied Stochastic Models and Data Analysis* **8**, 159–178.

Dueck, G. and T. Scheuer (1990). Threshold accepting: A general purpose optimization algorithm. *Journal of Computational Physics* **90**, 161–175.

Efron, B. and R.J. Tibshirani (1993). *An Introduction to the Bootstrap.* Chapman & HAll.

Fabozzi, F.J., P.N. Kolm, D.A. Pachamanova and S.M. Focardi (2007). *Robust Portfolio Optimization and Management.* Wiley. New Jersey.

Fang, K.-T. and P. Winker (1997). Application of threshold accepting to the evaluation of the discrepancy of a set of points. *SIAM Journal on Numerical Analysis* **34(5)**, 211–223.

Fang, K.-T., S. Kotz and K.W. Ng (1990). *Symmetric Multivariate and Related Distributions.* London. Chapman & HAll.

Frost, P. and J.E. Savarino (1988). For better performance: Constrain portfolio weights. *Journal of Portfolio Management* **15(1)**, 29–34.

Genton, M.G. and E. Ronchetti (2008). Robust prediction of beta. In: *Computational Methods in Financial Engineering* (E. Kontoghiorghes, B. Rustem and P. Winker, Eds.). pp. 147–161. Springer. Berlin.

Gilli, M. and E. Këllezi (2002). A global optimization heuristic for portfolio choice with VaR and expected shortfall. In: *Computational Methods in Decision-making, Economics and Finance, Applied Optimization Series* (E. Kontoghiorghes, B. Rustem and S. Siokos, Eds.). pp. 165–181. Kluwer Academic Publishers.

Gilli, M. and I. Roko (2008). Using economic and financial information for stock selection. *Computational Management Science* pp. 317–335.

Gilli, M. and P. Winker (2009). Review of heuristic optimization methods in econometrics. *in: Handbook of Computational Econometrics, Ed. D. Beasley and E. Kontoghiorghes. Wiley. Chichester* pp. 81–120.

Goldfarb, D. and G. Iyengar (2003). Robust portfolio selection problems. *Mathematics of Operations Research* **28(1)**, 1–38.

Kirkpatrick, S., C.D. Gelatt and M.P. Vecchi (1983). Optimization by simulated annealing. *Science* **220**, 671–680.

Lauprete, G.J., A.M. Samarov and R.E. Welsch (2002). Robust portfolio optimization. *Metrika* **55**, 139–149.

Maringer, D. (2005). *Portfolio Management with Heuristic Optimization.* Vol. 8 of *Advances in Computational Management Science.* Springer. Dordrecht.

Maringer, D. and H. Kellerer (2003). Optimization of cardinality constrained portfolios with a hybrid local search algorithm. *OR Spectrum* **25**, 481–495.

Maringer, D. and P. Winker (2007). The threshold accepting optimization algorithm in economics and statistics. In: *Optimisation, Econometric and Financial Analysis* (E. Kontoghiorghes and C. Gatu, Eds.). pp. 107–125. Springer. Berlin.

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

Maronna, R., R. Martin and V. Yohai (2006). *Robust Statistics.* Wiley. New Jersey.

Michaud, R.O. (1989). The Markowitz optimization enigma: Is "optimized" optmial?. *Financial Analysts Journal* **45**, 31–45.

Michaud, R.O. (1998). *Efficient Asset Management.* Havard Business School Press. Boston, MA.

Perret-Gentil, C. and M.-P. Victoria-Feser (2004). Robust mean-variance portfolio selection. Technical report. FAME Research paper no 140.

Stubbs, R.A. and P. Vance (2005). Computing return estimation error matrices for robust optimization. *Axioma Research Papers* **1**, 1–9.

Tütüncü, R.H. and M. Koenig (2004). Robust asset allocation. *Annals of Operations Research* **132**, 157–187.

Welsch, R.E. and X. Zhou (2007). Application of robust statistics to asset allocation models. *REVSTAT - Statistical Journal* **5(1)**, 97–114.

Winker, P., M. Lyra and C. Sharpe (2009). Least median of squares estimation by optimization heuristics with an application to the CAPM and multi factor models. *Forthcomming in: Computational Management Science.*

Zymler, S., B. Rustem and D. Kuhn (2009). Robust portfolio optimization with derivative insurance guarantees. Working paper. Department of Computing, Imperial College London.