Härdle, Wolfgang; Klochkov, Yegor; Petukhina, Alla; Zhivotovskiy, Nikita

Working Paper

Robustifying Markowitz

IRTG 1792 Discussion Paper, No. 2021-018

Provided in Cooperation with:
Humboldt University Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series"

Suggested Citation: Härdle, Wolfgang; Klochkov, Yegor; Petukhina, Alla; Zhivotovskiy, Nikita (2021) : Robustifying Markowitz, IRTG 1792 Discussion Paper, No. 2021-018, Humboldt-Universität zu Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series", Berlin

This Version is available at:
https://hdl.handle.net/10419/243167

IRTG 1792 Discussion Paper 2021-018

Robustifying Markowitz

Wolfgang Karl Härdle [*] [*2] [*3] [*4] [*5]
Yegor Klochkov [*6]
Alla Petukhina [*7]
Nikita Zhivotovskiy [*8]

[*] Humboldt-Universität zu Berlin, Germany
[*2] Xiamen University, China
[*3] Singapore Management University, Singapore
[*4] Charles University, Czech Republic
[*5] National Chiao Tung University, Taiwan
[*6] University of Cambridge, United Kingdom
[*7] Hochschule fur Technik und Wirtschaft Berlin, Germany
[*8] ETH, Zürich

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

Robustifying Markowitz

Wolfgang Karl Härdle[*], Yegor Klochkov[†], Alla Petukhina[‡], Nikita Zhivotovskiy[§]

September 30, 2021

Abstract

Markowitz mean-variance portfolios with sample mean and covariance as input parameters feature numerous issues in practice. They perform poorly out of sample due to estimation error, they experience extreme weights together with high sensitivity to change in input parameters. The heavy-tail characteristics of financial time series are in fact the cause for these erratic fluctuations of weights that consequently create substantial transaction costs. In robustifying the weights we present a toolbox for stabilizing costs and weights for global minimum Markowitz portfolios. Utilizing a projected gradient descent (PGD) technique, we avoid the estimation and inversion of the covariance operator as a whole and concentrate on robust estimation of the gradient descent increment. Using modern tools of robust statistics we construct a computationally efficient estimator with almost Gaussian properties based on median-of-means uniformly over weights. This robustified Markowitz approach is confirmed by empirical studies on equity markets. We demonstrate that robustified portfolios reach higher risk-adjusted performance and the lowest turnover compared to shrinkage based and constrained portfolios.

[*]Blockchain Research Center, Humboldt-Universität zu Berlin, Germany. Wang Yanan Institute for Studies in Economics, Xiamen University, China. Sim Kee Boon Institute for Financial Economics, Singapore Management University, Singapore. Faculty of Mathematics and Physics, Charles University, Czech Republic. Yushan Scholar National Yangming Jiaotong University, Taiwan. haerdle@hu-berlin.de

[†]Cambridge-INET, Faculty of Economics, University of Cambridge yk376@cam.ac.uk

[‡]Hochschule für Technik und Wirtschaft Berlin, School of Computer Science, Communication and Economics alla.petukhina@htw-berlin.de

[§]Department of Mathematics ETH, Zürich nikita.zhivotovskii@math.ethz.ch

1 Introduction

The cornerstone mean-variance portfolio theory proposed by Markowitz (1952) plays a significant role in research and practice. Efficient mean-variance portfolios (MV) experience a number of attractive properties and have a simple and straightforward analytical solution with only two input parameters: the expected mean and covariance matrix of asset returns. Mean-variance analysis is naturally connected to the Capital Asset Pricing Model (CAPM), a standard tool in asset pricing.

Despite its simplicity and theoretical appeal, implementation of mean-variance portfolios is often impractical. The traditional approach to use the sample moments as input parameters leads to extreme negative and positive weights, and extensive literature documents poor out-of-sample performance of such plug-in approach, see (Frost and Savarino; 1986, 1988; Best and Grauer; 1991; Chopra and Ziemba; 1993; Broadie; 1993; Litterman; 2004; Merton; 1980). The problem might be seen as an inverse problem, and it simply has high sensitivity to even small perturbations of the input estimates: the mean and the covariance matrix. It is possibly surprising that the MV portfolios are more sensitive to changes in the mean estimate, Jagannathan and Ma (2003) spell this out explicitly by writing that the error of mean estimation is so large that nothing is lost when one ignores the mean at all, and Michaud (1989) describes the influence of the mean error as "error-maximization." Following majority of research on this topic, we focus here on *global minimum variance* portfolios (GMV), which only depend on the covariance.

However, even with the mean left out of the equation, traditional policies suffer from extreme instability, which means that the portfolio weights fluctuate significantly over time. Drastic changes in the portfolio composition lead to increasing management and transaction costs and consequently reducing the popularity of MV policies among investors. In order to improve upon the stability of portfolio weights, one has to resort to alternative, robust estimation techniques. A *robust estimator* is one that performs well even when the observations do not follow the standard (normality) assumptions, have heavy tails, or are even subject to contamination. Although in case of normal distributions, sample moment estimators are asymptotically optimal MLEs, they are

not necessarily the best choice when the data deviates from normality (Huber; 2004). This is of particular importance in financial applications, where it is well known that the data is not only non-Gaussian, but also exhibits heavy tails.

To tackle the problem of heavy tails, DeMiguel and Nogales (2009) construct a portfolio optimization procedure based on M- and S-estimation technique and analyze the stability of the estimator analytically; they also demonstrate empirically that their approach reduces portfolio *turnover*, whereas it slightly improves the out-of-sample performance. Fan et al. (2019) construct an elementwise covariance estimator through an M-estimation procedure with Huber loss, providing statistical high-probability guarantees. Robust portfolio optimization problem has gained significant attention, Xidonas et al. (2020) categorize 148 researches conducted during the last 25 years and focused on this topic.

Failures of MV portfolios become even more pronounced with a growing investment universe, especially for cases when a sample size is less than the number of assets. Evidence was investigated by Kan and Zhou (2007), Bai et al. (2009), Karoui (2012), and Chen and Yuan (2016). To overcome this curse of dimensionality, structured covariance matrix estimators are proposed for asset return data. Fan et al. (2008) considered estimators based on factor models with observable factors. Stock and Watson (2002), Bai and Li (2012), Fan et al. (2013) studied covariance matrix estimators based on latent factor models. Ledoit and Wolf (2003), Ledoit and Wolf (2004b), Ledoit and Wolf (2004a) proposed linear and Ledoit and Wolf (2017) non-linear shrinkage of sample eigenvalues. These estimators are commonly based on the sample covariance matrix, and sub-Gaussian tail assumptions are required to guarantee consistency.

The goal of our robustifying Markowitz approach is to tackle both problems at the same time: how to optimize GMV portfolio when the dimension is possibly higher than the sample size and the distribution of the returns has heavy tails? Moreover, even if the returns are not heavy-tailed, how can one avoid the usual Gaussian assumption in the theoretical analysis?

Our theoretical and algorithmic contributions dwell on some recent breakthroughs

in statistical literature with regard to robust estimation. Lugosi and Mendelson (2019b) constructed a multivariate mean estimator based on the idea of median-of-means that dates back to Nemirovsky and Yudin (1983). The remarkable property of their estimator is that it pertains favorable properties of the Gaussian sample mean: it allows deviation bounds with high probability without much loss in the accuracy. Their only condition is that the second moment of each component of the random vector is bounded, which is the minimal possible condition to have a square-root convergence, even on average. However, their original estimator was not computationally tractable and in the past years the problem has attracted a lot of attention. Hopkins (2018) first proposed an estimator with polynomial computational (efficient) complexity, and subsequent research led to nearly linear-time algorithms (Depersin and Lecué; 2019; Hopkins et al.; 2020), thus making practical applications possible. As for the covariance estimation, Mendelson and Zhivotovskiy (2020) proposed an abstract algorithm that achieves performance of Gaussian sample covariance estimator under four bounded moments assumption. So far, it remains an open question whether such performance can be achieved with a polynomial algorithm, with some conjecturing that the answer is no (Cherapanamjeri et al.; 2020).

We bypass these algorithmic problems appearing in robust estimation of the covariance matrix. In fact, our approach does not require estimating the covariance operator directly. It is based on a simple iterative gradient descent, that requires estimating only the action of the covariance operator on a current approximation at each step.

Our contribution to robustifying Markowitz is threefold:

- Based on the algorithm from (Hopkins et al.; 2020), we introduce a robust and computationally tractable algorithm that achieves nearly Gaussian performance under only four moments assumption on the distribution of the return vector. This means, in particular, that the estimator works with almost any distribution with four bounded moments as good as it works with Gaussian data.

- We provide theoretical guarantees for our method in two cases. In the first case, we assume that the covariance matrix is well-conditioned, which means that the objective of our optimization problem enjoys the strong convexity property, and

the convergence guarantees are provided even for the mean-variance objective. However, in that case we require that the the dimension of the investment universe (N) is much smaller than the size of the sample (T). Moreover, the assumption that the covariance is well-conditioned is impractical due to the presence of strong factors in financial panel data, and we provide this result merely out of theoretical curiosity.

In the general case where we have no control over the small eigenvalues of the covariance matrix, we only consider the GMV objective. However, we can take advantage of possibly small *effective rank* of the covariance matrix, which allows the dimension N to be a lot larger than the sample size T.

- In an empirical study we compare behavior of the proposed portfolio estimator to the traditional portfolio benchmarks on equity data for two cases: when the size of the sample T is comparable with the dimension N, and when $T < N$. For the first case, we consider the S&P100 data, and for the second case we take the constituents of the Russell3000 index. In both cases we take daily data over the course of one year, which corresponds to $T = 252$. We demonstrate that our approach enjoys more stable weights than the traditional portfolios, while preserving (or slightly improving) their out-of-sample performance.

Let us also recall a well known hypothesis of Green and Hollifield (1992) that extreme portfolio weights appear not entirely due to high estimation errors, but rather due to the population optimal portfolios themselves having extreme weights and being poorly diversified. Specifically, they show that asset returns generated by a model with a single dominant factor result in excessive short and long positions. This leads to the study of restricted portfolio policies. In a seminal work, Jagannathan and Ma (2003) consider portfolios with non-negative constraints. Despite considering a lesser class of portfolios, they demonstrate a better out-of-sample performance. Furthermore, Fan et al. (2012) introduce *Gross Exposure Constraints*, which work similarly but allow negative allocation weights. Contrary to these ideas, we demonstrate that applying our robust procedure leads to desirable properties of weights without any constraints enforced a priori, which contradicts the original hypothesis of Green and Hollifield.

The remaining of the paper is organized as follows. In the beginning of Section 2, we describe the setup and give definitions of the MV and GMV portfolios, and give informal statements of our main results. In Section 2.1, we give introduction into recent advances in robust statistic. In Section 2.2, we describe the projected gradient descent technique applied to our setup and provide some motivation for the estimation of actions of the covariance operator. Then we describe how to estimate them. In Section 3 we first present the results for the well-conditioned case, that is, when the eigenvalues of the covariance matrix have the same order, and in Section 4 we consider a general case that allows poorly invertible covariance matrix as well as high dimensionality. Section 5 defines the performance measures for comparison of different benchmarks. Empirical results are presented in Section 6 where S&P100 and Russell3000 portfolios are compared to multiple standard benchmarks. Finally, the last section is devoted to conclusions and final remarks.

1.1 Notation

Throughout the paper we write $a \lesssim b$ and $b \gtrsim a$ if there is a constant C such that $a \leq Cb$. If we have both $a \lesssim b$ and $b \lesssim a$, we write $a \sim b$.

For a vector $x \in \mathbb{R}^d$, we denote by $\|x\| = \sqrt{x_1^2 + \cdots + x_d^2}$ its Euclidean norm. If A is a matrix, we denote $\|A\| = \sup_{u,v \in \mathbb{S}^{d-1}} u^\top A v$ its spectral norm, where $\mathbb{S}^{d-1}$ is a sphere in $\mathbb{R}^d$. If $A \in \mathbb{R}^{d \times d}$ is symmetric, we write $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_d(A)$ to denote its eigenvalues in descending order. We also denote $\lambda_{\max}(A) = \lambda_1(A)$ and $\lambda_{\min}(A) = \lambda_d(A)$ — its largest and smallest eigenvalues, respectively. In particular, we have that $\|A\| = \max\{|\lambda_{\max}(A)|, |\lambda_{\min}(A)|\}$. We say that a symmetric matrix A is positive semi-definite (PSD) if $v^\top A v \geq 0$ for all v. We also write $A \preceq B$ and $B \succeq A$ if $B - A$ is a PSD.

For a PSD matrix $A \in \mathbb{R}^{d \times d}$ we denote its *effective rank* by

$$\mathbf{r}(A) = \mathrm{Tr}(A)/\lambda_{\max}(A) = \sum_{j=1}^{d} \lambda_j(A)/\lambda_{\max}(A).$$

The effective rank is clearly always smaller than both the dimension d and the matrix

rank of A. This quantity plays an important role in covariance estimation problems. In particular, it was shown by Koltchinskii and Lounici (2017) that in the Gaussian case, the performance of the sample covariance matrix is governed by the effective rank of the covariance matrix and is not sensitive to a potentially larger dimension of the ambient space.

Finally, we denote $\mathbf{1} = (1, \dots, 1)^\top$ of dimension N, so that $w^\top \mathbf{1} = \sum_{i=1}^N w_i$.

2 Mean-Variance and Global Minimum Variance portfolios

Suppose we have an opportunity to invest into N assets and $r_1, \dots, r_N$ denote their log-returns. Let $X = (r_1, \dots, r_N)^\top$ be the multivariate return vector with mean μ and covariance Σ. Then a portfolio with allocation weights $w = (w_1, \dots, w_N)^\top$ has returns with expectation $\mu^\top w$ and variance $w^\top \Sigma w$.

One of the fundamental portfolio policies, the *mean-variance* portfolio (MV), is based on maximizing the utility

$$M_\gamma(w; \mu, \Sigma) = \mu^\top w - \frac{\gamma}{2} w^\top \Sigma w \qquad \text{subject to } \; w^\top \mathbf{1} = 1\,,$$

which takes as input the mean μ and the covariance operator Σ. Moreover, γ is a fixed *risk aversion* parameter provided by the investor. The quadratic term in the above expression represents the variance of the portfolio return $\mathrm{Var}(w^\top X)$, and the linear term is its mean $\mathsf{E}(w^\top X)$.

Some researchers often discard the dependence on mean and concentrate on an alternative portfolio policy that minimizes the risk measure

$$R(w; \Sigma) = \frac{1}{2} w^\top \Sigma w \qquad \text{subject to } w^\top \mathbf{1} = 1, \tag{1}$$

which corresponds to finding a *global minimum variance* portfolio (GMV). The quantity $w^\top \Sigma w = \mathrm{Var}(w^\top X)$ is often regarded as *realized risk* of a portfolio allocation w in the financial literature.

Suppose we have an i.i.d. sample $X_1, \dots, X_T$ that comes from the distribution with mean μ and covariance Σ. Our goal is to construct portfolio allocation weights $\hat{w}$ that are

as close to optimum as possible. Below we provide theoretical high-probability guarantees in terms of the gap between the estimator and the population optimal solution, that is,

$$R(\hat{w}; \Sigma) - \min_{w^\top \mathbf{1}} R(w; \Sigma)$$

in the GMV case, or

$$\max_{w^\top \mathbf{1}} M_\gamma(w; \mu, \Sigma) - M_\gamma(\hat{w}; \mu, \Sigma)$$

for the mean-variance portfolio. Notice that both of these entities are non-negative.

We analyze two different situations, focusing on high-dimensional non-asymptotic bounds. Firstly, we consider the hypothetical situation where the covariance matrix is well-conditioned. Such situation is unlikely in practice, and we present the following result mainly out of theoretical curiosity.

Theorem (Informal)**.** *Suppose that Σ is well-conditioned, i.e. its condition number is bounded by a constant. Then, for each δ, there is a computationally efficient estimator $\hat{w}_\delta$ that satisfies, with probability at least $1 - \delta$,*

$$\max_{w^\top \mathbf{1}=1} M_\gamma(w; \Sigma, \mu) - M_\gamma(\hat{w}_\delta; \Sigma, \mu) \lesssim \frac{N \log N + \log(1/\delta)}{T}$$

even when the distribution is non-Gaussian and has heavy tails.

The above result demonstrates that the MV portfolio can be robustly estimated as long as the ratio $N \log N/T$ remains small. It is known that for convergence to the optimal risk one has to have that $N/T = o(1)$. For instance, Karoui (2013) considers a "large N, large T" situation where N/T converges to some constant $\gamma \in (0, 1)$, and shows that there is a constant gap between the realized risks of the empirical and the optimal solutions. More recently, Bartl and Mendelson (2021) studied a similar portfolio optimization setup in a well-conditioned case. Although their algorithm is robust with respect to heavy-tailed data and achieves similar rates of convergence, their estimator is not computationally feasible.

Secondly, we consider the case where Σ is allowed to be ill-conditioned. This case corresponds to a less regular optimization problem and we provide slower convergence guarantees with respect to the number of observations.

Theorem (Informal)**.** *There is a computationally efficient estimator $\hat{w}_\delta$, such that, with probability at least $1-\delta$,*

$$R(\hat{w}_\delta;\Sigma) - \min_{w^\top \mathbf{1}=1} R(w;\Sigma) \lesssim \sqrt{\frac{\mathbf{r}(\Sigma)\log \mathbf{r}(\Sigma) + \log(1/\delta)}{T}}\,,$$

even when the distribution is non-Gaussian and has heavy tails.

This result suggests that the GMV portfolio converges to optimum as long as $\mathbf{r}(\Sigma)$ is much smaller than T, which is a rather adequate assumption. For example, for the S&P100 dataset, we evaluate that $\mathbf{r}(\Sigma) \approx 3$ and for the Russell3000 constituents, $\mathbf{r}(\Sigma) \approx 7$. Notice also that the condition number $\kappa \stackrel{\text{def}}{=} \lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$ is bounded from below by $N/\mathbf{r}(\Sigma)$, hence the covariance matrix is indeed ill-conditioned in these two applications.

2.1 Recent advances in robust statistics

The covariance matrix and the mean are not known in practice and must be estimated based on the observed log-returns.

In an idealized situation where $X_i \sim \mathcal{N}(\mu,\Sigma)$ are Gaussian, we have that the standard empirical mean estimator $\hat{\mu} = T^{-1}\sum_{i=1}^T X_i$ provides one with optimal high probability deviation bounds. In particular, for all $\delta \in (0,1)$, we have that, with probability at least $1-\delta$,

$$\|\hat{\mu} - \mu\| \le \sqrt{\frac{\mathrm{Tr}(\Sigma)}{T}} + \sqrt{\frac{2\|\Sigma\|\log(1/\delta)}{T}} \tag{2}$$

See (Boucheron et al.; 2013, Example 5.7) for derivation. The sharp deviation term $\sqrt{\frac{2\|\Sigma\|\log(1/\delta)}{T}}$ is very specific to the Gaussian assumption and could not be expected for less regular distributions. In particular, here the dependence on the confidence level is logarithmic and additive, in the sense that the bound separates into the *strong term* scaled with $\sqrt{\mathrm{Tr}(\Sigma)}$ and corresponding to the error on average, and the *weak term* that is scaled with $\sqrt{\|\Sigma\|}$. The weak term can potentially be a lot smaller than the strong one even for very small values of δ.

Similarly, Koltchinskii and Lounici (2017) proved that in the case of i.i.d. zero mean Gaussian observations, the sample covariance $\hat{\Sigma} = T^{-1}\sum_{i=1}^T X_i X_i^\top$ satisfies the

following deviation bound. For any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\| \sqrt{\frac{\mathbf{r}(\Sigma) + \log(1/\delta)}{T}}, \tag{3}$$

whenever $n \gtrsim \mathbf{r}(\Sigma) + \log(1/\delta)$. For the version of this inequality with explicit constants we refer to (Zhivotovskiy; 2021).

Having these performance bounds in mind, one is interested if the same bounds can be achieved under milder assumptions in a computationally efficient manner. Lugosi and Mendelson (2019b) developed an estimator that matches the bound (2) under the only assumption that the covariance exists (i.e., two moments assumption). Loosely speaking, they propose to control the deviations of the median-of-means of the projections $X_i^\top v$ uniformly in all directions $v \in \mathbb{R}^N$. Based on this bound, they came up with a tournament-based procedure, which however is too complicated to perform computationally. Further developments were made in (Lugosi and Mendelson; 2021), see also (Lugosi and Mendelson; 2019a) for a thourough review on this topic.

Hopkins (2018) first proposed an estimator that can be computed in polynomial time, and the time complexity was subsequently improved to nearly linear (Cherapanamjeri et al.; 2019; Depersin and Lecué; 2019; Hopkins et al.; 2020). An alternative method called *spectral sample reweighing* was developed in the context of robust estimation with outliers. Given data points $\{x_i\}_{i=1,\dots,k}$ the goal is to reweigh the points x_i with some weights $u_i \in [0,1]$ and find a center $\nu \in \mathbb{R}^N$ such that the largest eigenvalue of the weighted covariance $\sum_i u_i (x_i - \nu)(x_i - \nu)^\top$ is small. Hopkins, Li and Zhang (2020) develop an algorithm that does this in nearly linear time; see also (Diakonikolas et al.; 2017; Zhu et al.; 2020). More importantly for us, Hopkins, Li and Zhang establish a direct connection between the sample reweighing and the method developed in (Lugosi and Mendelson; 2019b), which makes this approach applicable in the heavy-tailed setup as well. We discuss this connection in greater detail in Section 8.1.

The problem of robust covariance estimation is more challenging. Mendelson and Zhivotovskiy (2020) construct an abstract estimator that matches the bound (3) up to some logarithmic factors. Unfortunately, it is not known whether such performance can be achieved with a computationally efficient algorithm. Existing computationally effi-

cient implementations are showing sub-optimal statistical guarantees (Ke et al.; 2019; Ostrovskii and Rudi; 2019; Hopkins et al.; 2020) and sometimes require additional assumptions such as the so-called SoS hypercotractivity that are hard to verify in the non-Gaussian situation (Hopkins et al.; 2020). Moreover, Hopkins et al. (2020) conjecture that as long the median-of-means approach is used, it is algorithmically hard to robustly estimate the sample covariance matrix in the presence of heavy-tailed data.

After this short excursion to some recent results in robust estimation, let us now come back to our portfolio optimization problem. Since we cannot get our hands on a robust covariance estimator, we take another route by observing that both MV and GMV are convex optimization problems.

2.2 Gradient descent for portfolio optimization

Our goal is to avoid the estimation of the whole covariance matrix, but rather resort to the estimation of the action of this operator Σw on some limited set of vectors w. We will use a procedure based on *projected gradient descent* (PGD), which is a standard convex optimization method. For instance, if we want to minimize the GMV objective with known Σ, the following sequence of approximations converges to an optimal solution (which is not necessarily unique): we start with arbitrary initial vector w_0 and then take the update steps,

$$w_s = \Pi_1[w_{s-1} - \eta \nabla_w R(w_{s-1}; \Sigma)], \qquad s = 1, 2, \dots, \tag{4}$$

where Π_1 is the orthogonal projector onto the restricted (convex) set $\{w : \ w^\top \mathbf{1} = 1\}$, which can be explicitly defined by the mapping,

$$\Pi_1 x = (I - N^{-1}\mathbf{1}\mathbf{1}^\top)x + N^{-1}\mathbf{1}\,.$$

It is straightforward to see that $R(\,\cdot\,;\Sigma)$ is convex (since the covariance operator is positive semi-definite), and $\|\Sigma\|$-smooth. By Theorem 3.7 from (Bubeck; 2014) the sequence (4) converges to a minimum at a rate $1/s$ as long as $\eta \leq 1/\|\Sigma\|$. Moreover, in the case where Σ is non-degenerate, the objective becomes strongly convex, and the sequence converges at a faster exponential rate under the same requirement on the step size, see (Bubeck; 2014, Theorem 3.10).

The case of MV portfolio is similar, only this time we need to maximize a concave function instead of minimizing a convex one. If we replace $\nabla_w R(w_{s-1};\Sigma)$ with $-\nabla_w M_\gamma(w_{s-1};\Sigma)$ in (4), then by the same reasons, the sequence converges to the maximum of M_γ as long as $\eta \le (\gamma\|\Sigma\|)^{-1}$, with exponential rate when Σ is non-degenerate.

The PGD iterations require computation of the following gradients,

$$\nabla_w R(w;\Sigma) = \Sigma w \qquad \text{or} \qquad \nabla_w M_\gamma(w;\Sigma) = \mu - \gamma\Sigma w.$$

where the mean μ and covariance Σ are typically replaced with their empirical counterparts that are calculated using given historical observations $X_1,\dots,X_T$. As discussed in the previous section, there is a practical robust mean estimator in (Hopkins et al.; 2020) with all desired properties. Since such an estimator is not available for the covariance operator, we instead produce an estimator $\hat{a}_\delta(w)$ for the PGD increment that estimates Σw for each w separately, and plug it into the update steps (4).

To see how it can be done, suppose for a moment that the expectation vanishes. Then, we can represent this product as a mean of a random vector as follows,

$$\Sigma w = \mathsf{E}(X^\top w)X\,.$$

We therefore can apply the robust mean algorithm to the vectors $(X_i^\top w)X_i$ and obtain a robust estimator of Σw. However, we need to take additional care to ensure that the estimator is an appropriate approximation uniformly in all directions. For this, we slightly adjust the procedure in the spirit of (Mendelson and Zhivotovskiy; 2020). In the latter work, the only assumption used is the equivalence of the fourth and the second moments in all directions sometimes called the *bounded kurtosis* assumption.

Assumption 2.1 (Bounded kurtosis)**.** *The return vectors $X_1,\dots,X_T$ are i.i.d. observations of a random vector X, that has mean μ, covariance Σ, and satisfies for all $u\in\mathbb{R}^N$,*

$$\mathsf{E}^{1/4}|u^\top(X-\mu)|^4 \le K\mathsf{E}^{1/2}|u^\top(X-\mu)|^2, \tag{5}$$

where $K\ge 1$ is some fixed constant.

All our results will be stated under this assumption. We remark that under the Gaussian distribution one has control over all higher moments, see e.g. (Koltchinskii

and Lounici; 2017). In the context of covariance estimation, a robust estimator is one that has Gaussian deviation bounds (i.e., as in (3)) but only requires the underlying distribution to follow the bounded kurtosis assumption. The next step is to provide a robust estimator of Σw that works simultaneously in all directions.

Proposition 2.1. *Suppose the bounded kurtosis assumption holds. There is a computationally efficient estimator $\hat{a}_\delta(w)$ that depends on direction w and T i.i.d. observations and such that, with probability at least $1-\delta$,*

$$\|\hat{a}_\delta(w) - \Sigma w\| \lesssim \|\Sigma\| \sqrt{\frac{\mathbf{r}(\Sigma)\log \mathbf{r}(\Sigma) + \log(1/\delta)}{T}} \|w\|$$

uniformly for all vectors w. Recall that $\mathbf{r}(\Sigma)$ denotes the effective rank

$$\mathbf{r}(\Sigma) = \mathrm{Tr}(\Sigma)/\|\Sigma\| \,.$$

We postpone the proof and detailed description of the estimation algorithm until Section 8.1.

Remark 2.1. *For technical reasons, the estimator $\hat{a}_\delta(w)$ depends on a norm-truncation parameter R that needs to be of order $\left(\frac{T\mathbf{r}(\Sigma)}{\log \mathbf{r}(\Sigma)}\right)^{1/4}$, which is unknown in general. It appears that since R increases with T, in many natural situations this truncation parameter is of a much larger order than $\max_i \|X\|$ and can be mostly ignored in practice. For more details see Section 8.1.*

Remark 2.2. *We point out that when one has access to some covariance estimator $\hat{\Sigma}$, one can simply take a family of estimators $\hat{a}(w) = \hat{\Sigma} w$. For instance, in the Gaussian case, taking the standard empirical covariance estimator would yield thanks to (3),*

$$\|\hat{a}(w) - \Sigma w\| \lesssim \|\Sigma\| \sqrt{\frac{\mathbf{r}(\Sigma) + \log(1/\delta)}{T}} \|w\|$$

with probability at least $1-\delta$, uniformly for all w. The estimator of Proposition 2.1 achieves the same rate of convergence under minimal distributional assumptions.

Now we can plug in the estimator of Σw (appearing in Proposition 2.1) into the update rule (4). To be precise, in the case of GMV optimization, our updates look as follows,

$$w_s = \Pi_1 \left[w_{s-1} - \eta \hat{a}_\delta(w_{s-1}) \right], \qquad s = 1, 2, \ldots$$

Naturally, the error may accumulate with each update, and we need to carefully analyze how the resulting solution differs from the optimum, to which the sequence (4) converges. We analyze this update rule in two separate cases.

First, we consider the case of well-conditioned matrix Σ, meaning that the ratio of its maximal and minimal eigenvalues is bounded by a constant. This means that the problem of maximizing the MV utility is a strongly-convex optimization problem, so that the gradient descent sequence enjoys exponential convergence rate and as we show below, the error of estimation does not accumulate. However, in that situation the effective rank $\mathbf{r}(\Sigma) = \mathrm{Tr}(\Sigma)/\|\Sigma\|$ is of order N, so the convergence only works in the case where N/T is small. Moreover, in typical applications, the covariance matrix is ill-conditioned, which is one of the reasons the MV portfolio performs so poorly. This comes in not enjoy this properties the covariance matrix is poorly invertible. For instance, this can be checked through evaluation of the effective rank: for the S&P100 dataset we estimate $\mathbf{r}(\Sigma) \approx 3$ and for the Russell3000 set we estimate that $\mathbf{r}(\Sigma) \approx 7$, in both cases much smaller than the dimension N. This brings us to the second part of our GD analysis, where we only consider the case of GMV optimization with ill-conditioned covariance matrix that has small effective rank. This scenario corresponds to non-strongly convex optimization and has weaker convergence rate. However, it enjoys dimension-free bounds, meaning that the convergence is guaranteed as long as the number observations is much larger than $\mathbf{r}(\Sigma)$, regardless of how high the total number of assets is. We also point out that in this case, one has to stop after appropriate number of steps to avoid overfitting.

3 Well-conditioned case

For maximizing the MV utility $M_\gamma(w; \mu, \Sigma)$, we consider the following updates,

$$w_s = \Pi_1 \left[w_{s-1} + \eta(\hat{\mu} - \gamma \hat{a}(w_{s-1})) \right], \qquad s = 1, 2, \ldots \tag{6}$$

where $\hat{\mu}$ is some estimator of mean μ, and $\hat{a}(w)$ is some family of estimators for the action of covariance operator Σw. We first show a deterministic result that controls the convergence through the errors of estimators $\hat{\mu}$ and $\hat{a}(w)$.

Lemma 3.1. *Denote,* $w^* = \arg\max_{w^\top \mathbf{1}=1} M_\gamma(w;\mu,\Sigma)$. *Suppose that we have an access to an estimator* $\hat{\mu}$ *satisfying*

$$\|\hat{\mu} - \mu\| \le \Delta_\mu,$$

and an access to a family of estimators $\hat{a}(w)$ *satisfying uniformly for all* $w \in \mathbb{R}^N$,

$$\|\hat{a}(w) - \Sigma w\| \le \Delta_\Sigma \|w\|\,.$$

Let $\lambda_{\max}$, $\lambda_{\min}$ *denote the maximal and minimal eigenvalues of* Σ, *respectively. Assume that* $\eta \le 1/(\gamma\lambda_{\max})$ *and* $\Delta_\Sigma < \lambda_{\min}$. *Then, the sequence* (6) *satisfies*

$$\|w_s - w^*\| < \left(1 - \frac{\gamma\eta(\lambda_{\min} - \Delta_\Sigma)}{2}\right)^s \|w_0 - w^*\| + \frac{\Delta_\mu + \gamma\Delta_\Sigma\|w^*\|}{\lambda_{\min} - \Delta_\Sigma}\,.$$

Proof. Let us first calculate w^* explicitly. Since $-M_\lambda(w;\Sigma,\mu)$ is strongly convex, and adding a Lagrangian multiplier $-l(w^\top\mathbf{1} - 1)$ corresponding to the restriction $w^\top\mathbf{1} = 1$, we have that w^* is the solution to

$$-\mu + \gamma\Sigma w - l\mathbf{1} = 0 \qquad \Rightarrow \qquad w = \frac{1}{\gamma}\Sigma^{-1}(\mu + l\mathbf{1})\,.$$

Since $w^\top\mathbf{1} = 1$ we find that $l = (\gamma - \mathbf{1}^\top\Sigma^{-1}\mu)/(\mathbf{1}^\top\Sigma^{-1}\mathbf{1})$. Therefore,

$$w^* = \gamma^{-1}\Sigma^{-1}\mu + \frac{1 - \gamma^{-1}\mathbf{1}^\top\Sigma^{-1}\mu}{\mathbf{1}^\top\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}\,.$$

Denote $\Pi_0 = I - N^{-1}\mathbf{1}\mathbf{1}^\top$ the orthogonal projector onto the subspace of $\{w:\ w^\top\mathbf{1} = 0\}$, so that $\Pi_1(x+y) = \Pi_1 x + \Pi_0 y$. It is straightforward to check that $\Pi_0(\gamma\Sigma w^* - \mu)$ vanishes, which is all we need to know about w^* for the remaining of the proof.

Write $\Delta(w) = \gamma\{\Sigma w - \hat{a}(w)\} - (\mu - \hat{\mu})$. Then,

$$\begin{aligned} w_{s+1} - w^* &= w_s - w^* - \eta\Pi_0[\gamma\Sigma w_s - \mu] + \eta\Pi_0\Delta(w_s) \\ &= (1 - \eta\gamma\Pi_0\Sigma)\,(w_s - w^*) + \eta\Pi_0\Delta(w_s) \\ &= \Pi_0(1 - \eta\gamma\Sigma)\Pi_0(w_s - w^*) + \eta\Pi_0\Delta(w_s)\,, \end{aligned}$$

where for the last equality we used the fact that $w_s - w^* = \Pi_0(w_s - w^*)$ since both sum up to one. Since $\eta \le 1/(\gamma\lambda_{\max})$, we have that $1 - \eta\gamma\Sigma$ is positive definite and $\|1 - \eta\gamma\Sigma\| = 1 - \eta\gamma\lambda_{\min}$. In addition, due to the requirement of the theorem for the estimators $\hat{\mu}$ and $\hat{a}(w)$, we have that

$$\|\Delta(w)\| \le \Delta_\mu + \gamma\Delta_\Sigma\|w\|\,.$$

Denoting $\delta_s = \|w_s - w^*\|$, we have the recursive inequality,

$$\delta_{s+1} \leq (1 - \eta\gamma\lambda_{\min})\delta_s + \eta\Delta_\mu + \eta\gamma\Delta_\Sigma\|w_s\|\,.$$

We can link $\|w_s\|$ to δ_s through a simple triangle inequality $\|w_s\| \leq \|w^*\| + \delta_s$. We obtain,

$$\delta_{s+1} \leq \{1 - \eta\gamma(\lambda_{\min} - \Delta_\Sigma)\}\delta_s + \eta\Delta', \qquad \Delta' \stackrel{\text{def}}{=} \Delta_\mu + \gamma\Delta_\Sigma\|w^*\|\,.$$

Denoting $\kappa = 1 - \eta\gamma(\lambda_{\min} - \Delta_\Sigma) < 1$ and $x = \eta\Delta'$, we expand our recursive inequality as follows,

$$\begin{aligned}\delta_{s+1} &\leq \kappa\delta_s + x \leq \kappa^2\delta_{s-1} + \kappa x + x \leq \kappa^{s+1}\delta_0 + (\kappa^s + \cdots + 1)x \\ &< \kappa^{s+1}\delta_0 + \frac{x}{1-\kappa}\,.\end{aligned}$$

Substituting κ and x back, we obtain the result. □

We now apply this lemma to the case where we use $\hat{\mu}_\delta$ from (Hopkins et al.; 2020) and $\hat{a}_\delta(w)$ from Proposition 2.1. From (2) we get that, with probability at least $1 - \delta$,

$$\|\hat{\mu}_\delta - \mu\| \lesssim \|\Sigma\|^{1/2}\sqrt{\frac{\mathbf{r}(\Sigma) + \log(1/\delta)}{T}} \tag{7}$$

Furthermore, by Proposition 2.1 we get that, with probability at least $1 - \delta$, simultaneously for all $w \in \mathbb{R}^N$,

$$\|\hat{a}_\delta(w) - \Sigma w\| \lesssim \|\Sigma\|\sqrt{\frac{\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma) + \log(1/\delta)}{T}}\|w\|\,. \tag{8}$$

Substituting these two error terms into the above lemma, we arrive at the following result.

Corollary 3.1. *Suppose, we are given independent $X_1, \ldots, X_T$ that have mean μ and covariance Σ, and the distribution satisfies the bounded kurtosis assumption (5). Let $\kappa = \lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$ denote the condition number. There is an absolute constant $C > 0$, such that the following holds. If $\delta \in (0, 1)$ satisfies,*

$$T \geq C\kappa^2\left(\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma) + \log(1/\delta)\right), \tag{9}$$

then there is an estimator $\hat{w}_\delta$ that depends on T observations such that, with probability at least $1 - \delta$,

$$\max_{w^\top\mathbf{1}=1} M_\gamma(w;\Sigma,\mu) - M_\gamma(\hat{w}_\delta;\Sigma,\mu) \lesssim \gamma\kappa^2\left(1 + \gamma^2\|\Sigma\|\|w^*\|^2\right)\frac{\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma) + \log(1/\delta)}{T}\,.$$

The proof is a simple substitution of error bounds (7), (8) into Lemma 3.1. We postpone the derivations to Section 8.3.

The above result has a number of favorable properties:

- The estimator only requires $\mathcal{O}(\log T)$ gradient descent updates. In addition, the amount of steps only has to be sufficiently large, i.e., there is no danger of overfitting by running the gradient descent for too long;
- the bound scales with $1/T$ when all the other parameters are fixed. In the optimization literature, this is regarded as a fast rate convergence. This rate is typical for strongly convex stochastic optimization problems;
- the value $\|w\|^2$ is often considered as a diversification measure of an allocation strategy, see (Strongin et al.; 2000). For instance, for the EW portfolio its value is $1/N$. One may expect it to be very small for the optimal portfolio.

However, the dependence on the condition number of the covariance matrix outweights all some of these useful properties. It is straightforward to verify that $\kappa \mathbf{r}(\Sigma) \geq N$. Hence, the above result only works in the setting, where the number of observations T is greater than the dimension N. The remaining term κ may further worsen the bound, so our result is rather limited to well conditioned covariance matrices. Unfortunately, it is rarely the case in practice: for our two datasets we estimate that $\mathbf{r}(\Sigma) < 3$ for S&P100 and $\mathbf{r}(\Sigma) < 7$ for Russell3000. The naive lower bound $\kappa \geq N/\mathbf{r}(\Sigma)$ yields that $\kappa \geq 27$ for S&P100 and $\kappa > 250$ for Russell3000. Therefore, our result does not contradict a commonly accepted evidence that MV portfolios perform poorly even when T is moderately larger than N (Ao et al.; 2019).

4 Ill-conditioned case

We now consider the case where we have no control over the condition number of Σ and it can even be degenerate. We will state the bound in the regime where only the effective

rank $\mathbf{r}(\Sigma)$ has to be small, and no requirements on the total dimension N are needed. In this section, we only consider the GMV portfolio.

For minimizing the GMV risk $R(w;\Sigma)$, we consider the following updates,

$$w_s = \Pi_1\left[w_{s-1} - \eta\hat{a}(w_{s-1})\right], \qquad s = 1,\dots,S$$

where $\hat{a}(w)$ is some family of estimators for the action of covariance operator Σw. Similarly to the previous section, we first show a deterministic result that controls the convergence through the error of this estimator.

Lemma 4.1. *Denote, $w^* = \arg\min_{w^\top\mathbf{1}=1} R(w;\Sigma)$. Suppose that we have an access to a family of estimators $\hat{a}(w)$ satisfying uniformly for all $w \in \mathbb{R}^d$,*

$$\|\hat{a}(w) - \Sigma w\|_2 \le \Delta_\Sigma \|w\|_2\,.$$

Assume that $\eta \le 1/\lambda_{\max}$ and let the number of steps S satisfies $S\Delta_\Sigma \le 1$. Then,

$$R(w_S;\Sigma) - R(w^*;\Sigma) \lesssim \max\{\|w_0 - w^*\|, \|w^*\|\}^2 \left(\frac{1}{\eta S} + \eta\Delta_\Sigma^2 S\right).$$

For the optimal choice $S \sim 1/(\eta\Delta_\Sigma)$, we have

$$R(w_s;\Sigma) - R(w^*;\Sigma) \lesssim \max\{\|w_0 - w^*\|, \|w^*\|\}^2 \Delta_\Sigma.$$

Proof. It is well known that the true minimum of the risk is $w^* = \Sigma^{-1}\mathbf{1}/(\mathbf{1}^\top\Sigma^{-1}\mathbf{1})$. Observe also that $\Pi_0\Sigma w^*$ vanishes. Write $\Delta(w) = \Pi_0[\Sigma w - \hat{a}(w)]$. Then,

$$\begin{aligned}
\|w_{s+1} - w^*\|^2 &= \|w_s - w^* - \eta\Pi_0\Sigma w_s - \eta\Delta(w_s)\|^2 \\
&= \|w_s - w^*\|^2 - 2\eta[\Sigma w_s]^\top(w_s - w^*) + \eta^2\|\Pi_0\Sigma w_s\|^2 + \eta^2\|\Delta(w_s)\|^2 \\
&\quad - 2\eta\Delta(w_s)^\top(w_s - w^* - \eta\Pi_0\Sigma w_s) \\
&= \|w_s - w^*\|^2 - 2\eta[\Sigma w_s]^\top(w_s - w^*) + \eta^2\|\Pi_0\Sigma w_s\|^2 - \eta^2\|\Delta(w_s)\|^2 \\
&\quad - 2\eta\Delta(w_s)^\top(w_{s+1} - w^*)\,.
\end{aligned}$$

It is easy to see that (which in more general terms is due to the convexity and smoothness of our objective to be minimized),

$$\begin{aligned}
-2\eta[\Sigma w_s]^\top(w_s - w^*) + \eta^2\|\Pi_0\Sigma w_s\|^2 &= -2\eta(w_s - w^*)^\top\Sigma(w_s - w^*) + \eta^2\|\Pi_0\Sigma w_s\|^2 \\
&\le \left(-2\frac{\eta}{\|\Sigma\|} + \eta^2\right)\|\Pi_0\Sigma w_s\|^2\,.
\end{aligned}$$

Using the condition $\eta \leq 1/\|\Sigma\|$ this sums up to at most zero. Combining the remaining terms we arrive at the inequality,

$$\|w_{s+1} - w^* + \Delta(w_s)\|^2 \leq \|w_s - w^*\|^2 .$$

Applying further the triangle inequality, we have that

$$\begin{aligned}\|w_{s+1} - w^*\| &\leq \|w_s - w^*\| + \|\Delta(w_s)\| \leq (1 + \Delta_\Sigma)\|w_s - w^*\| + \Delta_\Sigma\|w^*\| \\ &\leq \dots \\ &\leq (1 + \Delta_\Sigma)^{s+1} \left(\|w_0 - w^*\| + s\Delta_\Sigma\|w^*\|\right).\end{aligned}$$

Assume that $n\Delta_\Sigma \leq 1$. Then, using the inequality $(1+1/n)^n \leq e$, for each $s = 0, 1, \dots, n$,

$$\max(\|w_s\|, \|w_s - w^*\|) \leq (e+1)\left(\|w_0 - w^*\| + \|w^*\|\right) \stackrel{\text{def}}{=} M . \tag{10}$$

Further, we apply a standard trick for convex smooth optimization, see e.g., Theorem 3.5 in Bubeck (2014). Let us denote $R^*(w) = R(w; \Sigma)$, which is a convex and $\|\Sigma\|$-smooth function. Therefore, it holds that for any u, w,

$$0 \leq R^*(u) - R^*(w) - \nabla R^*(w)(u - v) \leq \frac{\|\Sigma\|}{2}\|w - u\|^2 . \tag{11}$$

Applying this inequality for w_s and $w_{s+1} = w_s - \eta\Pi_0\Sigma w_s - \eta\Delta(w_s)$, we first obtain that

$$\begin{aligned}R^*(w_{s+1}) - R^*(w_s) &\leq -\eta(\Sigma w_s)^\top \left[\Pi_0\Sigma w_s + \Delta(w_s)\right] + \frac{\eta^2\|\Sigma\|}{2}\left\|\Pi_0\Sigma w_s + \Delta(w_s)\right\|^2 \\ &\leq -\eta(\Sigma w_s)^\top \left[\Pi_0\Sigma w_s + \Delta(w_s)\right] + \frac{\eta}{2}\left\|\Pi_0\Sigma w_s + \Delta(w_s)\right\|^2 \\ &= -\frac{\eta}{2}\|\Pi_0\Sigma w_s\|^2 - \frac{\eta}{2}(\Sigma w_s)^\top\Delta(w_s) + \frac{\eta}{2}[2\Pi_0\Sigma w_s + \Delta(w_s)]^\top\Delta(w_s) \\ &= -\frac{\eta}{2}\|\Pi_0\Sigma w_s\|^2 + \frac{\eta}{2}[\Pi_0\Sigma w_s + \Delta(w_s)]^\top\Delta(w_s) \\ &\leq -\frac{\eta}{2}\|\Pi_0\Sigma w_s\|^2 + \frac{\eta}{2}\|\Pi_0\Sigma w_s\|\|\Delta(w_s)\| + \frac{\eta}{2}\|\Delta(w_s)\|^2 \\ &\leq -\frac{\eta}{4}\|\Pi_0\Sigma w_s\|^2 + \frac{3\eta}{4}\|\Delta(w_s)\|^2 .\end{aligned}$$

Observe that due to (11),

$$R^*(w_s) - R^*(w^*) \leq (\Sigma w_s)^\top(w_s - w^*) \leq \|\Pi_0\Sigma w_s\|\|w_s - w^*\| \leq M\|\Pi_0\Sigma w_s\|,$$

where in the last inequality we also use the bound (10). Furthermore, $\|\Delta(w_s)\| \leq \Delta_\Sigma M$. Denoting $\delta_s = R^*(w_s) - R^*(w^*)$, we obtain the recursive inequality,

$$\delta_{s+1} \leq \delta_s - \frac{\eta}{4M^2}\delta_s^2 + \eta\Delta_\Sigma^2 M^2 .$$

Denoting additionally $\alpha_s = \max\{0, \delta_s - s\eta\Delta_\Sigma^2 M^2\}$, we can easily derive that $\alpha_{s+1} \leq \max\left\{0, \alpha_s - \frac{\eta}{4M^2}\alpha_s\right\}$. It is straightforward to check that $\alpha_0 \leq \|\Sigma\|M^2 \leq \frac{4M^2}{\eta}$ and $\alpha_{s+1} \leq \alpha_s$. Therefore, we conclude that we can drop the positive part, so that $\frac{1}{\alpha_s} \leq \frac{1}{\alpha_{s+1}} - \frac{\eta}{4M^2}\frac{\alpha_s}{\alpha_{s+1}} \leq \frac{1}{\alpha_{t+1}} - \frac{\eta}{4M^2}$. Hence, follows the bound $\frac{1}{\alpha_t} \geq \frac{\eta}{4M^2}t$. Therefore, the following inequality holds

$$R^*(w_s) - R^*(w^*) = \delta_t \leq \alpha_t + t\eta\Delta_\Sigma^2 M^2 \leq \frac{4M^2}{\eta t} + t\eta\Delta_\Sigma^2 M^2\,,$$

which completes the proof. □

Once again, we plug our estimator $\hat{a}_\delta(w)$ into the update rule. In addition, we take the initial approximation to be an EW portfolio. Namely, our sequence is as follows

$$w_0 = N^{-1}\mathbf{1}, \qquad w_s = \Pi_1[w_{s-1} - \hat{a}_\delta(w_{s-1})], \qquad s = 1, \ldots, S\,. \tag{12}$$

Corollary 4.1. *Suppose that Assumption 2.1 holds. Take* $\eta = 1/\lambda_{\max}$ *and* $S \sim T(\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma) + \log(1/\delta))$*, and set* $\hat{w}_\delta = w_S$*. Then, with probability at least* $1-\delta$*,*

$$R(\hat{w}_\delta; \Sigma) - R(w^*; \Sigma) \lesssim \|\Sigma\|\|w^*\|^2\sqrt{\frac{\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma) + \log(1/\delta)}{T}}\,.$$

Proof. Simply substitute the bound (8) into Lemma 4.1. We also notice that $\|w^* - w_0\|^2 \leq 2\|w^*\|^2 + 2\|w_0\|^2$, and that $\|w^*\|^2 \geq (\mathbf{1}^\top w^*)^2/N = \|w_0\|^2$. □

Remark 4.1. *We remark that the scaling value* $\|\Sigma\|\|w^*\|^2$ *is only an upperbound on the optimal risk* $R(w^*;\Sigma) = \frac{1}{2}(w^*)^\top\Sigma w^*$ *and we cannot guarantee a ratio-type bound of the form* $R(\hat{w}_\delta;\Sigma) = (1+o(1))R(w^*;\Sigma)$*. However, this is not uncommon. For instance, Fan et al. (2012) shows that a portfolio with GEC constraints* $\sum_i |w_i| \leq C$ *satisfies,*

$$R(\hat{w};\Sigma) - R(w^*;\Sigma) \leq (1+C)^2 \max_{ij}|\hat{\Sigma}_{ij} - \Sigma_{ij}|,$$

where one typically has a bound $\max_{ij}|\hat{\Sigma}_{ij} - \Sigma_{ij}| \lesssim \sqrt{(\log N)/T}$*. Our bound may be a lot more beneficial if the optimal portfolio is well-diversified (i.e.,* $\|w^*\|^2 \sim 1/N$*), even though we do not impose any restrictions on the selected portfolio.*

5 Evaluation of empirical results

To test the performance of our approach, we apply it to two data sets of stocks. The first data set consists of 81 constituents of S&P100 index (as on January 1, 2021) and covers time span from January 2, 2000 to December 31, 2020 summing up to 5282 daily log-returns. These 81 stocks have a continuous return time-series over the period of our study. The second data set consists of 600 random constituents of Russell3000 index as on January 1, 2021, period of analysis is limited by 11 years: from January 2, 2010 to December 31, 2020. The length of analyzed time series is 2768 observations.

For the portfolio construction, we employ a rolling-window approach with monthly rebalancing. Specifically, we choose an estimation window of length T days starting on date $T+1$, for each rebalancing period l ($l = 1 \dots L$, with L the number of rebalancing periods) we use the data in the previous T days to estimate the parameters required to implement a particular strategy. For the S&P100 data set, the input parameters are estimated using the most recent 12 months' daily returns, corresponding roughly to 252 daily returns of past data (with the length of estimation window $T = 252$). As for Russell3000 data set, the input parameters are estimated according to benchmark portfolio policies, using the past 24 months of daily returns or roughly 500 daily returns ($T = 500$). Thus, the out-of-sample period for S&P100 data set starts on January 2, 2011 with length 5031 observations what corresponds to number of rebalancing periods $L = 228$ and for Russell3000 data set - on January 3, 2012 with 2265 out-of-sample observations corresponding to $L = 108$. The source for both data sets is Thompson Reuters.

5.1 Benchmark portfolios

Here we present the empirical results for GMV portfolio and evaluate its relative performance. The allocation rules included into the empirical study with corresponding reference and abbreviation are listed in Table 1.

Equally weighted (EW). DeMiguel and Nogales (2009) argue that a naive allocation strategy with weights $w_i = 1/N$ is hard to outperform in practice. It is often used as a benchmark for comparative analysis.

Sample-based Global minimum portfolio (GMV). This is the most straightforward way to GMV optimization. The sample covariance matrix $\hat{\Sigma}$ is plugged into the objective in (1). We should note that this strategy is only included for S&P100 data set, since for the Russell3000 we have $N > T$, and the sample covariance matrix is not invertible.

Global minimum portfolio with short-sale constraint (GMV_long). This portfolio is a sample-based GMV portfolio with only long positions allowed. This means that GMV objective corresponding to the empirical covariance is optimized subject to the constraints $w_j \geq 0$.

Global minimum portfolio with linear shrinkage estimator (GMV_lin). Ledoit and Wolf (2004b) propose an asymptotically optimal convex linear combination of the sample covariance matrix $\hat{\Sigma}$ with the identity matrix. Optimality is meant with respect to a quadratic loss function, asymptotically, as the number of observations and the number of assets go to infinity together. Ledoit and Wolf (2004b) use as a covariance matrix estimator as a convex linear combination of the sample covariance matrix and the identity matrix (shrinkage target) as follows:

$$\hat{\Sigma} = \rho I + (1 - \rho)S,$$

where ρ is the shrinkage intensity parameter and S is the sample covariance matrix. Their R package code is used in this horse race exercise.

Global minimum portfolio with non-linear shrinkage estimator (GMV_nlin). Ledoit and Wolf (2017) use the spectral decomposition for the empirical covariance

$$\hat{\Sigma} \stackrel{\text{def}}{=} U \widehat{D} U^{\top}$$

Model	Reference	Abbreviation
Equally weighted	DeMiguel et al. (2009)	EW
Robust Global Minimum Variance		GMV_robust
GMV with sample covariance	Merton (1980)	GMV
GMV with linear shrinkage cov estimator	Ledoit and Wolf (2004b)	GMV_lin
GMV with non-linear shrinkage cov estimator	Ledoit and Wolf (2017)	GMV_nlin
GMV with short sale constraint	Jagannathan and Ma (2003)	GMV_long

Table 1: Benchmark portfolios

where $\widehat{D} \overset{\text{def}}{=} \text{diag}\left(\widehat{d}(\lambda_1), \ldots, \widehat{d}(\lambda_N)\right)$, where $\lambda_1, \ldots, \lambda_N$ are the sample eigenvalues, and $\hat{d}$ is some nonlinear cutoff threshold based on N/T and the magnitude of the eigenvalues λ_j.

5.2 Performance measures

We report the following five out-of-sample performance measures for each benchmark portfolio rule.

- *Cumulative wealth (CumWealth)*

  CumWealth generated by each benchmark strategy with initial investment $W_0 = 1USD$ is computed as follows:

$$W_{l+1} = W_l + \hat{w}_l^\top X_{l+1}.$$

- *Sharpe ratio (SR)*

  To measure a risk adjusted performance, we compute Sharpe ratios (SR) for every strategy as a fraction of annualized average return of out-of-sample returns series to annualized standard deviation, showing the excess wealth which investor earns for accepting of every additional unit of risk. Such approach assumes implicitly setting the risk-free rate to 0, see e.g. (Ledoit and Wolf; 2017).

$$SR = \frac{AV}{SD},$$

where AV and SD are average out-of-sample returns and their standard deviations for each strategy multiplied by 252 and $\sqrt{252}$ respectively to annualize.

- *Turnover (TO)*

  The main practical objective of the introduced methodology is stabilizing of portfolio weights, aiming at reduction of transaction costs. To assess the impact of potential trading costs associated with portfolio rebalancing, we calculate two measures for turnover. First, following DeMiguel et al. (2009) and DeMiguel and Nogales (2009), we present Turnover, which is defined as an average sum of the absolute value of the rebalancing trades across the N assets of the investment universe and over the L rebalancing months (13).

$$TO = L^{-1} \sum_{l=1}^{L} \sum_{j=1}^{N} |\hat{w}_{j,l+1} - \hat{w}_{j,l+}| \,. \tag{13}$$

  where $\hat{w}_{j,l}$ and $\hat{w}_{j,l+1}$ are the weights assigned to the asset j for rebalancing periods l and $l+1$ and $\hat{w}_{j,l+}$ denotes its weight just before rebalancing at $l+1$. Thus, one accounts for the price change over the period, as one needs to execute trades to rebalance the portfolio towards the w_l target. High turnover will imply significant transaction costs; consequently, the lower TO of a strategy, the less its performance would be harmed by non-zero transaction costs.

- *Target Turnover (TTO)*

  Further, following Petukhina et al. (2021) we also calculate a target turnover, which is constructed as follows.

$$TTO = L^{-1} \sum_{l=1}^{L} \sum_{j=1}^{N} |\hat{w}_{j,l+1} - \hat{w}_{j,l}| \,.$$

  In contrast to equation (13) this definition of turnover implies by construction a value of zero for the EW portfolio. We provide this measure to focus on modifications of the target portfolio weights due to active management decisions and cleaned from the influence of assets' price dynamics.

Since the focus of this research is the reduction of portfolio weights' fluctuations, following Ledoit and Wolf (2017) we also compute the following five characteristics of weights' vectors $\hat{w}_t$ averaged through number of rebalancing periods. Thus, we calculate *minimum weight (min)* for every benchmark strategy as follows:

$$\text{min} = \frac{1}{T_{\max} - T} \sum_{t=T}^{T_{\max}} \min(\hat{w}_t) \,.$$

We similarly compute maximum weight *(max)*, standard deviation *(sd)*, and range of weights *(max-min)*.

In addition, we provide MAD from EW portfolio *(mad-ew)*, which is defined as:

$$\text{mad-ew} = \frac{1}{T_{\max} - T} \sum_{t=T}^{T_{\max}} N^{-1} \sum_{j=1}^{N} \left| \widehat{w}_{j,t} - \frac{1}{N} \right| \,.$$

6 Empirical study

Discussion of S&P100 data set results

First, we discuss portfolio weight stability, since it is the main focus of the research. Figure 2 demonstrates the dynamics of weights for S&P100. It can be observed that weights of plug-in GMV portfolio are characterized by a lot of extremes in comparison with all other policies and can vary from less than -40% to over 50 %. The least dispersed weights are observed for, introduced in this paper, GMV_robust approach. This visual result is confirmed by descriptive statistics of portfolio weights reported in the Table 2. It can be found that the average range of weights for GMV_robust 0.11 is the lowest one, what is twice less than the range of GMV_lin, GMV_nlin and GMV_long and almost four times less than plug-in GMV policy. $mad - ew$ also is the lowest for GMV_robust strategy, pointing out the more balanced distribution of weights. Table 2 reports these results, which can be summarized as follows.

First, the main two characteristics of the interest for this research would be *Turnover*

and *Target Turnover*. Not surprisingly, the best performing policy in this dimension is GMV_long with imposed non-negative constraints; it requires on average almost 14% (TO) of trading volume to rebalance the portfolio. GMV_long is followed by EW with 21% and GMV_robust with 37%. The highest turnover is reached by GMV with almost 95% of portfolio value to rebalance the portfolio to target weights $\hat{w}_t$. As for TTO we can conclude that cleaned from stochastic price dynamics the GMV_robust performs as good as GMV_long. Namely, 0.02 % of trading volume necessary for rebalancing is caused by change of computed GMV_long and GMV_robust portfolio policies, compared with 0.06% for GMV_lin or 0.09% for GMV.

Naturally, for investors, cumulative wealth (CumWealth) is of high interest as a measure of performance for the period considered. The best performing portfolio is GMV with 300% of gained value, followed by GMV_lin and GMV_nlin. GMV_robust earns 282% of initial portfolio value, what is higher than GMV_long and EW. The evolvement of cumulative wealth of all benchmark strategies for the considered period is plotted in Figure 1.

In terms of risk-adjusted performance, the winning strategies are shrinkage strategies with annual Sharpe ratio 41.9% for GMV_lin and 41.38% for GMV_nlin. Thus, GMV_robust gains a comparable 40.00% of excess return for taking an extra unit of risk.

Discussion of Russell3000 data set results

Outcomes of weights stability analysis are consistent with ones described for S&P100 data set. GMV_robust weights are characterized by harmonized weights without extreme short or long values. It is visible in the Figure 4 and in the Table 5: GMV_robust $MAD-EW$ and $max-min$ are the lowest in comparison with benchmark portfolios (excluding EW).

In terms of accumulated wealth, GMV_robust for large portfolios performs very close to shrinkage estimators, Table 4 summarizes investment performance characteristics. Thus, GMV_robust gains in the end of the period 201.7% of initial value while GMV_lin and GMV_nlin 209% and 201.9%. But considering non-zero trading fees would change the rank drastically. For example, transaction costs rate on the level of 50 basis points,

likewise in DeMiguel et al. (2009), would reduce the cumulative wealth of GMV_lin to 140% and to 172 % for GMV_robust or 177% for EW. Such disproportional reduction occurs due to the prominent difference in *Turnover* values: for GMV_robust it is 0.54 while for GMV_lin – 1.27.

As for risk-adjusted performance, the annual Sharpe ratio of GMV_robust is almost 9% lower than Shrinkage portfolio rules but at the same time much higher than EW or GMV_long policies.

Thus, according to outcomes of empirical experiments we can claim that GMV_robust portfolio policy achieves its goal and substantially reduces fluctuations of weights, leading to the lowest level of accumulated transaction costs. The risk adjusted performance is equal or slightly lower than shrinkage benchmarks and higher than constrained rules. This conclusion stays robust for small and large portfolios.

	CumWealth	SR	TTO	TO
EW	2.6635	24.9370	0.0000	0.2124
GMV_robust	2.8282	40.0069	0.0002	0.3695
GMV_long	2.6719	36.7571	0.0002	0.1393
GMV_lin	2.9701	41.9005	0.0006	0.7000
GMV_nlin	2.9038	41.3884	0.0005	0.6190

Table 2: Out-of-sample performance of benchmark portfolios, 81 stocks of S&P100, monthly rebalancing. Time period: 20010101 - 20201231
RobustM_PerformanceSP100

7 Conclusion and discussion

"Robustifying Markowitz" has seen many attempts that are mostly based on robustifying the original simple inversion formula for exact determination of optimal GMV weights. In bypassing this "error maximizing" technique, we have presented a tool fixing the portfolio weights in a low cost re-balancing ballpark. Using modern results from robust statistics, we have constructed an algorithm that provides a computationally effective

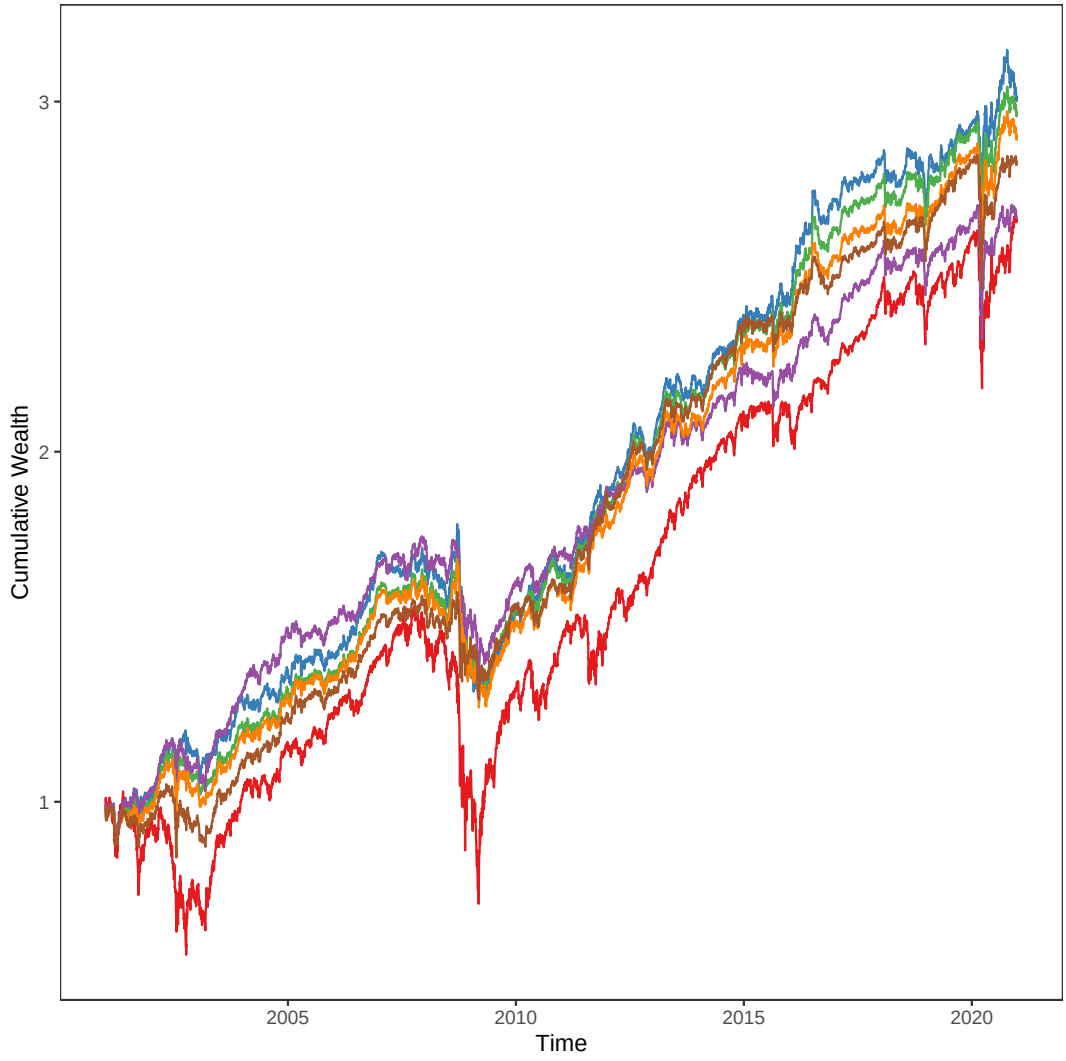

Figure 1: Cumulative wealth of benchmark portfolios EW, GMV, GMV_lin, GMV_long, GMV_nlin, GMV_robust for 81 stocks from S&P100, 20010101-20201231

RobustM_PerformanceSP100

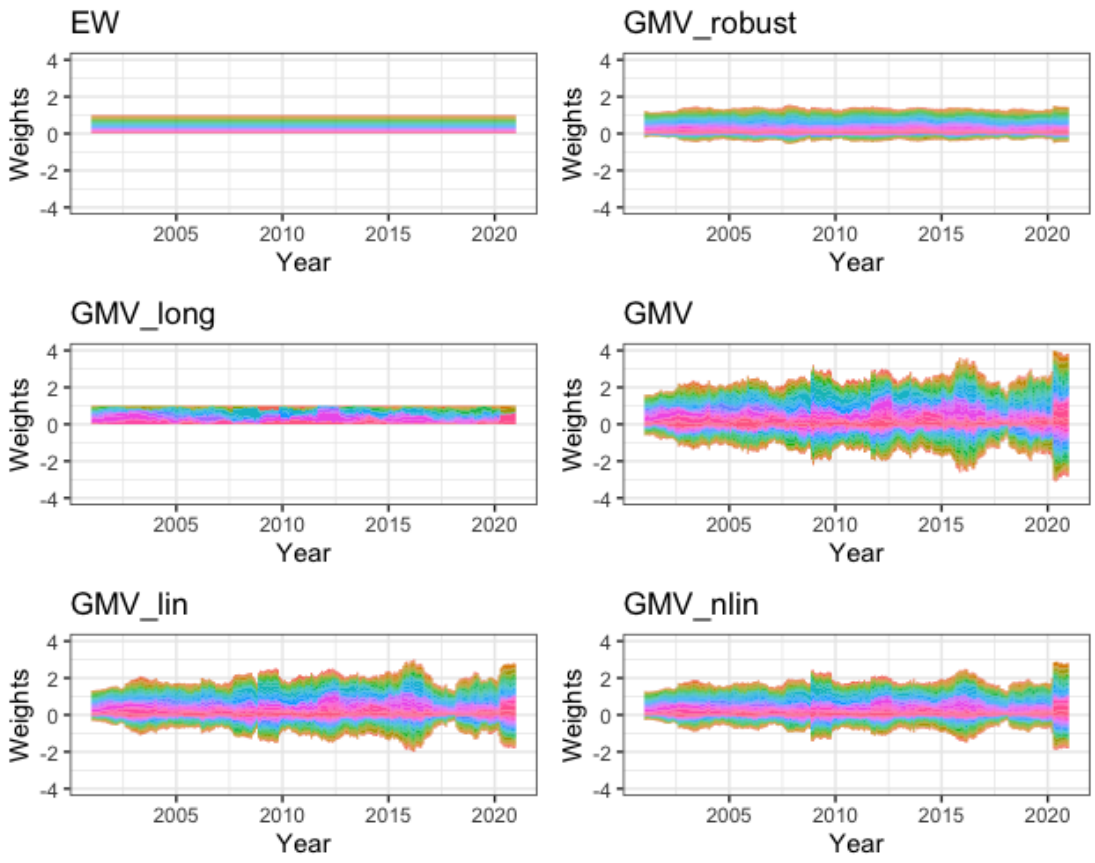

Figure 2: Weights of assets in GMV portfolios, 81 S&P100 constituents, 20010101 - 20201231

RobustM_PerformanceSP100

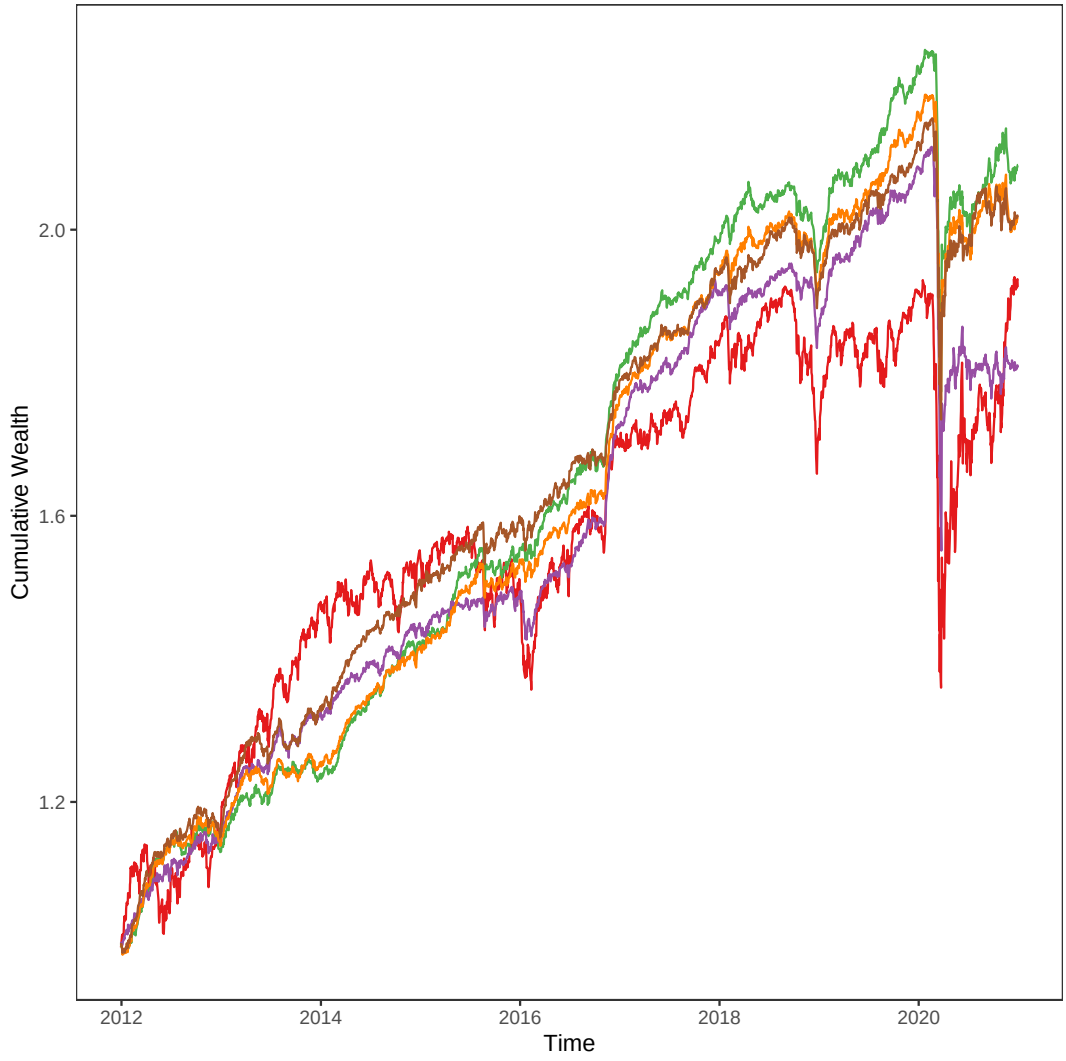

Figure 3: Cumulative wealth of benchmark portfolios EW, GMV_lin, GMV_long, GMV_nlin, GMV_robust, 600 random constituents of Russell3000, 20120101 - 20201231

RobustM_PerformanceRussell3000

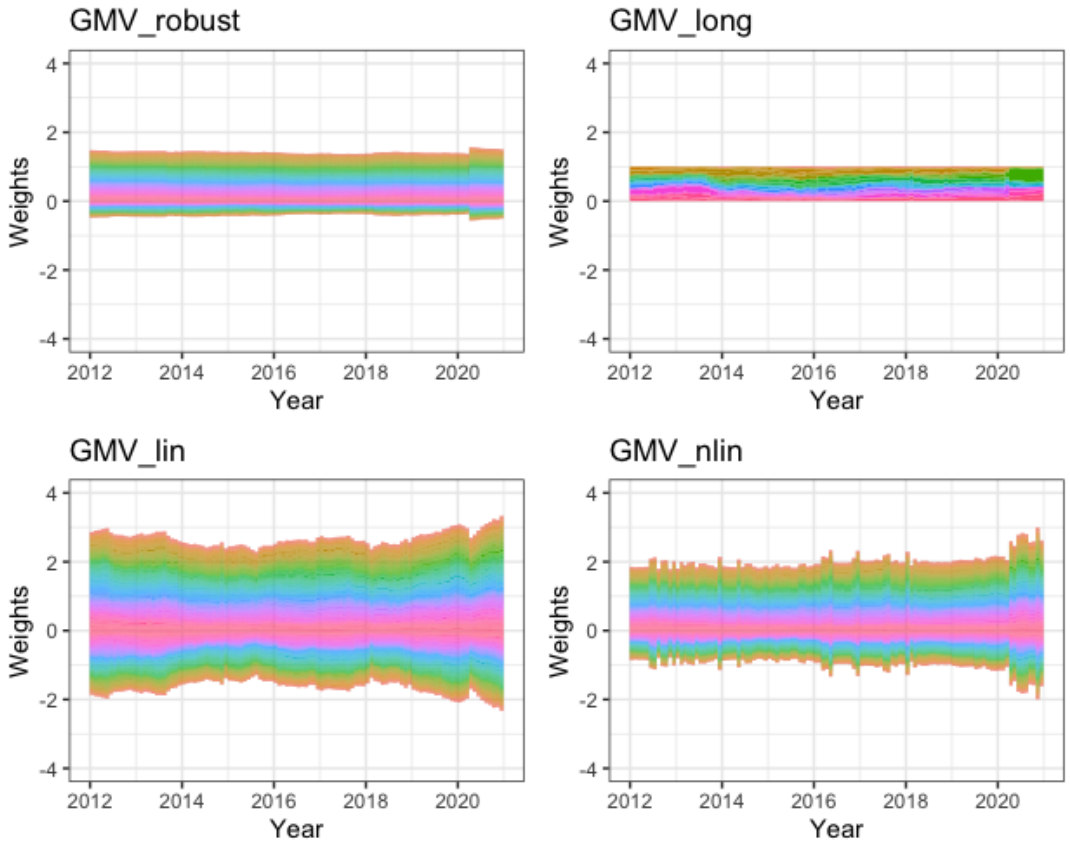

Figure 4: Weights of assets in GMV portfolios, 600 Russell3000 constituents, 20120101 - 20201231

RobustM_PerformanceRussell3000

	min	max	sd	mad-ew	max-min
EW	0.0123	0.0123	0.0000	0.0000	0.0000
GMV_robust	-0.0461	0.0691	0.0240	0.0192	0.1153
GMV_long	0.0000	0.2484	0.0389	0.0201	0.2484
GMV_lin	-0.0995	0.1793	0.0497	0.0374	0.2788
GMV_nlin	-0.0832	0.1478	0.0423	0.0323	0.2310

Table 3: Average characteristics of the weight vectors of GMV portfolios, 81 stocks of S&P100, monthly rebalancing. Time period: 20010101 - 20201231

RobustM_PerformanceSP100

	CumWealth	SR	TTO	TO
EW	1.9191	38.3565	0.0000	0.2682
GMV_robust	2.0172	70.5979	0.0000	0.5488
GMV_long	1.8065	57.8239	0.0000	0.1974
GMV_lin	2.0905	79.1200	0.0001	1.2729
GMV_nlin	2.0186	78.5869	0.0001	0.9181

Table 4: Out-of-sample performance of benchmark portfolios, 600 stocks of Russell3000, monthly rebalancing. Time period: 20120101 - 20201231

RobustM_PerformanceRussell3000

estimator for GMV Markowitz portfolios. We have shown that it suffices to utilize a PGD procedure to optimize the portfolio weights without estimating the covariance operator itself. The focus on just the PGD updates significantly distinguishes our approach from the previous techniques. We have successfully derived almost Gaussian properties of this estimator in nice (N/T small) and not so nice (N/T big) condition cases.

The weights developed with the robustified approach are less sensitive to deviations of the asset-return distribution from normality than those of the traditional minimum-variance policy. Empirical studies confirm that the proposed policies are indeed more stable and cost reducing. The stability of the proposed portfolios makes them a feasible alternative to traditional portfolios.

	min	max	sd	mad-ew	max-min
EW	0.0017	0.0017	0.0000	0.0000	0.0000
GMV_robust	-0.0103	0.0164	0.0038	0.0029	0.0267
GMV_long	0.0000	0.1382	0.0100	0.0031	0.1382
GMV_lin	-0.0290	0.0430	0.0095	0.0073	0.0720
GMV_nlin	-0.0185	0.0319	0.0067	0.0050	0.0503

Table 5: Average characteristics of the weight vectors of GMV portfolios, 600 stocks of Russell3000, monthly rebalancing. Time period: 20120101 - 20201231

- RobustM_PerformanceRussell3000

The proposed toolbox improves stability properties of weights, leading to better investment characteristics of allocation policies. The "Robustifying Markowitz" algorithm outperforms conventional minimum-variance portfolios in terms of their out-of-sample Sharpe ratios due to substantial reduction of trading volume measured by turnover. Finally, these performance results are confirmed across small and large portfolios. Even for dimensions of portfolio size larger than the length of estimation window (e.g. the Russell3000 data) the above claim pertains.

8 Proofs

8.1 Proof of Proposition 2.1

Before we proceed, let us recall some of the results and definitions from (Lugosi and Mendelson; 2019b) and (Hopkins et al.; 2020). We start by giving the definition of combinatorial centers, which is the central object in the original construction of Lugosi and Mendelson, but the definition itself is due to Hopkins, Li and Zhang.

Definition 8.1 (Combinatorial center)**.** *A point $\theta \in \mathbb{R}^N$ is called a (r,κ)-combinatorial center of $Y_1,\dots,Y_\ell$ if for all unit vectors $v \in \mathbb{R}^N$, the inequality*

$$|v^\top (Y_j - \theta)| \leq r$$

takes place for at least $(1-\kappa)\ell$ of indices $j = 1,\dots,\ell$.

Essentially, Lugosi and Mendelson (2019b) prove that for appropriately chosen r_δ, the true mean is a $(r_\delta, 1/4)$-combinatorial center with probability at least $1-\delta$, where Y_j being the bucket means. The estimation strategy is then executed by what is called a *median-of-means tournament*: one needs to pick an $(r, 1/4)$-combinatorial center with r as small as possible. The deviation bound then follows by a simple triangle inequality. One difficulty of implementing this strategy computationally is the lack of control on how these subsets of indices behave for different directions $v \in \mathbb{R}^N$.

In addition, Hopkins et al. (2020) define the *spectral center* of bucket means, which can serve as a relaxation of the combinatorial one.

Definition 8.2 (Spectral center)**.** *For $\varepsilon \in (0, 1/2)$, denote*

$$\Delta_{\ell,\varepsilon} = \left\{ u \in \mathbb{R}^\ell : \sum_{j=1}^{\ell} u_j = 1, \;\; 0 \le u_j \le 1/\{\ell(1-\varepsilon)\} \right\}.$$

A point $\theta \in \mathbb{R}^N$ is called a (r, ε)-spectral center if there are weights $(u_1, \dots, u_\ell) \in \Delta_{\ell,\varepsilon}$ such that

$$\left\| \sum_{j=1}^{\ell} u_j (Y_j - \theta)(Y_j - \theta)^\top \right\| \le r^2.$$

It is straightforward to see that if θ is a (r, ε)-spectral center with minimal r, then it has the form $\theta = \sum_{j=1}^{\ell} u_j Y_j$ for some $(u_1, \dots, u_\ell) \in \Delta_{\ell,\varepsilon}$, i.e. the solution should be a weighted mean of Y_j. The two definitions are "equivalent" in the following sense.

Lemma 8.1. *Suppose that θ is (r, κ)-combinatorial center. Then it is also a $(5r, 10\kappa)$-spectral center. Conversely, if θ is an (r, ε)-spectral center, then it is also a $(\sqrt{(1-\varepsilon)/\varepsilon}\, r, 2\varepsilon)$-combinatorial center.*

Hopkins et al. (2020) state this lemma for some particular constants ε and κ. Their proof consists of some arguments of the proof of Proposition 1 in (Depersin and Lecué; 2019). For the sake of completeness, we reproduce these arguments in Section 8.2, slightly changed.

We deal with both notions of centers for the following reason: it is easier to deal with the statistical properties of combinatorial centers, whereas the spectral centers are more

convenient from computational perspective. Hopkins, Li and Zhang (2020) develop an algorithm that finds a center with a spectral signature that is guaranteed at most (say) twice as large as minimal possible. Namely, we denote the output of their algorithm as $\mathsf{HLZ}(Y_1, \dots, Y_\ell; \varepsilon)$ and they show that the output $\hat{\mu}$ satisfies

$$\min_{u \in \Delta_{\ell,\varepsilon}} \left\| \sum_{j=1}^{\ell} u_j (Y_j - \hat{\mu})(Y_j - \hat{\mu})^\top \right\| \lesssim \min_{\theta} \min_{u \in \Delta_{\ell,\varepsilon}} \left\| \sum_{j=1}^{\ell} u_j (Y_j - \theta)(Y_j - \theta)^\top \right\| . \tag{14}$$

Hence our goal is to show that with high probability, the spectral signature of the true mean is sufficiently small, which we can do using the median-of-means analysis and switching back and forth between spectral and combinatorial centers.

Let us now give the description of the estimator $\hat{a}_\delta(w)$. It consists of the following steps:

1. First we centralize our observations. For this, consider the transformations $\widetilde{X}_1 = (X_1 - X_2)/\sqrt{2}, \widetilde{X}_2 = (X_3 - X_4)/\sqrt{2}, \dots$ Obviously, each of these new "observations" has zero mean and the same covariance as X_i, and moreover they are independent.

2. Fix $\varepsilon < 10/21$ and set $\ell = \lceil 2(\varepsilon/10)^{-2} \log(2/\delta) \rceil$. Split the observations $\widetilde{X}_1, \dots, \widetilde{X}_{\lfloor T/2 \rfloor}$ into ℓ non-intersecting buckets

$$B_1 \sqcup \dots \sqcup B_\ell = \{1, \dots, \lfloor T/2 \rfloor\}.$$

3. Next, using the data from each of the buckets, we construct the following covariance estimators,

$$\Sigma_j = \frac{1}{|B_j|} \sum_{i \in B_j} \widetilde{X}_i \widetilde{X}_i \mathbf{1}_{\|\widetilde{X}_i\| \le R}$$

4. For a given direction $w \in \mathbb{R}^N$, we output the result of the HLZ algorithm applied to the bucket means

$$\hat{a}_\delta(w) \stackrel{\text{def}}{=} \mathsf{HLZ}(\Sigma_1 w, \dots, \Sigma_\ell w; \varepsilon) .$$

Remark 8.1. *Notice that given $\varepsilon < 10/21$ and fixing, say, $\delta = 0.05$, we have that the number of buckets $\ell = \lceil 2(\varepsilon/10)^{-2} \log(2/\delta) \rceil$ has to be at least 1500, which makes the algorithm rather impractical. Unfortunately, the theory that we use does not allow more adequate constants. In the empirical study we heuristically choose $\varepsilon = 1/3$ and $\ell = 10$.*

Transforming the observations as $(X_1 - X_2)/\sqrt{2}, (X_3 - X_4)/\sqrt{2}, \dots$ is done for the sole purpose of centralization. We have done so by reducing the size of the sample by at most two. To avoid the notation overloading, we assume below that $\mu = 0$ and proceed to work with the original X_i.

Set $Z_i = X_i X_i^\top \mathbf{1}[\|X_i\| < R]$. According to step 2 of the algorithm, we split the data into ℓ blocks $B_1, \dots, B_\ell$ and consider the trimmed covariances,

$$\Sigma_j = \frac{1}{|B_j|} \sum_{i \in B_j} Z_i \,.$$

Below we derive the following bound: with probability $1 - \delta$, we have that for any directions v, w, the inequality

$$|u^\top \Sigma_j w - u^\top \Sigma w| \lesssim \Delta_\Sigma \stackrel{\text{def}}{=} \|\Sigma\| \sqrt{\frac{\mathbf{r}(\Sigma) \log \mathbf{r}(\Sigma) + \ell}{T}} \tag{15}$$

holds for at least $1 - \kappa$ fraction of the indices $j = 1, \dots, \ell$, where $\kappa = \varepsilon/10$ is fixed. Let us first complete the proof given this inequality.

At step 5 of the algorithm, we produce the vectors $Y_j = \Sigma_j w$. On the event from (15), we have that for any unit $u \in \mathbb{R}^N$, the inequality

$$|u^\top (Y_j - \Sigma w)| \le C \Delta_\Sigma$$

holds for at least $1 - \kappa$ a fraction of indices. Hence, Σw is a $(C\Delta_\Sigma, \kappa)$-combinatorial center of Y_j. By Lemma 8.1, it also means that Σw is a $(5C\Delta_\Sigma, 10\kappa)$-spectral center of Y_j. Hence, by (14), the output of $\hat{a}_\delta(w) = \mathsf{HLZ}(Y_1, \dots, Y_j; \varepsilon)$ is a $(C'\Delta_\Sigma, 10\kappa)$-spectral center, and using the second part of Lemma 8.1, we conclude that it is also a $(C'\sqrt{10/\kappa}\Delta_\Sigma, 20\kappa)$-combinatorial center. Since $21\kappa < 1$, we get that $1 - 20\kappa + 1 - \kappa > 1$, which means that for any direction $u \in \mathbb{R}^d$, by the pigeonhole principle, we can pick a single Y_j that is close to both combinatorial centers Σw and $\hat{a}_\delta(w)$ in this direction. Therefore, by the triangle inequality,

$$|u^\top (\hat{a}_\delta(w) - \Sigma w)| \lesssim \Delta_\Sigma,$$

and since the bound holds in arbitrary direction u, we get the required bound in Euclidean norm.

It remains to prove the bound (15).

Let $\widetilde{\Sigma} = \mathsf{E} Z_i$. We have by Lemma 2.1 of Mendelson and Zhivotovskiy (2020),

$$\|\Sigma - \widetilde{\Sigma}\| \lesssim \frac{\|\Sigma\|^2 \mathbf{r}(\Sigma)}{R^2}\,. \tag{16}$$

Let $Quant_\alpha(z_1 \dots z_\ell)$ of a sequence of real numbers denotes an order statistics $z_{(\lceil \alpha \ell \rceil)}$, where $z_{(1)} \dots z_{(k)}$ is a non-decreasing rearrangement of the original sequence. Then, we can rewrite (15) as follows,

$$\max\left\{Quant_{1-\kappa}(u^\top \Sigma_j w) - u^\top \Sigma w, u^\top \Sigma w - Quant_\kappa(u^\top \Sigma_j w)\right\} \lesssim \Delta_\Sigma\,.$$

Let us apply (Klochkov, Kroshnin and Zhivotovskiy; 2020, Lemma 2.3) to the class of functions $\{f_{u,w}(Y) = u^\top Y w\}$. We have that with probability at least $1 - 2e^{-\kappa^2 \ell/2}$,

$$\begin{aligned}\max\left\{Quant_{1-\kappa}(u^\top \Sigma_j w) - u^\top \Sigma w, u^\top \Sigma w - Quant_\kappa(u^\top \Sigma_j w)\right\} \\ \lesssim \mathsf{E} \sup_{u,w}\left(\frac{1}{T}\sum_{i=1}^N \varepsilon_i u^\top Y_i w\right) + \sqrt{\sup_{u,w} \mathsf{E}(u^\top Y_1 w)^2 \frac{\ell}{T}} + \|\widetilde{\Sigma} - \Sigma\|\,,\end{aligned}$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent Rademacher signs, i.e. taking ± 1 with probability $1/2$. The supremum in both terms of the RHS is over unit vectors u, w in $\mathbb{R}^N$. Let us first bound the second, weak term. We have that

$$\mathsf{E}(u^\top Y_1 w)^2 \le \mathsf{E}(u^\top X_1)^2 (w^\top X_1)^2 \le \mathsf{E}^{1/2}(u^\top X_1)^4 \mathsf{E}^{1/2}(w^\top X_1)^4\,.$$

By the L_4–L_2 equivalence assumption we get that $\mathsf{E}^{1/2}(u^\top X_1)^4 \lesssim \mathsf{E}(u^\top X_1)^2 \le \|\Sigma\|$. The weak term is therefore bounded by $C\|\Sigma\|\sqrt{\ell/T}$.

Now let us deal with the first, strong term. We rewrite it as follows,

$$\mathsf{E} \sup_{u,w}\left(\frac{1}{T}\sum_{i=1}^N \varepsilon_i u^\top Y_i w\right) = \mathsf{E} \sup_{u,w} u^\top \left(\frac{1}{T}\sum_{i=1}^N \varepsilon_i Y_i\right) w = \mathsf{E}\left\|\frac{1}{T}\sum_{i=1}^N \varepsilon_i Y_i\right\|$$

The right-most expression is the expected value of the norm of a sum of centered matrices $\varepsilon_i Y_i$, which are bounded by R^2 pointwise. We therefore can apply the Matrix Bernstein inequality, the details are carried out by Mendelson and Zhivotovskiy (2020) in Section 3. They show that this leads eventually to the bound

$$\mathsf{E}\left\|\frac{1}{T}\sum_{i=1}^T \varepsilon_i Y_i\right\| \lesssim \|\Sigma\|\sqrt{\frac{\mathbf{r}(\Sigma)\log \mathbf{r}(\Sigma)}{T}} + \frac{R^2 \log \mathbf{r}(\Sigma)}{T}\,.$$

Recalling the bound (16) we get that, with probability $1 - 2e^{-\kappa^2\ell/2}$,

$$\max\left\{Quant_{1-\kappa}(u^\top\Sigma_j w) - u^\top\Sigma w, u^\top\Sigma w - Quant_\kappa(u^\top\Sigma_j w)\right\} \lesssim \Delta_\Sigma + \frac{R^2\log\mathbf{r}(\Sigma)}{T} + \frac{\|\Sigma\|^2\mathbf{r}(\Sigma)}{R^2}\,.$$

For $R \sim \|\Sigma\|^{1/2}\left(\frac{T\mathbf{r}(\Sigma)}{\log\mathbf{r}(\Sigma)}\right)^{1/4}$ the RHS simplifies to $\Delta_\Sigma = \|\Sigma\|\sqrt{\frac{\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma)+\ell}{T}}$. It remains to notice that $1 - 2e^{-\kappa^2\ell/2} \geq 1-\delta$ as long as $\ell \geq 2\kappa^{-2}\log\left(\frac{2}{\delta}\right)$.

8.2 Proof of Lemma 8.1

Let us first recall the following basic fact from linear algebra: for a symmetric matrix A, its largest eigenvalue satisfies $\lambda_{\max}(A) = \sup_{M\succeq 0,\mathrm{Tr}(M)=1}\mathrm{Tr}(MA)$. Hence, we can rewrite

$$\min_{w\in\Delta_{\ell,\varepsilon}}\left\|\sum_{j=1}^{\ell} w_j(Y_j-\theta)(Y_j-\theta)^\top\right\| = \min_{w\in\Delta_{\ell,\varepsilon}}\max_{M\succeq 0,\mathrm{Tr}(M)=1}\sum_{j=1}^{\ell} w_j(Y_j-\theta)^\top M(Y_j-\theta).$$

The latter can be seen as *semi-definite program* (SDP) and using the strong duality of SDP one can show that the minimum over w and the maximum over M can be swapped (see formula (5.2) in (Hopkins et al.; 2020); see also (Depersin and Lecué; 2019; Diakonikolas et al.; 2020)):

$$\min_{w\in\Delta_{\ell,\varepsilon}}\max_{M\succeq 0,\mathrm{Tr}(M)=1}\sum_{j=1}^{\ell} w_j(Y_j-\theta)^\top M(Y_j-\theta) = \max_{M\succeq 0,\mathrm{Tr}(M)=1}\min_{w\in\Delta_{\ell,\varepsilon}}\sum_{j=1}^{\ell} w_j(Y_j-\theta)^\top M(Y_j-\theta)$$

The right-hand side form is closer to what Lugosi and Mendelson (2019b) do: for any direction $M = vv^\top$ we can pick its own weights. This property allows to show the equivalence.

We write $y_j = Y_j - \theta$ for short everywhere in this section.

First, assume that θ is a (r,κ)-combinatorial center. We will show by contradiction that it is also a (R,ε)-spectral center, where $R = 5r$ and $\varepsilon = 10\kappa$. Suppose it is not, so that for some $M \succeq 0$ with $\mathrm{Tr}(M) = 1$ we have that

$$\min_{w\in\Delta_{\ell,\varepsilon}}\sum_j w_j y_j^\top M y_j \geq R^2\,.$$

If $w \in \Delta_{\ell,\varepsilon}$ delivers the minimum it must put non-zero weights to at least $\lceil \ell(1-\varepsilon) \rceil$ terms. Since the weights sum up to one, we conclude that for at least $\lceil \ell\varepsilon \rceil$ indices $j = 1, \ldots, \ell$, it holds that $y_j^\top M y_j \geq R^2$. We denote this set of indices as B. Now, let $M = \sum_{k=1}^{\ell} \lambda_k u_k u_k^\top$ be its spectral decomposition. Since $M \succeq 0$ and $\mathrm{Tr}(M) = 1$, we have that $\sum_k \lambda_k = 1$ and $\lambda_k \geq 0$.

Let us take a random unit vector $v = \sum_k \sqrt{\lambda_k} u_k \varepsilon_k$, where ε_k are independent random signs, so that the equality $\sum_k \lambda_k = 1$ ensures that it is indeed a unit vector. Moreover,

$$y_j^\top v = \sum_k (\sqrt{\lambda_k} y_j^\top u_k)\varepsilon_k = \sum_k a_k^{(j)} \varepsilon_k,$$

where we denote $a_k^{(j)} = \sqrt{\lambda_k} y_j^\top u_k$, and we also denote by $a^{(j)} \in \mathbb{R}^\ell$ the vector with corresponding coordinates. Observe that for $j \in B$, we have that

$$\|a^{(j)}\|^2 = \sum_k \lambda_k y_j^\top u_k u_k^\top y_j = y_j^\top M y_j \geq R^2.$$

The Khintchin inequality due to Szarek (1976) states that,

$$\frac{1}{\sqrt{2}} \|a^{(j)}\| \leq \mathsf{E} \left| \sum_k a_k^{(j)} \varepsilon_j \right| \leq \|a^{(j)}\|.$$

Furthermore, the lower tail of the bounded differences inequality due to Theorem 6.9 in Boucheron et al. (2013) implies that

$$\mathsf{P}\left(\left| \sum_k a_k^{(j)} \varepsilon_j \right| < \frac{1}{\sqrt{2}} \|a^{(j)}\| - t \right) \leq e^{-\frac{t^2}{2(\|a^{(j)}\|^2 + t\|a^{(j)}\|/3)}}$$

Taking $t = \frac{1-c}{\sqrt{2}} \|a^{(j)}\|$, we get that

$$\mathsf{P}\left(\left| \sum_k a_k^{(j)} \varepsilon_j \right| \geq \frac{c}{\sqrt{2}} \|a^{(j)}\| \right) \geq 1 - e^{-\frac{(1-c)^2}{4(1+(1-c)/(3\sqrt{2}))}},$$

which for $c = \sqrt{2}/5$ is greater than 0.1. Hence, we can find a unit vector v such that for at least one tenth of the indices $j \in B$,

$$|y_j^\top v| \geq \frac{1}{5} R = r.$$

One tenth of B accounts for $0.1\varepsilon = \kappa$, hence θ cannot be an (r, κ)-combinatorial center.

Suppose that θ is a (r, ε)-spectral center. Again we will prove that it is also a (R, κ)-combinatorial center by contradiction, with $R = \sqrt{(1-\varepsilon)/\varepsilon}\, r$, $\kappa = 2\varepsilon$. Suppose it is not.

Then, there is a unit vector v, such that for strictly more than $\ell\kappa$ indices j, $|y_j^\top v| > R$. Denote this set of indices as B. Since θ is a spectral center, we get that for $M = vv^\top$,

$$\min_{w \in \Delta_{\ell,\varepsilon}} \sum_{j=1}^{\ell} w_j |y_j^\top v|^2 \le r^2$$

The minimum puts weight $1/(\ell(1-\varepsilon))$ for $\lfloor \ell(1-\varepsilon) \rfloor$ with the smallest values $|y_j^\top v|$. By the pigeonhole principle, strictly more than $\ell\kappa - \lceil \ell\varepsilon \rceil$ of them are in the set B. Hence,

$$\min_{w \in \Delta_{\ell,\varepsilon}} \sum_{j=1}^{\ell} w_j |y_j^\top v|^2 > \frac{\kappa - \varepsilon}{1-\varepsilon} R^2 = \frac{\varepsilon}{1-\varepsilon} R^2 = r^2.$$

This completes the proof by contradiction.

8.3 Proof of Corollary 3.1

Simply substitute $\Delta_\mu = C\|\Sigma\|^{1/2}\sqrt{\frac{\mathbf{r}(\Sigma)+\log(1/\delta)}{T}}$ and $\Delta_\Sigma = C\|\Sigma\|\sqrt{\frac{\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma)+\log(1/\delta)}{T}}$ into Lemma 3.1, and take $\eta = 1/(\gamma\lambda_{\max})$. The condition (9) ensures that $\Delta_\Sigma \le \lambda_{\min}/2$. We get that

$$\|w_s - w^*\| \le \left(1 - \frac{1}{2\kappa}\right)^s \|w_0 - w^*\| + C' \frac{\|\Sigma\|^{1/2} + \gamma\|\Sigma\|\|w^*\|}{\lambda_{\min}} \sqrt{\frac{\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma) + \log(1/\delta)}{T}}$$

Taking $s \sim \log T$ steps, the first term will be dominated by second one. Furthermore, since the objective is quadratic and w^* its optimum, we have that

$$\begin{aligned} M_\gamma(w^*;\Sigma,\mu) - M_\gamma(w_s;\Sigma,\mu) &= \frac{\gamma}{2}\|\Sigma^{1/2}(w_s - w^*)\|^2 \\ &\lesssim \frac{\gamma\|\Sigma\|^2 + \gamma^3\|\Sigma\|^3\|w^*\|^2}{\lambda_{\min}^2} \frac{\mathbf{r}(\Sigma)\log\mathbf{r}(\Sigma) + \log(1/\delta)}{T}, \end{aligned}$$

hence follows the bound.

Acknowledgments

Wolfgang Härdle and Alla Petukhina gratefully acknowledge the financial support of the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training programme" under the grant

agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA), the European Cooperation in Science & Technology COST Action grant CA19130 - Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry, the Deutsche Forschungsgemeinschaft's IRTG 1792 grant, Wolfgang Härdle - the Yushan Scholar Program of Taiwan, the Czech Science Foundation's grant no. 19-28231X / CAS: XDA 23020303.

References

Ao, M., Yingying, L. and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios, *The Review of Financial Studies* **32**(7): 2890–2919.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension, *Annals of Statistics* **40**(1): 436–465.

Bai, Z., Liu, H. and Wong, W.-K. (2009). Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory, *Mathematical Finance* **19**: 639 – 667.

Bartl, D. and Mendelson, S. (2021). On Monte-Carlo methods in convex stochastic optimization, *arXiv preprint arXiv:2101.07794* .

Best, M. and Grauer, R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results, *Review of Financial Studies* **4**: 315–42.

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters, *Annals of Operations Research* **45**(1): 21–58.

Bubeck, S. (2014). Convex optimization: Algorithms and complexity, *arXiv preprint arXiv:1405.4980* .

Chen, J. and Yuan, M. (2016). Efficient Portfolio Selection in a Large Market, *Journal of Financial Econometrics* **14**(3): 496–524.
URL: *https://doi.org/10.1093/jjfinec/nbw003*

Cherapanamjeri, Y., Flammarion, N. and Bartlett, P. L. (2019). Fast mean estimation with sub-Gaussian rates, *Conference on Learning Theory*, PMLR, pp. 786–806.

Cherapanamjeri, Y., Hopkins, S. B., Kathuria, T., Raghavendra, P. and Tripuraneni, N. (2020). Algorithms for heavy-tailed statistics: Regression, covariance estimation, and beyond, *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 601–609.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice, *The Journal of Portfolio Management* **19**(2): 6–11.
URL: *https://jpm.pm-research.com/content/19/2/6*

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The review of Financial Studies* **22**(5): 1915–1953.

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

Depersin, J. and Lecué, G. (2019). Robust subgaussian estimation of a mean vector in nearly linear time, *arXiv preprint arXiv:1906.03058* .

Diakonikolas, I., Kamath, G., Kane, D. M., Li, J., Moitra, A. and Stewart, A. (2017). Being robust (in high dimensions) can be practical, *International Conference on Machine Learning*, PMLR, pp. 999–1008.

Diakonikolas, I., Kane, D. M. and Pensia, A. (2020). Outlier robust mean estimation with subgaussian rates via stability, *arXiv preprint arXiv:2007.15618* .

Fan, J., Fan, Y. and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics* **147**(1): 186–197.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements, *Journal of the Royal Statistical Society. Series B, Statistical methodology* **75**(4).

Fan, J., Wang, W. and Zhong, Y. (2019). Robust covariance estimation for approximate factor models, *Journal of Econometrics* **208**(1): 5–22.

Fan, J., Zhang, J. and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints, *Journal of the American Statistical Association* **107**(498): 592–606.

Frost, P. A. and Savarino, J. E. (1986). An empirical bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* pp. 293–305.

Frost, P. A. and Savarino, J. E. (1988). For better performance: Constrain portfolio weights, *The Journal of Portfolio Management* **15**(1): 29–34.
URL: *https://jpm.pm-research.com/content/15/1/29*

Green, R. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified?, *The Journal of Finance* **47**(5): 1785–1809.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1992.tb04683.x*

Hopkins, S. B. (2018). Sub-Gaussian mean estimation in polynomial time, *arXiv preprint arXiv:1809.07425* p. 120.

Hopkins, S. B., Li, J. and Zhang, F. (2020). Robust and heavy-tailed mean estimation made simple, via regret minimization, *arXiv preprint arXiv:2007.15839* .

Huber, P. J. (2004). *Robust statistics*, Vol. 523, John Wiley & Sons.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps, *The Journal of Finance* **58**(4): 1651–1683.

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

Karoui, N. (2012). High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation, *Annals of Statistics* **38**.

Karoui, N. E. (2013). On the realized risk of high-dimensional markowitz portfolios, *SIAM Journal on Financial Mathematics* **4**(1): 737–783.

Ke, Y., Minsker, S., Ren, Z., Sun, Q. and Zhou, W.-X. (2019). User-friendly covariance estimation for heavy-tailed distributions, *Statistical Science* **34**(3): 454–471.

Klochkov, Y., Kroshnin, A. and Zhivotovskiy, N. (2020). Robust k-means clustering for distributions with two moments, *to appear in Annals of Statistics* .

Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators, *Bernoulli* **23**(1): 110–133.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* **10**(5): 603–621.

Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix, *The Journal of Portfolio Management* **30**(4): 110–119.

Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis* **88**(2): 365–411.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks, *The Review of Financial Studies* **30**(12): 4349–4388.

Litterman, B. (2004). *Modern investment management: an equilibrium approach*, Vol. 246, John Wiley & Sons.

Lugosi, G. and Mendelson, S. (2019a). Mean estimation and regression under heavy-tailed distributions: A survey, *Foundations of Computational Mathematics* **19**(5): 1145–1190.

Lugosi, G. and Mendelson, S. (2019b). Sub-Gaussian estimators of the mean of a random vector, *The annals of statistics* **47**(2): 783–794.

Lugosi, G. and Mendelson, S. (2021). Robust multivariate mean estimation: the optimality of trimmed mean, *The Annals of Statistics* **49**(1): 393–410.

Markowitz, H. (1952). Portfolio selection, *Journal of Finance* **7**(1): 77–91.
URL: *https://EconPapers.repec.org/RePEc:bla:jfinan:v:7:y:1952:i:1:p:77-91*

Mendelson, S. and Zhivotovskiy, N. (2020). Robust covariance estimation under L_4-L_2 norm equivalence, *Annals of Statistics* **48**(3): 1648–1664.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* **8**(4): 323–361.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal?, *Financial Analysts Journal* **45**(1): 31–42.

Nemirovsky, A. S. and Yudin, D. B. (1983). Problem complexity and method efficiency in optimization.

Ostrovskii, D. M. and Rudi, A. (2019). Affine invariant covariance estimation for heavy-tailed distributions, *Conference on Learning Theory*, PMLR, pp. 2531–2550.

Petukhina, A., Trimborn, S., Härdle, W. K. and Elendner, H. (2021). Investing with cryptocurrencies–evaluating their potential for portfolio allocation strategies, *Quantitative Finance* pp. 1–29.
URL: *https://doi.org/10.1080/14697688.2021.1880023*

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors, *Journal of the American statistical association* **97**(460): 1167–1179.

Strongin, S., Petsch, M. and Sharenow, G. (2000). Beating benchmarks, *The Journal of Portfolio Management* **26**(4): 11–27.

Szarek, S. (1976). On the best constants in the khinchin inequality, *Studia Mathematica* **2**(58): 197–208.

Xidonas, P., Steuer, R. and Hassapis, C. (2020). Robust portfolio optimization: A categorized bibliographic review, *Annals of Operations Research* **292**(1): 533–552.

Zhivotovskiy, N. (2021). Dimension-free bounds for sums of independent matrices and simple tensors via the variational principle, *arXiv preprint arXiv:2108.08198* .

Zhu, B., Jiao, J. and Steinhardt, J. (2020). Robust estimation via generalized quasi-gradients, *arXiv preprint arXiv:2005.14073* .

IRTG 1792 Discussion Paper Series 2021

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Surrogate Models for Optimization of Dynamical Systems" by Kainat Khowaja, Mykhaylo Shcherbatyy, Wolfgang Karl Härdle, January 2021.

002 "FRM Financial Risk Meter for Emerging Markets" by Souhir Ben Amor, Michael Althof, Wolfgang Karl Härdle, February 2021.

003 "K-expectiles clustering" by Bingling Wang, Yingxing Li, Wolfgang Karl Härdle, March 2021.

004 "Understanding Smart Contracts: Hype or Hope?" by Elizaveta Zinovyev, Raphael C. G. Reule, Wolfgang Karl Härdle, March 2021.

005 "CATE Meets ML: Conditional Average Treatment Effect and Machine Learning" by Daniel Jacob, March 2021.

006 "Coins with benefits: on existence, pricing kernel and risk premium of cryptocurrencies" by Cathy Yi-Hsuan Chen, Dmitri Vinogradov, April 2021.

007 "Rodeo or Ascot: which hat to wear at the crypto race?" by Konstantin Häusler, Wolfgang Karl Härdle, April 2021.

008 "Financial Risk Meter based on Expectiles" by Rui Ren, Meng-Jou Lu, Yingxing Li, Wolfgang Karl Härdle, April 2021.

009 "Von den Mühen der Ebenen und der Berge in den Wissenschaften" by Annette Vogt, April 2021.

010 "A Data-driven Explainable Case-based Reasoning Approach for Financial Risk Detection" by Wei Li, Florentina Paraschiv, Georgios Sermpinis, July 2021.

011 "Valuing cryptocurrencies: Three easy pieces" by Michael C. Burda, July 2021.

012 "Correlation scenarios and correlation stress testing" by Natalie Packham, Fabian Woebbeking, July 2021.

013 "Penalized Weigted Competing Risks Models Based on Quantile Regression" by Erqian Li, Wolfgang Karl Härdle, Xiaowen Dai, Maozai Tian, July 2021.

014 "Indices on Cryptocurrencies: an Evaluation" by Konstantin Häusler, Hongyu Xia, August 2021.

015 "High-dimensional Statistical Learning Techniques for Time-varying Limit Order Book Networks" by Shi Chen, Wolfgang Karl Härdle, Melanie Schienle, August 2021.

016 "A Time-Varying Network for Cryptocurrencies" by Li Guo, Wolfgang Karl Härdle, Yubo Tao, August 2021.

017 "Green financial development improving energy efficiency and economic growth: a study of CPEC area in COVID-19 era" by Linyun Zhang, Feiming Huang, Lu Lu, Xinwen Ni, September 2021.

018 "Robustifying Markowitz" by Wolfgang Karl Härdle, Yegor Klochkov, Alla Petukhina, Nikita Zhivotovskiy, September 2021.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.