Risk Reduction and Efficiency Increase in Large Portfolios:
Leverage and Shrinkage

Zhao Zhao, Olivier Ledoit and Hui Jiang

Revised version, January 2020

Risk Reduction and Efficiency Increase in Large Portfolios: Leverage and Shrinkage

Zhao Zhao[†], Olivier Ledoit[‡] and Hui Jiang[*]

[†]School of Economics, Huazhong University of Science and Technology,
Wuhan, Hubei, China. zhaozhao@hust.edu.cn

[‡]Department of Economics, University of Zurich, CH-8032 Zurich, Switzerland,
olivier.ledoit@econ.uzh.ch; and AlphaCrest Capital Management, New York, NY, USA.

[*]School of Mathematics and Statistics, Huazhong University of Science and Technology,
Wuhan, Hubei, China. jianghui@hust.edu.cn

Abstract

We investigate the effects of constraining leverage and shrinking covariance matrix in constructing large portfolios, both theoretically and empirically. Considering a wide variety of setups that involve conditioning or not conditioning the covariance matrix estimator on the recent past (multivariate GARCH), smaller vs. larger universe of stocks, alternative portfolio formation objectives (Global Minimum Variance vs. exposure to profitable factors), and various transaction cost assumptions, we find that a judiciously-chosen shrinkage method always outperforms an arbitrarily-determined leverage constraint. By extending the mathematical connection between leverage and shrinkage from static to dynamic, we provide a new theoretical explanation for our finding from the perspective of degrees of freedom. In addition, both simulation and empirical analysis show that the DCC-NL estimator results in risk reduction and efficiency increase in large portfolios as long as a small amount of leverage is allowed, whereas tightening the leverage constraint often hurts a DCC-NL portfolio.

KEY WORDS: DCC; Nonlinear shrinkage; Leverage constraint; Large portfolios; Risk reduction; Markowitz mean-variance efficiency.

JEL CLASSIFICATION NOS: C13, C58, G11.

---

[*]Corresponding Author.

1 Introduction

It is well established that using sample covariance matrix is inappropriate in constructing large unconstrained portfolios, which will lead to extreme positions and poor out-of-sample performance. To improve the performance of portfolios, two different methods have been proposed and become very popular. The first one is to directly constrain (e.g., DeMiguel et al., 2009a; Jagannathan and Ma, 2003; Li, 2015) or shrink (e.g., DeMiguel et al., 2013; Kan and Zhou, 2007; Tu and Zhou, 2011) the portfolio weights, and the second is to use a shrinkage estimator of covariance matrix to indirectly constrain the portfolio weights (e.g., Ledoit and Wolf, 2017).

There is a general perception that constraining the leverage is a simple alternative to shrinking the covariance matrix. Studies share the same theoretical analysis on why the two approaches are beneficial to large portfolios: the extreme sample covariances between stocks that lead to extreme weights are usually caused by estimation errors (see footnote 8 in DeMiguel et al., 2009b, for illustration); thus, imposing constraints on portfolio weights or shrinking the sample covariances can reduce the sampling error and improve the out-of-sample performance. While some people may have read this connection as a confirmation that "shrinkage is not necessary", it could equally be interpreted as saying that shrinkage works because *we* (meaning the asset management and finance professors who impose leverage constraints) have been effectively shrinking all along without even knowing it.

This paper aims at providing a comparative analysis of the two approaches and getting the best of both worlds. In particular, we work in the realm of fully-invested portfolios, i.e., portfolios whose weights sum up to one, which is the default choice for the bulk of the asset management industry (as opposed to weights summing to zero). Even though the weights sum up to one, there is some leeway to take on some negative weights, and an interesting question is how much. The major candidates are the strategy without any constraint on leverage, the so-called "150/50" strategy (meaning that for $100 million of capital, the prime broker enables you to go $150 million long if you go $50 million short at the same time, for a net exposure to the stock market of $100 million, which is exactly equal to the capital invested), the "130/30" strategy, and the "100/0" strategy (no short sales at all). For a more solid analysis, we also consider a wide choice for the degree of the leverage constraint, with the gross-exposure parameter continuously increasing from 1 (the "100/0" strategy) to 16 (the "1300/300" strategy). Moreover, at the same time, instead of linearly shrinking the unconditional covariance matrix as the previous literature, we allow each of the eigenvalues of the sample covariance matrix to have its own shrinkage intensity, optimally determined under large-dimensional asymptotics, while also incorporating Multivariate GARCH effects.

Our paper makes the following contributions to the literature that deals with risk reduction and efficiency increase in large portfolios. First, we extend the mathematical equivalency of imposing leverage constraints and shrinking covariances from the conventional static framework to dynamic, where Multivariate GARCH effects are incorporated. Second, we find that a judiciously-chosen shrinkage method on covariance matrix always outperforms the strategy

of imposing an arbitrarily-determined constraint on leverage. We provide a new theoretical explanation from the perspective of degrees of freedom to bridge the gap between theory and practice. Third, we reveal the best of both shrinkage on covariances and constraint on weights through Monte Carlo simulations and empirical exercises, which can provide enlightenments to quantitative investors and analysts.

While if imposing leverage constraints and shrinking covariances are essentially equivalent, then why is shrinkage still beneficial for portfolios with leverage constraints? The answer is that constraining the leverage has only one degree of freedom, i.e., the amount of leverage permitted; whereas shrinkage of the nonlinear kind — as utilized here — has as many degrees of freedom as there are variables in the system, which is assumed to be a large number, and these are individually optimized in an asymptotic sense as matrix dimension and sample size go to infinity together. Thus, the influence of shrinkage is fine-tuned automatically, whereas leverage applies uniformly across the board.[1] The leverage constraint, therefore, is redundant when the optimal shrinkage is provided by the nonlinear shrinkage method.

Our Monte Carlo simulation and empirical analysis demonstrate that loosening the leverage constraint from zero leverage (100/0) to 130/30 or even 150/50 delivers better performance. Moreover, we show that using a 1000-dimensional Multivariate GARCH model with built-in nonlinear shrinkage is independently and additively beneficial. In addition, we find that as long as there is *some* leverage — which can be done in many fully-invested funds as long as the prime broker allows it —, then shrinkage starts to help, and implementing the most advanced covariance matrix shrinkage formula helps even more. However, tightening the leverage constraint often hurts the out-of-sample performance of a pure DCC-NL portfolio. Our findings are stable in simulations, where both normally-distributed and t-distributed disturbance terms are considered, as well as in empirical exercises, where both the "Global Minimum Variance" (GMV) portfolio and the mean-variance efficient portfolio are constructed.

We know that since the groundbreaking work of Markowitz (1952), statistics and optimization techniques have been used to develop diversified investment strategies that either: (i) minimize risk, subject to exposure to the stock market (the GMV portfolio); or (ii) are efficient in terms of risk-return trade-off (the mean-variance efficient portfolio). The first type is a *purer* test of the covariance matrix estimator and the optimization program used, while the second one requires, in addition, a good predictive model for expected returns, which is notoriously hard to obtain. During the past decades, on the one hand, various attempts have been made to deal with the parameter uncertainty and estimation risk in optimizing portfolios (e.g., Brandt et al., 2009; Branger et al., 2019; Garlappi et al., 2007). On the other hand, hundreds of signals have been proposed to predict expected returns (see Green et al., 2017; Harvey et al., 2016; Hou et al., 2015, and the references therein). Given that the two streams of literature have attracted a large amount of attention, we consider both exercises. Specifically, we use the signal return-on-equity as a proxy for the expected return, which has been proven to have statistically significant explanatory power for cross-sectional anomalies (Feng et al., 2020). For robustness

---

[1]Behr et al. (2013) extend the arbitrarily-chosen constraints on weights to flexible *ex-ante* constraints that can better suit the data, but it is still infeasible to guarantee that all individual weights are optimally shrunk.

check, we consider the widely-used signal earnings-to-price as an alternative proxy. We also consider different setups for transaction costs.

Besides the vast literature on return predictive signals, our work is most closely related to the literature that attempts to provide a unified analysis framework for constraints on weights and advanced estimators for covariance matrix. Jagannathan and Ma (2003) find with the no-short-sale constraint in place, global minimum variance portfolios constructed based on sample covariance matrix perform as well as those constructed using the linear shrinkage estimator. However, if a shrinkage estimator of the covariance matrix is used, the no-short-sale constraint would then hurt the out-of-sample performance. They explain the similar performances and the "either-or" dichotomy by the analogous mechanisms of constraining weights and shrinking covariances. Brodie et al. (2009) and DeMiguel et al. (2009a) use the norm framework to unify the shrinkage effects in weights and in covariance matrix. Britten-Jones (1999) converts the portfolio selection problem to a regression problem. On this basis, Fan et al. (2012) and Li (2015) show that constraining portfolio norms amounts to constraining estimation risks. Further, Callot et al. (2020) use a nodewise regression approach to estimate the inverse covariance matrix directly.

Note that shrinkage estimators of covariance matrix that have been compared with constraints on portfolio weights are all linear ones, with the shrinkage target being a covariance matrix implied by the Sharpe (1963) one-factor model proposed by Ledoit and Wolf (2003) (DeMiguel et al., 2009a; Jagannathan and Ma, 2003; Li, 2015), an identity matrix proposed by Ledoit and Wolf (2004b) (DeMiguel et al., 2009a), or a constant-correlation model proposed by Ledoit and Wolf (2004a) (Li, 2015). The estimation is substantially equivalent to linearly shrinking the sample eigenvalues towards a more centralized set of eigenvalues by a unified shrinkage intensity. Ledoit and Wolf (2012) extend the linear shrinking to the nonlinear transformation of the sample eigenvalues, and obtain the nonlinear shrinkage (NL) estimator of the covariance matrix, which has been proven to have better out-of-sample performance (Ledoit and Wolf, 2015, 2017). As the shrinkage mechanism is improved from a procedure with an exogenous target and a unified intensity to an endogenous optimization algorithm, it is important to compare its effect with that of imposing varying degrees of leverage constraint.

Further, Bollerslev et al. (2018) claim that the shrinkage intensity should be time-varying to consider the dynamic variation of the covariances. Consistent with this idea, Engle et al. (2019) propose the DCC-NL estimator of the covariance matrix, which uses the nonlinear shrinkage estimator to replace the sample covariance matrix in the "correlation targeting" maximum likelihood estimation of the dynamic conditional correlation (DCC) model. The DCC-NL estimator of the covariance matrix turns out to perform better than previous estimators based on conventional DCC model. Since the DCC model works in capturing the conditional heteroscedasticity, which is totally different from the shrinkage mechanism, we conjecture that the use of DCC model would also help improve the out-of-sample performance of portfolios with leverage constraints.

Thus, a key difference between our work and the previous literature is that we focus on the most advanced estimator for covariance matrix, which not only applies the nonlinear shrinkage

procedure to account for the individual shrinkage intensity, but also considers the dynamic variation of the covariances using the GARCH model. On this basis, we perform a comparative analysis and obtain new findings for the effects of shrinkage covariance matrix and leverage constraint on risk reduction and efficiency increase.

The rest of the paper is organized as follows. Section 2 provides the methodologies and the background, including the nonlinear shrinkage estimator of the covariance matrix, its combination with the DCC model, and our theoretical findings in constructing leverage-constrained GMV and mean-variance efficient portfolios. In Section 3, we use Monte Carlo simulations to verify our theoretical results. In Section 4, we describe our data, report the results for out-of-sample performance of the GMV and the mean-variance efficient portfolios, and conduct robustness checks. Section 5 concludes.

2 Methodology and Background

2.1 Shrinkage and DCC

It is widely known that the sample covariance matrix performs poorly out-of-sample in large dimensions due to overfitting. Without imposing any additional structure on the data, shrinkage methods improve the estimation precision by rectifying the bias of the sample eigenvalues. The basic idea behind shrinkage methods is to pull the extreme sample eigenvalues towards the grand mean of all sample eigenvalues, since the smallest sample eigenvalues are biased downwards and the largest ones upwards. Ledoit and Wolf (2003, 2004a,b) propose the linear shrinkage estimators, which are the first-order approximation solutions to a nonlinear optimization problem, as all sample eigenvalues adjust with the same shrinkage intensity. The nonlinear shrinkage estimators proposed by Ledoit and Wolf (2012, 2015) allow the sample eigenvalues to adjust with heterogeneous shrinkage intensities and should generally perform better than the linear ones.

To determine the optimal shrinkage intensity for every sample eigenvalue (in regard to a particular loss function), Ledoit and Wolf (2015) discretize the famous Marčenko and Pastur (1967) equation and construct the QuEST (Quantized Eigenvalues Sampling Transform) function. By numerically inverting the QuEST function, the consistent estimators for the population eigenvalues can be obtained. Specifically, let $(\lambda_1, \ldots, \lambda_N)$ denote a set of eigenvalues of the $N \times N$ sample covariance matrix S, sorted in descending order, and $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$ be the corresponding eigenvectors. Let $Q_{T,N}(\boldsymbol{t}) := \left(q_{T,N}^1(\boldsymbol{t}), \ldots, q_{T,N}^N(\boldsymbol{t})\right)'$ denote the QuEST function, which turns the set of population eigenvalues $\boldsymbol{t} := (t_1, \ldots, t_N)$ into the set of sample eigenvalues. Thus, given the set of sample eigenvalues, the population eigenvalues can be consistently estimated by inverting the QuEST function:

$$\widehat{\boldsymbol{\tau}} := \underset{\mathbf{t} \in [0, +\infty)^N}{\operatorname{argmin}} \ \frac{1}{N} \sum_{i=1}^{N} \left(q_{T,N}^i(\mathbf{t}) - \lambda_i\right)^2 \ , \tag{2.1}$$

Then, the nonlinear shrinkage estimator (denoted by NL) of the covariance matrix is

$$\widehat{\Sigma} := \sum_{i=1}^{N} \widehat{\lambda}_i\left(\widehat{\boldsymbol{\tau}}\right) \cdot \boldsymbol{u}_i \boldsymbol{u}_i' \,, \tag{2.2}$$

where $\widehat{\lambda}_i\left(\widehat{\boldsymbol{\tau}}\right)$ for $i = 1, \ldots, N$ denote the shrunk eigenvalues based on $\widehat{\boldsymbol{\tau}}$. The basic idea of this shrinkage formula is that

$$\widehat{\lambda}_i\left(\widehat{\boldsymbol{\tau}}\right) \approx \boldsymbol{u}_i' \, \Sigma \boldsymbol{u}_i \,, \tag{2.3}$$

where Σ represents the unconditional population covariance matrix. The approximation is valid asymptotically as matrix dimension and sample size go to infinity together in the manner detailed by Ledoit and Wolf (2015). Equation (2.2)–(2.3) are very similar to $\lambda_i = \boldsymbol{u}_i' \, \Sigma \boldsymbol{u}_i$ and $S = \sum_{i=1}^{N} \lambda_i \cdot \boldsymbol{u}_i \boldsymbol{u}_i'$: all we have done is replace the *in-sample* variance of a portfolio whose weights are determined by eigenvector $\boldsymbol{u}_i$ with the *true* variance of the same portfolio. This is a substantial improvement because the fact that the eigenvectors $(\boldsymbol{u}_i)_{i=1,\ldots,N}$ are extracted from the same dataset as the eigenvalues $(\lambda_i)_{i=1,\ldots,N}$ generates tremendous over-fitting bias.

In addition, to capture the volatility-clustering feature of asset returns, Engle (2002) uses the Dynamic Conditional Correlation (DCC) model to describe the time-varying structure in variances and covariances. Let $\Sigma_t := (\sigma_{ijt})$ denote the conditional covariance matrix of asset returns $r_t := (r_{it})$ (N-dimensional column vector) at time t, where $t = 1, \ldots, T$. Let $D_t := \mathrm{diag}\left(\sigma_{11t}^{1/2} \ldots \sigma_{NNt}^{1/2}\right)$ denote the volatility matrix, $Q_t := (q_{ijt})$ the pseudo-correlation matrix, and $P_t := (\rho_{ijt})$ the correlation matrix, satisfying

$$P_t := \mathrm{diag}\left(q_{11t}^{-1/2} \ldots q_{NNt}^{-1/2}\right) Q_t \, \mathrm{diag}\left(q_{11t}^{-1/2} \ldots q_{NNt}^{-1/2}\right) \,. \tag{2.4}$$

The DCC model is defined as

$$\Sigma_t = D_t P_t D_t \,. \tag{2.5}$$

A GARCH(1,1) model is used to describe the dynamic of every univariate volatility:

$$\sigma_{ii,t}^2 = \sigma_{ii,0}^2 \left(1 - \alpha_i - \beta_i\right) + \alpha_i \, r_{i,t-1}^2 + \beta_i \, \sigma_{ii,t-1}^2 \,, \tag{2.6}$$

and the pseudo-correlation matrix Q_t is specified as

$$Q_t = \bar{Q}(1 - \alpha - \beta) + \alpha \, s_{t-1} s_{t-1}' + \beta \, Q_{t-1} \,, \tag{2.7}$$

where α_i, β_i, α, and β are non-negative scalars satisfying $\alpha_i + \beta_i < 1$ for every $i \in \{1, 2, \ldots, N\}$ and $\alpha + \beta < 1$. $\sigma_{ii,0}$ is the long-run volatility of asset return for individual i, $s_t = D_t^{-1} r_t$ is the devolatilized returns at time t, and $\bar{Q}$ is the long-run covariance matrix of s_t.

By combining the nonlinear shrinkage estimator of $\bar{Q}$ with the DCC model, Engle et al. (2019) propose the DCC-NL estimator of the covariance matrix. To avoid inverting matrices with large dimensions, they also use the 2MSCLE method (Pakel et al., 2020) in estimating the DCC model, which is the composite likelihood estimation bonding the individual likelihoods generated by 2×2 blocks of all contiguous pairs. To sum up, NL aims to improve the estimation precision of covariance matrix by shrinking eigenvalues and thus reducing sampling errors. Meanwhile,

DCC takes the conditional heteroscadasticity into consideration by dynamic modeling. In view of these strengths, the DCC-NL estimator is supposed to have better out-of-sample performance than the DCC estimator, the NL estimator, and the sample covariance matrix (denoted by S), especially in large dimensions.

2.2 Constructing GMV Portfolios with Leverage Constraint

Based on the estimator $\widehat{\Sigma}_t$ of the time-varying covariance matrix Σ_t, constructing GMV portfolio with gross-exposure constraints is equivalent to the following minimization problem given by

$$\min_{\boldsymbol{w}_t} \boldsymbol{w}_t' \widehat{\Sigma}_t \boldsymbol{w}_t \tag{2.8}$$

$$\text{subject to} \qquad \boldsymbol{w}_t' \mathbf{1} = 1 \text{ and } \sum_{i=1}^{N} |w_{i,t}| \leq \gamma.$$

The constraint $\sum_{i=1}^{N} |w_{i,t}| \leq \gamma$ could be expressed as $\|\boldsymbol{w}_t\|_1 \leq \gamma$. Note that $\gamma \geq 1$, and the constraint becomes weaker with the increase of γ. When $\gamma = 1$, the constraint is equivalent to the extreme situation considered in Jagannathan and Ma (2003) that no short sales are allowed. $\gamma = 1.6$ corresponds to fully-invested portfolios of the 130/30 type, and $\gamma = 2$ to 150/50. When $\gamma = \infty$, short exposure is unconstrained.

Define the Lagrangian as

$$L(\boldsymbol{w}_t, \mu, \lambda) \quad = \quad \boldsymbol{w}_t' \widehat{\Sigma}_t \boldsymbol{w}_t - \mu(\boldsymbol{w}_t' \mathbf{1} - 1) - \lambda(\gamma - \|\boldsymbol{w}_t\|_1), \tag{2.9}$$

and let $\mathbf{g}_t$ be the subgradient vector of $\|\boldsymbol{w}_t\|_1$. Then, for $w_{i,t} \neq 0$, the i-th element of $\mathbf{g}_t$ is unique, i.e., $g_{i,t} = sign(w_{i,t})$; for $w_{i,t} = 0$, $g_{i,t}$ could be any values in $[-1, 1]$.

Consequently, the Karush-Kuhn-Tucker (KKT) conditions for the above leverage-constrained optimization problem (2.8) are

$$\begin{cases} 2\widehat{\Sigma}_t \boldsymbol{w}_t - \mu \mathbf{1} + \lambda \mathbf{g}_t = 0, \\ \lambda(\gamma - \|\boldsymbol{w}_t\|_1) = 0, \quad \lambda \geq 0, \\ \|\boldsymbol{w}_t\|_1 \leq \gamma, \quad \boldsymbol{w}_t' \mathbf{1} - 1 = 0, \end{cases} \tag{2.10}$$

where $\mathbf{1}$ is the column vector of ones, λ and μ are Lagrange multipliers. Denote a solution to (2.10) as $\boldsymbol{w}_t^*$. The following result shows that constructing the leverage-constrained minimum variance portfolio from the DCC estimator $\widehat{\Sigma}_t$ is equivalent to constructing a (unconstrained) minimum variance portfolio from a shrunk version of $\widehat{\Sigma}_t$.

Theorem 1. (i) *Let $\tilde{\Sigma}_{\gamma,t} = \widehat{\Sigma}_t + \frac{1}{2}\lambda(\mathbf{g}_t^* \mathbf{1}' + \mathbf{1}\mathbf{g}_t^{*'})$, where $\mathbf{g}_t^*$ is the subgradient at $\boldsymbol{w}_t^*$, and λ is the Lagrange multiplier defined in (2.10). Then $\tilde{\Sigma}_{\gamma,t}$ is positive definite if $\widehat{\Sigma}_{\gamma,t}$ is a positive definite DCC covariance matrix estimator.*

(ii) *The partial constrained portfolio optimization problem (2.8) is equivalent to the optimization problem*

$$\min_{\boldsymbol{w}_t' \mathbf{1} = 1} \boldsymbol{w}_t' \tilde{\Sigma}_{\gamma,t} \boldsymbol{w}_t \tag{2.11}$$

with the regularized covariance matrix $\tilde{\Sigma}_{\gamma,t}$.

It can easily be seen that the approach of imposing constraints on leverage has only one degree of freedom: the Lagrange multiplier λ (or, equivalently, the gross exposure constraint γ, as these two are in one-to-one correspondance, holding everything else equal). It is not obvious how to choose this parameter optimally if the goal is to maximize covariance matrix accuracy, but we will leave this issue aside for a moment. By contrast, nonlinear shrinkage has N degrees of freedom $\widehat{\lambda_1}(\widehat{\tau}), \ldots, \widehat{\lambda_N}(\widehat{\tau})$, each one chosen optimally through an automatic procedure under large-dimensional asymptotics. This ability to 'locally fine-tune' is a huge advantage when population eigenvalues can be dispersed, clustered, or otherwise unruly, which is the general case.

2.3 Constructing Mean-Variance Portfolios with Leverage Constraint

Given the estimator of the covariance matrix $\widehat{\Sigma}_t$ and the gross-exposure parameter γ, the Markowitz mean-variance efficient portfolio based on a return predictive signal $\boldsymbol{m}_t := (m_{1t}, \ldots, m_{Nt})'$ is formulated as:

$$\min_{\boldsymbol{w}_t} \boldsymbol{w}_t' \widehat{\Sigma}_t \boldsymbol{w}_t \tag{2.12}$$

$$\text{subject to} \quad \boldsymbol{w}_t' \mathbf{1} = 1 , \tag{2.13}$$

$$\boldsymbol{w}_t' \boldsymbol{m}_t = b_t \ \text{ and} \tag{2.14}$$

$$\sum_{i=1}^{N} |w_{i,t}| \leq \gamma, \tag{2.15}$$

where b_t is a selected target exposure to the signal $\boldsymbol{m}_t$. In our empirical study, b_t is determined by the sorting portfolios. In particular,

$$b_t = \boldsymbol{w}Q_t' \boldsymbol{m}_t \tag{2.16}$$

where $\boldsymbol{w}Q_t$ is the weight vector of quantile-based portfolios. Let $\{(1), (2), \ldots, (N)\}$ be the permutation of $\{1, 2, \ldots, N\}$ that results in descending order of scores for the signal $\boldsymbol{m}_t$. Then, $wQ_t^{(1)} = \ldots = wQ_t^{(d)} := 1/d$ and $wQ_t^{(d+1)} = \ldots = wQ_t^{(N)} := 0$, where d is the largest integer that is smaller than or equal to the ratio of portfolio size N to the number of quantiles B. We consider quintiles ($B = 5$) in our empirical analysis.

Denote the solution to problem (2.12) as $\boldsymbol{w}_{b,t}^*$, then we could obtain the following theorem similar to Theorem 1.

Theorem 2. *The partial constrained portfolio optimization problem (2.12) is equivalent to the optimization problem*

$$\min_{\boldsymbol{w}_t} \boldsymbol{w}_t' \tilde{\Sigma}_{\gamma,t} \boldsymbol{w}_t \tag{2.17}$$

$$\text{subject to} \quad \boldsymbol{w}_t' \mathbf{1} = 1 \ \text{and} \ \boldsymbol{w}_t' \boldsymbol{m}_t = b_t.$$

with the regularized covariance matrix $\tilde{\Sigma}_{\gamma,t}$. Here $\tilde{\Sigma}_{\gamma,t} = \widehat{\Sigma}_t + \frac{1}{2}\lambda(\mathbf{g}_{b,t}^* \mathbf{1}' + \mathbf{1}\mathbf{g}_{b,t}^{*\prime})$, $\mathbf{g}_{b,t}^*$ is the subgradient at $\boldsymbol{w}_{b,t}^*$, and λ is the Lagrange multiplier.

3 Monte Carlo Simulations

3.1 Data Generating Process

In the last section, we demonstrate that imposing the gross-exposure constraint on portfolio weights is equivalent to using the shrinkage estimator when the time-varying structure of the covariance matrix is captured by the DCC model. Moreover, we consider varying levels of constraint on the gross-exposure instead of imposing the nonnegative constraints on all weights.

In this section, we use Monte Carlo simulations to verify our theoretical results and to quantify the finite sample performance with varying levels of leverage constraint and different covariance matrix estimators. We generate Monte Carlo data that can simulate the real data the best.

First, we estimate the unconditional covariance matrix from the most liquid stocks ($N = 500, 1000$) in the CRSP database based on the nonlinear shrinkage method using five years of daily data from 2010 to 2014. This matrix will be regarded as the true unconditional covariance matrix.

Second, we simulate the DCC time series with parameters $\alpha = 0.05$ and $\beta = 0.93$ for (2.7), and with parameters $\alpha_i = 0.05$ and $\beta_i = 0.90$ for all individual stocks $i = 1, \ldots, N$ for (2.6). The disturbance terms are drawn from a multivariate standard normal distribution or a multivariate Student t-distribution with 5 degrees of freedom. For each simulation, we thereby generate an $T \times N \times N$ time-varying covariance matrix and correspondingly an $T \times N$ matrix of simulated returns, where the time length T is fixed at 1250. We repeat each simulation for 100 times.

3.2 Simulation Results

We consider four different estimators for the covariance matrix, including the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) proposed by Ledoit and Wolf (2012, 2015), the covariance matrix estimator based on DCC model (Engle, 2002), and the DCC-NL estimator proposed by Engle et al. (2019). For each covariance matrix estimator, we allow a wide choice of the leverage-constraint parameter γ, ranging from 1 to 16. Here we construct GMV portfolios based on (2.8).

Following Fan et al. (2012), we calculate three different risks to evaluate the performance. First, the oracle risk (denoted by R_{orc}) corresponding to the oracle portfolio, based on the true time-varying covariance matrix, is obtained by (3.1). Second, the empirical risk (denoted by R_{emp}) corresponding to the empirical portfolio, based on the estimated time-varying covariance matrix, is obtained by (3.2). The empirical risk, however, is not the actual risk of the empirical portfolio, since it relies on the estimated covariance matrix. Replacing the estimated covariance matrix by the true covariance matrix, we get the actual risk (denoted by R_{act}, (3.3)), which is crucial for comparing the finite sample performance of different portfolios.

$$
\begin{align}
R_{orc} &= \boldsymbol{w}_t' \Sigma_t \boldsymbol{w}_t \tag{3.1} \\
R_{emp} &= \widehat{\boldsymbol{w}}_t' \widehat{\Sigma}_t \widehat{\boldsymbol{w}}_t \tag{3.2}
\end{align}
$$

$$R_{act} \quad = \quad \widehat{\boldsymbol{w}}_t' \Sigma_t \widehat{\boldsymbol{w}}_t \qquad\qquad (3.3)$$

Figures 1-4 show R_{orc}, R_{emp}, and R_{act} for four different covariance matrix estimators, each with varying constraints on the gross-exposure, where Figure 1 and Figure 3 are for the cases of $N = 500$, and Figure 2 and Figure 4 are for the cases of $N = 1000$. For Figures 1-2, the disturbance terms of the simulated data are drawn from a multivariate standard normal distribution, while for Figures 3-4, they are drawn from a multivariate "Student" t-distribution with 5 degrees of freedom. In all cases, the sample size is fixed at $T = 1250$. It is intuitive to see from Equations (3.1) to (3.3) that the gap between the actual and the oracle risks is caused by the estimation errors in weights or the wrong constraints, whereas the differences between the actual and the empirical risks are because of the estimation errors in the covariance matrix estimators.

The curve of the oracle risk shared by the four graphs in each figure indicates that the theoretical risk decreases sharply with the increase of the gross-exposure parameter γ before γ reaches 2, when the constraints on gross-exposure form the 150/50 strategy. This provides theoretical evidence supporting the outstanding performance of the 150/50 strategy. The sums of the absolute weights for oracle portfolios are 5.76 (5.75) and 5.85 (5.84), for $N = 500$ and $N = 1000$, respectively, in the cases with normally-distributed (t-distributed) disturbance terms, implying the corresponding exact inflection points on the curves of oracle risk.

The estimation of the covariance matrix becomes more and more difficult with the increase of γ and N, especially if no-shrinkage-involved covariance matrix estimator is used at the meantime. When $N = 500$, the gap between the actual and the empirical risks first becomes wider as γ increases, and then remains constant. In addition, the actual-empirical gap is noticeably larger for no-shrinkage-involved covariance matrix estimators. The difference between the actual and the oracle risks show very similar properties, except that it is much smaller when DCC model is used, suggesting the effectiveness of DCC model in forecasting and accordingly in reducing the actual risk. In the cases of $N = 1000$, the actual risks are remarkably larger when the constraints on the gross-exposure are weak and S or DCC estimator is used at the same time: the curves of actual risks turn up when $\gamma > 4$, making the actual-empirical gap and the actual-oracle gap increase. These findings all help confirm the positive effects of the shrinkage and the leverage constraint on reducing risks. However, in comparison with the consistent positive effects of the shrinkage, leverage constraint helps reduce portfolio risks only in moderate degree.

To better analyze the actual risks of portfolios constructed based on different covariance matrix estimators, we show them in Figure 5 (for normally-distributed disturbance terms) and Figure 6 (for t-distributed disturbance terms). On the one hand, for varying leverage constraints, it is always the DCC-NL estimator that brings the lowest risk. On the other hand, with the relaxation of the leverage constraint (the increase of γ), the actual risk first decreases and then increases (for S and DCC) or remains constant (for NL and DCC-NL). The actual risk is the lowest when γ is between 2 and 5. When $\gamma < 5$ (the leverage constraint is effective), the performance of estimators mainly depends on whether the DCC is used, whereas when $\gamma > 5$, the difference in performance largely depends on whether the shrinkage is used. Overall,

the DCC model, the shrinkage method, and the imposing of an appropriate leverage constraint all help reduce risks, and the effects increase with N.

We also present a subset of the simulation results in Table 1 (for normally-distributed disturbance terms) and Table 2 (for t-distributed disturbance terms). Specifically, the tables present the performance of the GMV portfolios based on various covariance matrix estimators with gross-exposure parameter $\gamma = \infty$, 2, 1.6, and 1, corresponding to four specific strategies, which are no constraint on weights, the 150/50 strategy, the 130/30 strategy, and the no-short-sale constraint. Results shown in Panels A, B are for portfolios with $N = 500$ and $N = 1000$, respectively.

Besides the aforementioned revelations, Tables 1-2 indicate that portfolios constructed using the DCC-NL estimator without imposing any leverage constraint have the minimum actual risks. This should not be surprising since the data-generating process is based on DCC model and the ex post standard deviation calculated is based on the real covariance matrix. The moderate leverage constraint with parameter $\gamma = 2$ or $\gamma = 1.6$ helps reduce the actual risks remarkably when no shrinkage is used in the estimation of the covariance matrix, but they are not as effective as using the DCC-NL estimator. Moreover, in line with the results shown in Figure 5 and Figure 6, the DCC-NL estimator is always suggested, even when an appropriate leverage constraint is imposed.

The numbers in the last two columns indicate that both the shrinkage and the leverage constraint largely reduce the standard deviations and the total short positions of weights. The oracle total short position, based on the true covariance matrix, approximately equals 237.9% (237.6%) when $N = 500$, and equals 242.3% (242.1%) when $N = 1000$, for normally-distributed (t-distributed) disturbance terms. As a result, the 130/30 strategy and 150/50 strategy are not diversified enough, leading to their underperformance compared to using the DCC-NL estimator without imposing any leverage constraint.

3.3   Combining Shrinkage with Leverage Constraint

As a final lesson, there are circumstances where the leverage constraint is externally imposed: for example, by the regulatory authorities, by the financing conditions extended by prime brokers, or by risk-management commitments advertized to fund investors in the marketing materials at the initial asset-gathering stage. In such cases, Tables 1–2 show that there is still incremental benefit to using DCC-NL, even *after* the leverage constraint has already been imposed. Indeed, for every panel and every value of the gross exposure constraint γ:

1. the conditional covariance matrix (DCC) is better than the unconditional one (S);
2. the shrunk conditional covariance matrix (DCC-NL) is better than the plain one (DCC).

Here we measure 'better' as having lower standard deviation of the returns on the Global Minimum Variance portfolio. Note, however, that the gains from shrinkage become monotonically weaker as the gross exposure constraint γ becomes more binding. This is because the leverage constraint applies a 'brute-force' one-size-fits-all *over*-shrinkage that leaves little

room for the benefits of a locally adaptive optimal nonlinear shrinkage formula to express themselves.

4 Empirical Results

4.1 Data

We examine the effects of the leverage constraint, the shrinkage estimation in covariance matrix, and the use of DCC model to capture the heteroscedasticity on the out-of-sample performance of the minimum variance portfolio and the mean-variance efficient portfolio. We use the same portfolio-construction rules as in Ledoit et al. (2019), but we impose gross-exposure constraints on portfolio weights.

Specifically, we focus on stocks traded on the NYSE, AMEX, and NASDAQ, with daily return data for all the immediately preceding 1250 days as well as the upcoming 21 days, and with correlations not exceeding 0.95[2]. The daily return data we use, which covers the period from 01/01/1980 to 12/31/2018, are from the Center for Research in Security Prices (CRSP) database. The out-of-sample period is from 01/08/1986 to 12/31/2018. We update the portfolios every 21 consecutive trading days, and thus form 396 investment dates from 01/08/1986 to 12/31/2018. At every investment date h, the covariance matrix is estimated based on the most recent 1250 daily returns (roughly equals five years).

For both the minimum variance portfolio and the efficient mean-variance portfolio, we consider two different portfolio sizes $N = 500, 1000$. For a given combination (h, N), in the set of stocks that satisfy the above conditions, we pick the largest N stocks (as measured by their market capitalization on investment date h) as our investment universe.

The covariance matrix estimators we consider here include the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2015), the covariance matrix estimator based on DCC model (Engle, 2002), and the DCC-NL estimator (Engle et al., 2019). To consider the effects of the leverage constraint and the shrinkage covariance matrix estimator together, we establish portfolios with a varying gross-exposure parameter γ, based on each covariance matrix estimator for each portfolio type and portfolio size.

4.2 Main Results

4.2.1 Results for GMV portfolios

Table 3 presents the out-of-sample performance measures of the GMV portfolios with a varying gross-exposure parameter $\gamma = \infty$, 2, 1.6, and 1 for each covariance matrix estimator. Specifically, we report the annualized average return (AvR), computed by the average out-of-sample returns multiply by 252, annualized standard deviations (StdR), computed by the standard deviation of the out-of-sample returns multiply by $\sqrt{252}$, and Information Ratios (IR), which is the ratio of AvR to StdR.

---

[2]The sample correlations are calculated based on the daily returns over the past 1250 days. We remove the stock with the lower volume in a pair on the investment date if the correlation of the two exceeds 0.95.

On the one hand, DCC-NL performs the best among four covariance matrix estimators considered in all cases with different leverage constraints. Judging by the StdR of GMV portfolios, the outperformance of DCC-NL is most obvious when $N = 1000$ and no leverage constraint is imposed: it reduces the out-of-sample standard deviation by 5.38 percentage points compared to the sample covariance matrix.

As the leverage constraint becomes tighter, the relative outperformance of NL declines, while that of the DCC improves. The decline in the outperformance of NL is consistent with the observations in Jagannathan and Ma (2003). According to them, imposing constraints on portfolio weights has a shrinkage-like effect, and thus it hurts the performance of the shrinkage estimator of covariance matrix. Nevertheless, the DCC-NL estimator is always preferred: it delivers the smallest out-of-sample standard deviation even when an appropriate leverage constraint is imposed. We use the prewhitened HAC_{PW} method described in Ledoit and Wolf (2011) to test if the outperformance of DCC-NL over DCC in terms of out-of-sample standard deviation is significant in cases with different leverage constraints. The results show that the outperformance is always significant at the 0.01 level, except when no leverage is allowed.

On the other hand, the moderate constraints with $\gamma = 2$, or $\gamma = 1.6$ outperform the extreme no-short-sale constraint with $\gamma = 1$ and the no constraint with $\gamma = \infty$ if no shrinkage is used in the covariance matrix. For example, when $N = 1000$, if sample covariance matrix is used, the 50% short-sale constraint ($\gamma = 2$) reduces the out-of-sample standard deviation by 4.27 and 1.67 percentage points compared to the no-constraint strategy and the no-short-sale strategy, respectively. However, the effect of imposing a leverage constraint is limited and becomes insiginificant if the nonlinear shrinkage estimator is used.

As expected, the effect of using DCC-NL estimator is better than the combined effect of using DCC and imposing an appropriate leverage constraint. Specifically, when $N = 1000$, the GMV portfolios constructed using the DCC-NL estimator without any leverage constraint and the GMV portfolios constructed using the DCC estimator with the constraint of $\gamma = 1.6$ have annualized out-of-sample standard deviations of 8.16 percent and 8.35 percent, respectively, both reducing the 13.54 percent from using the sample covariance matrix by over 5 percentage points.

Figure 7 reveals the change of the out-of-sample risks with the continuous increase in the parameter γ. The out-of-sample risks first decline and then increase for all portfolios constructed based on different covariance matrix estimators and for both cases with 500 and 1000 stocks. The optimal choice that corresponds to the lowest risk is around $\gamma = 2$ (the 150/50 strategy) for portfolios with 500 stocks and around $\gamma = 1.6$ (the 130/30 strategy) for portfolios with 1000 stocks, where in both cases the DCC-NL estimator is suggested.

Taken together, the DCC-NL estimator achieves the best performance in all cases no matter whether a leverage constraint is imposed; imposing the 30% to 50% constraint on leverage also improves portfolio performance, but not as effective as using DCC-NL estimator.

4.2.2 Results for Mean-Variance Efficient portfolios

Table 4 presents results for the Markowitz mean-variance efficient portfolios constructed based on the signal return-on-equity (ROE), which is a well-known profitability factor that can indicate the growth potential of a firm (Haugen and Baker, 1996) and has been proven to have statistically significant explanatory power for cross-sectional anomalies (Feng et al., 2020; Hou et al., 2015). We calculate ROE by the income before extraordinary divided by 1-quarter-lagged book equity. Judging by the Information Ratios and the corresponding significant test[3], the DCC-NL estimator still performs the best among all covariance matrix estimators and its outperformance over DCC is statistically significant when the leverage constraint is not too strong. When $N = 1000$ and no constraint imposed on weights, the Markowitz portfolio constructed using the DCC-NL estimator has a Information Ratio of 1.79, almost double that from using the sample covariance matrix. When no shrinkage is in the covariance matrix, the 130/30 and 150/50 strategies outperform the strategy without any constraint on weights or with the extreme no-short-sale constraint.

Moreover, it is not difficult to find that the effect of directly using the nonlinear shrinkage estimator exceeds that of imposing the 30% or 50% leverage constraint, both of which outperform imposing the nonnegative constraint on weights, while both are inferior to using the DCC-NL estimator without any constraint. These findings again support our conjecture that imposing constraints on portfolio weights has a shrinkage-like effect, no matter whether the DCC model is used for considering the dynamics in covariances and variances. But unlike the nonlinear shrinkage technique, it is difficult to achieve the optimal shrinkage level for the leverage constraint with one degree of freedom.

4.3 Portfolio Weights

For each investment period, we compute the minimum weight (MinW), the maximum weight (MaxW), the standard deviation of weights (StdW) and the total short positions in weights (ShortW) across the N stocks of the portfolio. We present the average values over the 396 investment dates from 01/08/1986 to 12/31/2018 for the four characteristics of portfolio weights in Tables 3-4.

We find that portfolios constructed based on the sample covariance matrix have the smallest minimum weight and the largest total short position, while portfolios constructed based on the DCC estimator have the largest maximum weight, and portfolios constructed based on the NL estimator have the smallest maximum weight and the smallest standard deviation in weights.

If no leverage constraint is in place, the total short positions are always large, especially when N is large and no shrinkage is used in the estimation of the covariance matrix. For example, when $N = 1000$, the total short position of GMV portfolio reaches 337.03 percent if the DCC estimator is used, and this number even comes up to 558.05 if the sample covariance matrix is used. For Markowitz portfolios, the corresponding short positions are even larger,

---

[3]We use the prewhitened HAC_{PW} method described in Ledoit and Wolf (2008) to test if the outperformance of DCC-NL over DCC in terms of out-of-sample Information Ratio is significant.

with 361.70 percent for the DCC estimator and 571.36 percent for the sample covariance matrix.

The total short position and the turnover of portfolios are largely reduced by the use of the nonlinear shrinkage estimators and the introduction of leverage constraints. This should not be surprising considering how the nonlinear shrinkage method works in improving the estimation precision of covariance matrix.

4.4   Robustness Checks

4.4.1   Alternative Predictors

For robustness check, we consider the Markowitz mean-variance efficient portfolios with an alternative signal. Instead of using the signal ROE, now we focus on a factor that can not only indicate growth potential but also reflect price level. We follow Basu (1983) and use earnings-to-price (E/P), measured as income before extraordinary divided by the market capitalization.

Table 5 presents the results for the Markowitz mean-variance efficient portfolios constructed based on the E/P signal, from which we draw similar conclusions to our main findings aforementioned. First, the DCC-NL estimator still performs the best among all estimators we considered, though its advantage over DCC is significant only when no constraint is imposed on weights. Second, the moderate constraints with $\gamma = 2$ and $\gamma = 1.6$ lead to better out-of-sample performance than no constraint or the extreme no-short-sale constraint when no shrinkage is in the covariance matrix. Third, the effect of using NL estimator is more remarkable than that of imposing the 30% or 50% leverage constraint. When $N = 1000$, the use of NL estimator and the imposing of the 50% leverage constraint raise the Information Ratio of the portfolio constructed from 1.03 to 1.74 and 1.62, respectively. Finally, the effect of using nonlinear shrinkage estimators gradually fades out as the intensity of the leverage constraint increases. The advantage of using nonlinear shrinkage estimators culminates when no constraint is imposed: the NL estimator and the DCC-NL estimator increase the Information Ratio based on sample covariance matrix by more than 70 percent and 90 percent, respectively.

4.4.2   Transaction Cost

Transaction cost is an important issue in practical implementations (Mei and Nogales, 2018; Mei et al., 2016). In Table 6, we present results for the Markowitz mean-variance efficient portfolios constructed based on the ROE signal, when the transaction cost is considered. Referring to Avramovic and Mackintosh (2013) and Webster et al. (2015), we set the bid-ask spread to be three or five basis-points to embody the transaction cost.

Unsurprisingly, the Information Ratio becomes lower with the increase of transaction costs. Features suggested by the pattern of Information Ratios are consistent with our main results. The DCC-NL estimator generates the best out-of-sample performance among the four covariance matrix estimators in all cases. The advantage of the DCC-NL estimator is most remarkable when no leverage constraint is imposed. When $N = 500$ and the bid-ask spread is 3 basis-points, using DCC-NL estimator increases the Information Ratio of using S estimator from 0.94 to 1.23,

if no constraint is imposed on weights. When constraints are imposed, the DCC-NL estimator still helps increase the Information Ratio. Moreover, owing to the parsimony of the shrinkage method in turnover, the outperformance of the DCC-NL (NL) estimator over the DCC (S) estimator becomes more significant and robust than when the transaction cost is ignored.

4.5 Combining Shrinkage with Leverage Constraint in Practice

In the process of researching the comparison of leverage vs. shrinkage, it so happens that we have also gathered evidence as to whether there are any benefits from combining both techniques. The question is: benefits to whom?

With respect to a leverage-constrained portfolio, upgrading from the sample covariance matrix to the DCC-NL estimator (while preserving the leverage constraint) yields benefits almost as across-the-board as in the Monte-Carlo simulations analyzed in Section 3.3. The pattern identified earlier still holds: shrinkage has more room for improvement if leverage is less binding of a constraint.

With respect to a pure DCC-NL portfolio, gradually tightening the leverage constraint often hurts, but not always, and may even help at the beginning. Thus, there are circumstances where imposing $\gamma = 2$, which corresponds to a 150/50 portfolio, actually results in better performance:

- Global Minimum Variance Portfolio, 500-stock universe
- Global Minimum Variance Portfolio, 1,000-stock universe
- ROE-optimized portfolio, 5bp transaction cost, 1,000-stock universe

In most other cases, moving from an absence of leverage constraint to a 150/50 portfolio generated very little loss. Thus, taking into account the practicalities of the problem, it seems that 150/50 is the 'sweet spot' where leverage does not start to hurt (much), provided we have a good nonlinear shrinkage estimator of the conditional covariance matrix such as DCC-NL.

5 Conclusion

Leverage constraints are often used by quantitative investors. Besides the nonnegative constraint, strategies limiting the total short position to be at most 30% and 50% of the portfolio value are implementable in practice: the so-called 130/30 strategy and 150/50 strategy. Previous research find that imposing the nonnegative constraint on weights can reduce the risks of the optimal portfolios constructed and explain it by the shrinkage-like effect. We extend the nonnegative constraint to different degrees of leverage constraints so that the prevalent 130/30 strategy and 150/50 strategy are also contained. On the other hand, we focus on the DCC-NL estimator to consider the dynamics and the estimation precision of covariance matrix at the same time, where NL represents the nonlinear shrinkage, which is an improvement to the linear shrinkage estimator considered in previous literature.

We provide mathematical connection between imposing the gross-exposure constraint and using the shrinkage covariance matrix estimator in a dynamic framework. Moreover, we give

theoretical insights into the finding that using the nonlinear shrinkage estimator outperforms imposing an arbitrarily-chosen leverage constraint. We demonstrate through Monte Carlo simulations and empirical studies that portfolios constructed based on the DCC-NL estimator are well diversified, even when leverage constraints are imposed. The 30% to 50% leverage constraint is better than the no-short-sale constraint and the no constraint when no shrinkage is used in the covariance matrix estimation. The good out-of-sample performance of the DCC-NL estimator is attributed to both the use of a DCC model, which captures the dynamic structure in variances and covariances, and the introduction of an appropriate shrinkage, which reduces the sampling errors in the estimation of covariance matrix.

Based on daily return data from stocks traded on the NYSE, AMEX, and NASDAQ, we construct the global minimum variance portfolios and the Markowitz mean-variance efficient portfolios based on return predictive signals with portfolio sizes $N = 500$ and 1000. The empirical results show that though imposing an appropriate leverage constraint has a similar effect as using the nonlinear shrinkage covariance matrix estimator in reducing risks, the latter always perform better, and using the DCC-NL estimator helps even more. The effects of DCC and NL both increase in portfolio size, and the latter decreases as the leverage constraint becomes tighter. Moreover, we find that using the DCC-NL estimator improves the out-of-sample performance even when a moderate leverage constraint is imposed, but the leverage constraint often hurts a pure DCC-NL portfolio. Finally, using the NL estimator, the DCC model, and an appropriate leverage constraint all help improve the out-of-sample performance, but only NL and leverage constraint help reduce the standard deviation of weights and the turnover of portfolios.

In our main study, we use the return-on-equity as a proxy for the mean return to construct the mean-variance efficient portfolio. For robustness check, we consider the earnings-to-price as an alternative proxy and also take the effect of transaction costs into consideration by assuming the bid-ask spread to be 3 or 5 basis-points. Neither of these changes affects the robustness of our findings.

Table 1: Actual Risk of Portfolio and Standard Deviation and Total Short Position of Weights.

	$\widehat{\Sigma}$	Panel A: $N = 500$			Panel B: $N = 1000$		
		StdR	StdW	ShortW	StdR	StdW	ShortW
$\gamma = \infty$	S	5.53	1.45	222.75	6.63	0.75	239.63
	NL	5.01	1.13	166.32	3.92	0.54	161.28
	DCC	5.11	1.31	195.77	6.43	0.68	209.17
	DCC-NL	4.40	1.11	155.17	3.69	0.54	149.89
$\gamma = 2$	S	6.75	0.81	47.88	5.88	0.43	48.76
	NL	6.74	0.77	47.84	5.85	0.41	48.74
	DCC	5.14	0.73	45.90	4.63	0.40	47.74
	DCC-NL	5.11	0.73	45.86	4.57	0.40	47.71
$\gamma = 1.6$	S	7.48	0.78	29.31	6.86	0.43	29.57
	NL	7.50	0.75	29.29	6.80	0.41	29.57
	DCC	5.62	0.69	28.33	5.13	0.39	29.16
	DCC-NL	5.59	0.69	28.30	5.11	0.39	29.13
$\gamma = 1$	S	9.66	0.78	0.00	9.76	0.45	0.00
	NL	9.64	0.75	0.00	9.80	0.44	0.00
	DCC	7.59	0.62	0.00	7.33	0.37	0.00
	DCC-NL	7.58	0.62	0.00	7.32	0.37	0.00

Notes: This table shows the simulation results for the empirical GMV portfolios constructed based on different covariance matrix estimators and facing various degrees of leverage constraints. The disturbance terms of the simulated data are drawn from a multivariate standard normal distribution. The covariance matrix is estimated using the most recent 1250 daily returns based on six different methods, which are the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2019). $\gamma = \infty, 2, 1.6, 1$ stands for an increasing restriction with the proportion of short-sale not exceeding $\infty, 50\%, 30\%, 0$, respectively. The standard deviation of return (StdR) is calculated using the true covariance matrix, and thus it represents the actual risk. The standard deviation of weights (StdW) and the total short position of weights (ShortW) of the empirical portfolios are also reported. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively.

Table 2: Actual Risk of Portfolio and Standard Deviation and Total Short Position of Weights.

	$\widehat{\Sigma}$	Panel A: $N = 500$			Panel B: $N = 1000$		
		StdR	StdW	ShortW	StdR	StdW	ShortW
$\gamma = \infty$	S	5.54	1.42	216.01	6.48	0.74	235.76
	NL	5.01	1.12	164.14	3.84	0.54	160.09
	DCC	5.07	1.25	183.60	6.24	0.67	207.46
	DCC-NL	4.37	1.07	148.57	3.65	0.53	147.35
$\gamma = 2$	S	7.05	0.78	46.75	5.92	0.42	48.27
	NL	7.04	0.75	46.63	5.90	0.40	48.19
	DCC	5.01	0.70	44.10	4.47	0.39	46.76
	DCC-NL	4.98	0.70	44.08	4.42	0.38	46.71
$\gamma = 1.6$	S	8.09	0.75	28.75	6.86	0.41	29.30
	NL	8.12	0.72	28.72	6.89	0.40	29.29
	DCC	5.51	0.66	27.53	4.97	0.37	28.57
	DCC-NL	5.48	0.66	27.48	4.94	0.37	28.56
$\gamma = 1$	S	10.90	0.73	0.00	10.60	0.42	0.00
	NL	10.90	0.71	0.00	10.63	0.41	0.00
	DCC	7.62	0.59	0.00	7.20	0.35	0.00
	DCC-NL	7.61	0.59	0.00	7.19	0.35	0.00

Notes: This table shows the simulation results for the empirical GMV portfolios constructed based on different covariance matrix estimators and facing various degrees of leverage constraints. The disturbance terms of the simulated data are drawn from a multivariate t-distribution with 5 degrees of freedom. The covariance matrix is estimated using the most recent 1250 daily returns based on six different methods, which are the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2019). $\gamma = \infty, 2, 1.6, 1$ stands for an increasing restriction with the proportion of short-sale not exceeding $\infty, 50\%, 30\%, 0$, respectively. The standard deviation of return (StdR) is calculated using the true covariance matrix, and thus it represents the actual risk. The standard deviation of weights (StdW) and the total short position of weights (ShortW) of the empirical portfolios are also reported. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively.

Table 3: Out-of-sample Performance, Characteristics of Weights and Average Turnover of the Global Minimum Variance Portfolio.

	$\widehat{\Sigma}$	AvR	StdR	IR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
	S	10.23	10.78	0.95	-6.22	10.22	1.73	255.37	6.50
	NL	10.63	9.75	1.09	-2.84	4.88	1.00	140.40	2.35
$\gamma = \infty$	DCC	13.22	10.44	1.27	-4.29	16.16	1.61	180.43	4.11
	DCC-NL	12.94	9.55***	1.35	-2.26	14.70	1.27	110.21	2.04
	S	10.78	9.99	1.08	-3.83	10.13	1.15	82.71	3.19
	NL	10.69	9.86	1.08	-2.81	5.65	0.85	66.30	1.39
$\gamma = 2$	DCC	12.22	9.82	1.24	-2.94	17.32	1.34	68.78	2.88
	DCC-NL	12.38	9.52***	1.30	-2.09	16.26	1.23	58.87	1.36
	S	10.71	10.11	1.06	-3.46	10.76	1.09	54.55	2.80
	NL	10.77	10.06	1.07	-2.85	6.27	0.83	48.11	1.39
$\gamma = 1.6$	DCC	12.02	9.78	1.23	-2.66	18.44	1.35	50.12	2.76
	DCC-NL	12.24	9.59***	1.28	-1.99	17.67	1.27	41.18	1.29
	S	10.67	11.34	0.94	0.00	13.08	1.02	0.00	2.36
	NL	10.97	11.34	0.97	0.00	8.25	0.81	0.00	0.36
$\gamma = 1$	DCC	11.13	10.19	1.09	0.00	24.72	1.52	0.00	1.29
	DCC-NL	11.24	10.17	1.10	0.00	24.61	1.51	0.00	0.17
Panel B: 1000 stocks contained in the portfolio									
	S	10.16	13.54	0.75	-7.28	9.72	1.69	558.05	12.62
	NL	10.64	8.81	1.21	-1.44	2.44	0.49	142.97	3.26
$\gamma = \infty$	DCC	10.28	10.51	0.98	-4.46	19.55	1.37	337.03	7.42
	DCC-NL	11.26	8.16***	1.38	-1.15	16.12	0.79	97.57	2.73
	S	10.78	9.27	1.16	-2.71	8.38	0.70	90.91	3.56
	NL	10.67	9.13	1.17	-1.65	3.19	0.43	67.58	1.69
$\gamma = 2$	DCC	10.71	8.52	1.26	-2.06	21.18	0.97	82.00	3.30
	DCC-NL	11.24	8.11***	1.39	-1.13	18.08	0.82	56.67	1.66
	S	10.95	9.37	1.17	-2.40	8.88	0.66	54.55	2.96
	NL	10.92	9.41	1.16	-1.78	3.62	0.44	49.40	1.65
$\gamma = 1.6$	DCC	10.90	8.35	1.31	-1.74	22.64	0.98	52.96	2.96
	DCC-NL	11.23	8.10***	1.39	-1.09	20.17	0.88	38.75	1.48
	S	12.02	10.94	1.10	0.00	11.82	0.67	0.00	2.47
	NL	12.04	10.98	1.10	0.00	5.53	0.46	0.00	0.55
$\gamma = 1$	DCC	9.96	8.62	1.16	0.00	31.61	1.23	0.00	1.42
	DCC-NL	9.66	8.53***	1.13	0.00	30.88	1.22	0.00	0.22

Notes: This table shows the empirical results for the GMV portfolios constructed based on different covariance matrix estimators and facing various degrees of leverage constraints. The covariance matrix is estimated using the most recent 1250 daily returns based on four different methods, which are the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2019). $\gamma = \infty, 2, 1.6, 1$ stands for an increasing restriction with the proportion of short-sale not exceeding $\infty, 50\%, 30\%, 0$, respectively. We hold the portfolios for 21 days and record their daily returns. We report their out-of-sample annualized average return (AvR), annualized standard deviations (StdR), and Information Ratios (IR). Four characteristics of portfolio weights, including the minimum weight (MinW), the maximum weight (MaxW), the standard deviation of weights (StdW), and the total short positions of weights (ShortW), and the average turnover (AvT) of portfolios are also reported. All numbers shown are in percentage except those for Information Ratios. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively. In the rows labeled DCC and DCC-NL, significant outperformance of one of the two portfolios over the other in terms of StdR is denoted by asterisks: *** ,** ,* denote significance at the 0.01, 0.05, 0.1 levels, respectively.

Table 4: Out-of-sample Performance, Characteristics of Weights and Average Turnover of the Markowitz Portfolio Constructed based on the ROE Signal.

	$\widehat{\Sigma}$	AvR	StdR	IR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
	S	12.79	11.03	1.16	-6.53	10.39	1.79	267.08	6.72
	NL	13.77	10.00	1.38	-3.07	5.04	1.05	150.65	2.50
$\gamma = \infty$	DCC	15.36	10.73	1.43	-4.69	16.35	1.68	195.82	4.27
	DCC-NL	15.27	9.78	1.56***	-2.65	14.86	1.32	122.55	2.19
	S	12.97	10.23	1.27	-3.73	10.15	1.07	50.00	2.67
	NL	13.32	10.22	1.30	-3.00	6.17	0.86	50.00	1.12
$\gamma = 2$	DCC	14.18	10.05	1.41	-3.10	18.36	1.35	50.00	2.55
	DCC-NL	14.42	9.81	1.47**	-2.48	17.41	1.29	50.00	1.20
	S	12.63	10.56	1.20	-3.38	11.06	1.06	30.00	2.35
	NL	12.98	10.57	1.23	-3.00	7.11	0.85	30.00	1.14
$\gamma = 1.6$	DCC	13.61	10.17	1.34	-2.79	19.62	1.36	30.00	2.40
	DCC-NL	13.84	10.01	1.38**	-2.39	19.09	1.33	30.00	1.13
	S	12.74	12.30	1.04	0.00	14.90	1.14	0.00	2.30
	NL	12.83	12.33	1.04	0.00	10.79	0.94	0.00	0.32
$\gamma = 1$	DCC	12.14	11.41	1.06	0.00	22.75	1.47	0.00	1.14
	DCC-NL	12.08	11.41	1.06	0.00	22.68	1.47	0.00	0.15
Panel B: 1000 stocks contained in the portfolio									
	S	13.05	13.82	0.94	-7.57	9.80	1.73	571.36	12.88
	NL	14.57	9.02	1.62	-1.55	2.48	0.51	151.45	3.39
$\gamma = \infty$	DCC	13.46	10.95	1.23	-4.83	19.52	1.42	361.70	7.82
	DCC-NL	15.00	8.37	1.79***	-1.45	16.67	0.83	110.62	2.89
	S	13.36	9.36	1.43	-2.50	8.52	0.64	50.00	2.86
	NL	13.74	9.46	1.45	-1.77	3.53	0.44	50.00	1.36
$\gamma = 2$	DCC	14.12	8.59	1.64	-1.98	21.72	0.96	50.00	2.69
	DCC-NL	14.34	8.33	1.72**	-1.47	19.77	0.88	49.99	1.41
	S	13.20	9.72	1.36	-2.22	9.32	0.64	30.00	2.48
	NL	13.55	9.82	1.38	-1.92	4.17	0.45	30.00	1.35
$\gamma = 1.6$	DCC	13.75	8.61	1.60	-1.81	23.19	0.99	30.00	2.54
	DCC-NL	13.73	8.45	1.62	-1.45	22.04	0.95	30.00	1.25
	S	14.14	11.66	1.21	0.00	12.68	0.73	0.00	2.42
	NL	13.77	11.77	1.17	0.00	7.17	0.53	0.00	0.49
$\gamma = 1$	DCC	12.03	10.06	1.20	0.00	25.05	1.07	0.00	1.24
	DCC-NL	11.95	9.97	1.20	0.00	24.97	1.07	0.00	0.17

Notes: This table shows the empirical results for the Markowitz Portfolios constructed based on the signal Return-on-Equity (ROE), using different covariance matrix estimators and facing various degrees of leverage constraints. The covariance matrix is estimated using the most recent 1250 daily returns based on four different methods, which are the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2019). $\gamma = \infty, 2, 1.6, 1$ stands for an increasing restriction with the proportion of short-sale not exceeding $\infty, 50\%, 30\%, 0$, respectively. We hold the portfolios for 21 days and record their daily returns. We report their out-of-sample annualized average return (AvR), annualized standard deviations (StdR), and Information Ratios (IR). Four characteristics of portfolio weights, including the minimum weight (MinW), the maximum weight (MaxW), the standard deviation of weights (StdW), and the total short positions of weights (ShortW), and the average turnover (AvT) of portfolios are also reported. All numbers shown are in percentage except those for Information Ratios. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively. In the rows labeled DCC and DCC-NL, significant outperformance of one of the two portfolios over the other in terms of IR is denoted by asterisks: *** ,** ,* denote significance at the 0.01, 0.05, 0.1 levels, respectively.

Table 5: Out-of-sample Performance, Characteristics of Weights and Average Turnover of the Markowitz Portfolio Constructed based on the E/P Signal.

	$\widehat{\Sigma}$	AvR	StdR	IR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
	S	13.86	10.89	1.27	-6.46	10.49	1.77	263.44	6.73
	NL	14.64	9.90	1.48	-3.06	5.36	1.04	148.22	2.47
$\gamma = \infty$	DCC	16.37	10.58	1.55	-4.82	16.27	1.66	192.81	4.24
	DCC-NL	16.23	9.69	1.68***	-2.67	14.67	1.30	120.38	2.16
	S	13.96	10.14	1.38	-3.65	10.67	1.07	50.00	2.67
	NL	14.30	10.16	1.41	-2.94	6.80	0.85	50.00	1.13
$\gamma = 2$	DCC	15.52	9.91	1.57	-3.13	18.23	1.34	50.00	2.55
	DCC-NL	15.49	9.72	1.59	-2.47	17.16	1.27	50.00	1.20
	S	14.07	10.46	1.35	-3.31	11.76	1.06	30.00	2.36
	NL	14.46	10.51	1.38	-2.91	7.97	0.86	30.00	1.15
$\gamma = 1.6$	DCC	15.16	10.04	1.51	-2.83	19.63	1.36	30.00	2.41
	DCC-NL	15.11	9.92	1.52	-2.39	18.97	1.32	30.00	1.13
	S	15.50	12.66	1.22	0.00	16.42	1.20	0.00	2.34
	NL	15.61	12.80	1.22	0.00	12.52	1.00	0.00	0.32
$\gamma = 1$	DCC	15.04	11.67	1.29	0.00	23.66	1.51	0.00	1.15
	DCC-NL	15.02	11.59	1.30	0.00	23.50	1.51	0.00	0.15
Panel B: 1000 stocks contained in the portfolio									
	S	14.04	13.65	1.03	-7.65	10.09	1.72	567.54	12.84
	NL	15.51	8.93	1.74	-1.51	2.63	0.51	149.02	3.35
$\gamma = \infty$	DCC	14.74	10.77	1.37	-5.04	19.48	1.40	355.07	7.71
	DCC-NL	15.94	8.27	1.93***	-1.42	16.46	0.82	107.99	2.85
	S	15.02	9.28	1.62	-2.50	9.03	0.64	50.00	2.87
	NL	15.13	9.40	1.61	-1.76	3.78	0.43	50.00	1.38
$\gamma = 2$	DCC	15.77	8.51	1.85	-2.06	21.64	0.95	50.00	2.72
	DCC-NL	15.67	8.25	1.90	-1.45	19.45	0.87	49.99	1.43
	S	15.38	9.66	1.59	-2.17	9.95	0.64	30.00	2.51
	NL	15.43	9.80	1.57	-1.88	4.53	0.44	30.00	1.37
$\gamma = 1.6$	DCC	15.42	8.53	1.81	-1.81	23.19	0.99	30.00	2.56
	DCC-NL	15.31	8.36	1.83	-1.41	21.81	0.94	30.00	1.26
	S	17.49	12.01	1.46	0.00	13.84	0.75	0.00	2.45
	NL	17.47	12.18	1.43	0.00	8.19	0.55	0.00	0.50
$\gamma = 1$	DCC	15.12	10.27	1.47	0.00	26.05	1.10	0.00	1.28
	DCC-NL	14.98	10.18	1.47	0.00	25.70	1.10	0.00	0.17

Notes: This table shows the empirical results for the Markowitz Portfolios constructed based on a Value Signal (E/P), using different covariance matrix estimators and facing various degrees of leverage constraints. The covariance matrix is estimated using the most recent 1250 daily returns based on four different methods, which are the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2019). $\gamma = \infty, 2, 1.6, 1$ stands for an increasing restriction with the proportion of short-sale not exceeding $\infty, 50\%, 30\%, 0$, respectively. We hold the portfolios for 21 days and record their daily returns. We report their out-of-sample annualized average return (AvR), annualized standard deviations (StdR), and Information Ratios (IR). Four characteristics of portfolio weights, including the minimum weight (MinW), the maximum weight (MaxW), the standard deviation of weights (StdW), and the total short positions of weights (ShortW), and the average turnover (AvT) of portfolios are also reported. All numbers shown are in percentage except those for Information Ratios. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively. In the rows labeled DCC and DCC-NL, significant outperformance of one of the two portfolios over the other in terms of IR is denoted by asterisks: *** ,** ,* denote significance at the 0.01, 0.05, 0.1 levels, respectively.

Table 6: Out-of-sample Performance of the Markowitz Portfolio Constructed based on the ROE Signal with Transaction Costs Considered.

	$\widehat{\Sigma}$	Spread = 3 basis-points			Spread = 5 basis-points		
		AvR	StdR	IR	AvR	StdR	IR
Panel A: 500 stocks contained in the portfolio							
$\gamma = \infty$	S	10.37	11.04	0.94	8.75	11.07	0.79
	NL	10.45	10.03	1.04	8.23	10.10	0.82
	DCC	11.40	10.77	1.06	8.77	10.87	0.81
	DCC-NL	12.06	9.81	1.23***	9.92	9.88	1.00***
$\gamma = 2$	S	9.59	10.26	0.93	7.33	10.33	0.71
	NL	10.49	10.23	1.03	8.61	10.28	0.84
	DCC	10.84	10.08	1.08	8.61	10.14	0.85
	DCC-NL	11.56	9.83	1.18***	9.66	9.88	0.98***
$\gamma = 1.6$	S	9.36	10.58	0.88	7.18	10.64	0.68
	NL	10.14	10.59	0.96	8.25	10.63	0.78
	DCC	10.32	10.20	1.01	8.13	10.26	0.79
	DCC-NL	11.01	10.03	1.10***	9.13	10.08	0.91***
$\gamma = 1$	S	9.49	12.32	0.77	7.33	12.37	0.59
	NL	10.29	12.34	0.83	8.59	12.37	0.69
	DCC	9.30	11.43	0.81	7.41	11.47	0.65
	DCC-NL	9.60	11.42	0.84***	7.95	11.45	0.69***
Panel B: 1000 stocks contained in the portfolio							
$\gamma = \infty$	S	8.75	11.07	0.79	5.31	13.97	0.38
	NL	8.23	10.10	0.82	4.81	9.40	0.51
	DCC	8.77	10.87	0.81	1.05	11.45	0.09
	DCC-NL	9.92	9.88	1.00***	5.54	8.75	0.63***
$\gamma = 2$	S	7.33	10.33	0.71	3.91	9.69	0.40
	NL	8.61	10.28	0.84	5.19	9.72	0.53
	DCC	8.61	10.14	0.85	4.78	8.94	0.53
	DCC-NL	9.66	9.88	0.98***	5.76	8.64	0.67***
$\gamma = 1.6$	S	7.18	10.64	0.68	3.98	10.02	0.40
	NL	8.25	10.63	0.78	5.00	10.08	0.50
	DCC	8.13	10.26	0.79	4.49	8.96	0.50
	DCC-NL	9.13	10.08	0.91***	5.25	8.75	0.60***
$\gamma = 1$	S	7.33	12.37	0.59	4.95	11.91	0.42
	NL	8.59	12.37	0.69	5.74	11.96	0.48
	DCC	7.41	11.47	0.65	3.55	10.30	0.34
	DCC-NL	7.95	11.45	0.69***	4.11	10.18	0.40***

Notes: This table shows the empirical results for the Markowitz Portfolios constructed based on the signal Return-on-Equity (ROE), using different covariance matrix estimators and facing various degrees of leverage constraints when transaction costs are considered. The covariance matrix is estimated using the most recent 1250 daily returns based on four different methods, which are the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2019). $\gamma = \infty, 2, 1.6, 1$ stands for an increasing restriction with the proportion of short-sale not exceeding $\infty, 50\%, 30\%, 0$, respectively. We hold the portfolios for 21 days and record their daily returns. We report their out-of-sample annualized average return (AvR), annualized standard deviations (StdR), and Information Ratios (IR). AvR and StdR are shown in percentage. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively. The left and right panels show results under the assumptions of 3 basis-points and 5 basis-points bid-ask spreads, respectively. In the rows labeled DCC and DCC-NL, significant outperformance of one of the two portfolios over the other in terms of IR is denoted by asterisks: *** ,** ,* denote significance at the 0.01, 0.05, 0.1 levels, respectively.
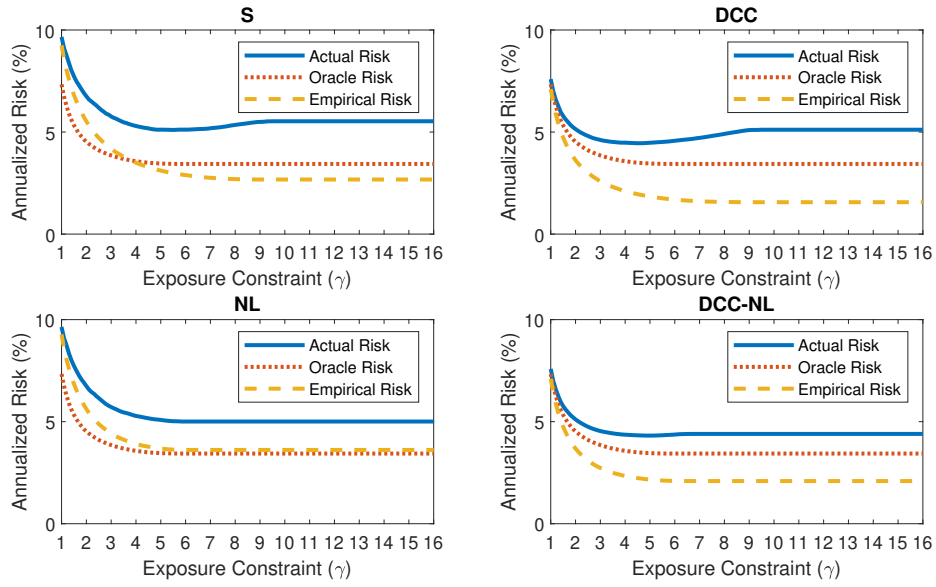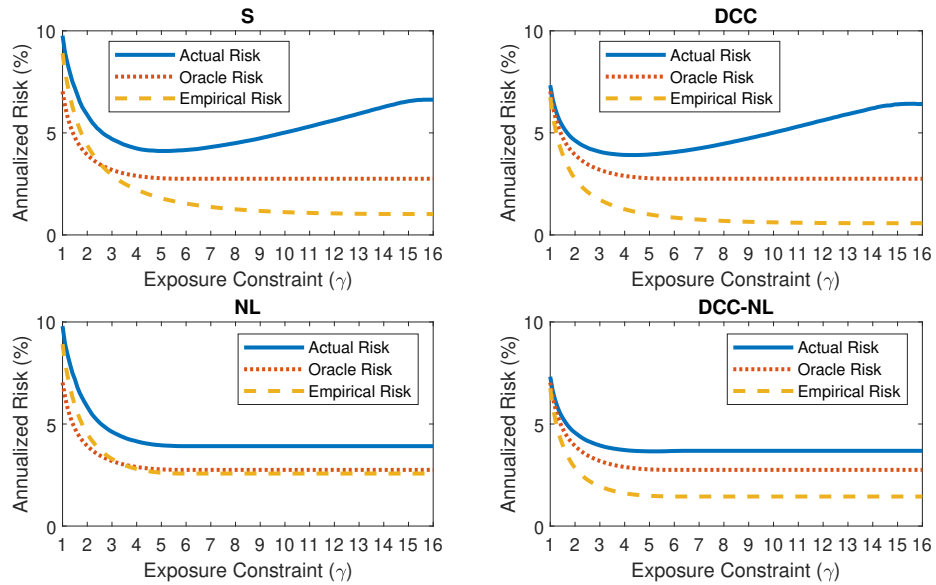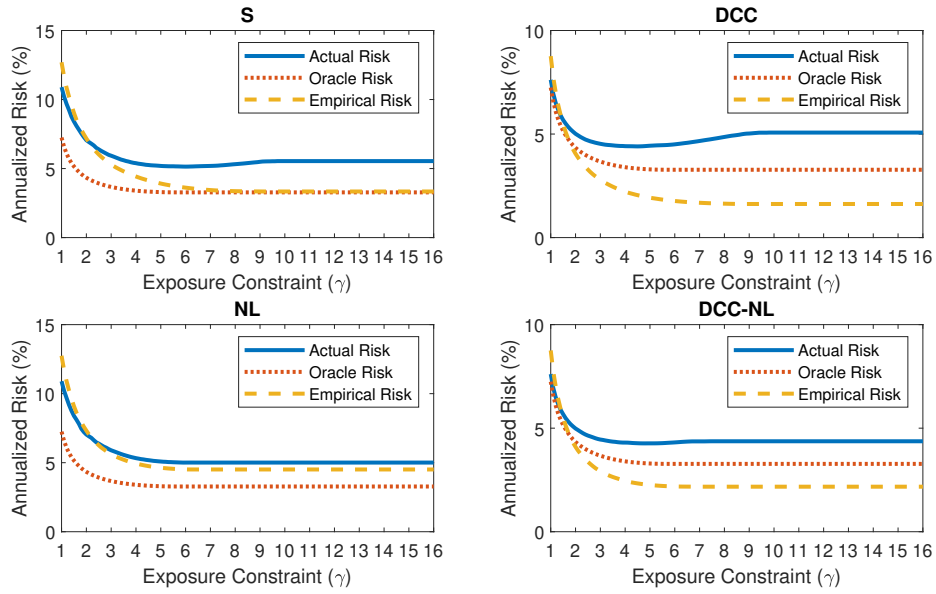
Figure 1: Comparisons of the oracle risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ). $N = 500$, and the disturbance terms are drawn from a multivariate standard normal distribution.
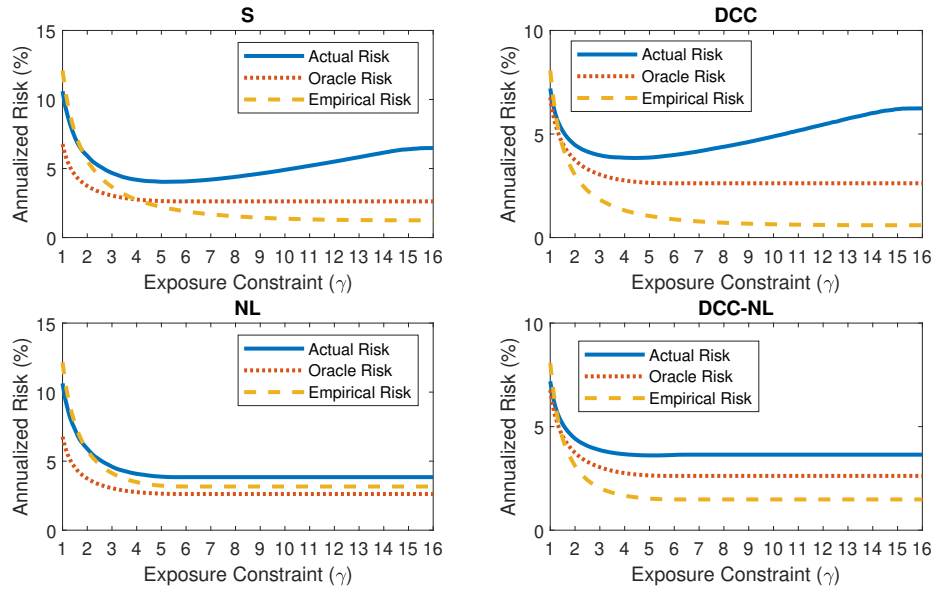

Figure 2: Comparisons of the oracle risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ). $N = 1000$, and the disturbance terms are drawn from a multivariate standard normal distribution.

Figure 3: Comparisons of the oracle risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ). $N = 500$, and the disturbance terms are drawn from a multivariate t-distribution with 5 degrees of freedom.

Figure 4: Comparisons of the oracle risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ). $N = 1000$, and the disturbance terms are drawn from a multivariate t-distribution with 5 degrees of freedom.
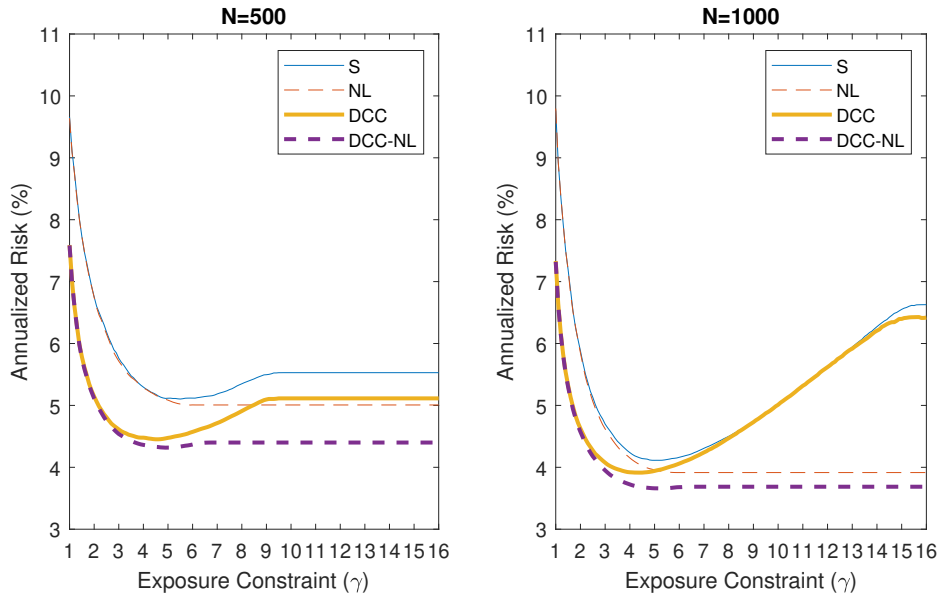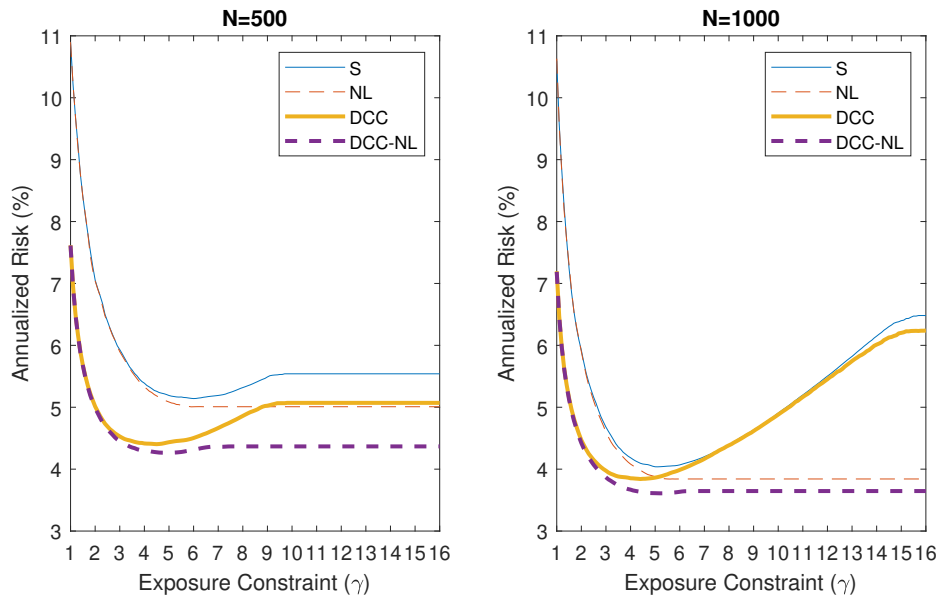
Figure 5: Comparisons of the average actual risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ). The disturbance terms of the simulated data are drawn from a multivariate standard normal distribution.

Figure 6: Comparisons of the average actual risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ). The disturbance terms of the simulated data are drawn from a multivariate t-distribution with 5 degrees of freedom.
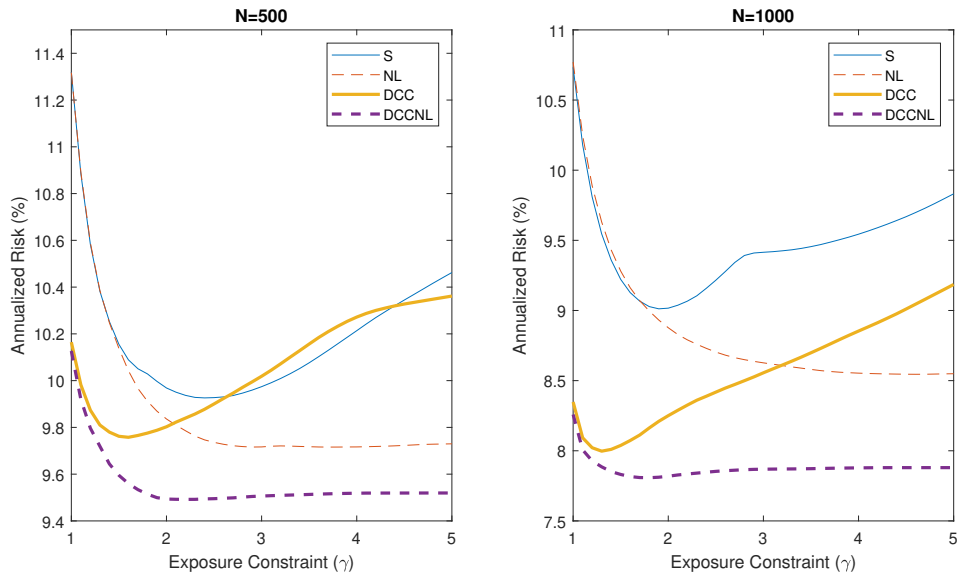
Figure 7: Comparisons of the out-of-sample risks of the GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ).

A Theoretical Justifications

Proof of Theorem 1 (i). Note that the matrix $\tilde{\Sigma}_{\gamma,t}$ is obviously symmetric and the solution to problem (2.8) is denoted by $\boldsymbol{w}_t^*$. For any vector x,

$$
\begin{aligned}
x'\tilde{\Sigma}_{\gamma,t}x &= x'\widehat{\Sigma}_t x + \frac{1}{2}\lambda(x'\mathbf{g}_t^*\mathbf{1}'x + x'\mathbf{1}\mathbf{g}_t^{*'}x)\\
&= x'\widehat{\Sigma}_t x + \lambda(x'\mathbf{g}_t^*)(\mathbf{1}'x).
\end{aligned}
\tag{1.1}
$$

Based on the KKT conditions in (2.10), $2\widehat{\Sigma}_t\boldsymbol{w}_t^* - \mu\mathbf{1} + \lambda\mathbf{g}_t^* = 0$. Therefore,

$$
\lambda(x'\mathbf{g}_t^*)(\mathbf{1}'x) = -2(x'\mathbf{1})(x'\widehat{\Sigma}_t\boldsymbol{w}_t^*) + \mu(x'\mathbf{1})^2.
\tag{1.2}
$$

Note that

$$
|(x'\mathbf{1})(x'\widehat{\Sigma}_t\boldsymbol{w}_t^*)| = |(x'\mathbf{1})(x'\widehat{\Sigma}_t^{\frac{1}{2}})(\widehat{\Sigma}_t^{\frac{1}{2}}\boldsymbol{w}_t^*)| \le |(x'\mathbf{1})(x'\widehat{\Sigma}_t x)^{\frac{1}{2}}(\boldsymbol{w}_t^{*'}\widehat{\Sigma}_t\boldsymbol{w}_t^*)^{\frac{1}{2}}|,
$$

where the equality holds because of the positive definiteness of the DCC estimator $\widehat{\Sigma}_t$, and the inequality could be obtained by Cauchy-Schwartz inequality.

In addition, because the DCC estimator $\widehat{\Sigma}_t$ is positive definite under some conditions, we have

$$
0 < \boldsymbol{w}_t^{*'}\widehat{\Sigma}_t\boldsymbol{w}_t^* = \frac{1}{2}\mu\boldsymbol{w}_t^{*'}\mathbf{1} - \frac{1}{2}\lambda\boldsymbol{w}_t^{*'}\mathbf{g}_t = \frac{1}{2}\mu - \frac{1}{2}\lambda\|\boldsymbol{w}_t^*\|_1 \le \frac{1}{2}\mu.
$$

Hence,

$$
|(x'\mathbf{1})(x'\widehat{\Sigma}_t\boldsymbol{w}_t^*)| \le |x'\mathbf{1}|(x'\widehat{\Sigma}_t x)^{\frac{1}{2}}(\frac{1}{2}\mu)^{\frac{1}{2}}.
\tag{1.3}
$$

Combining (1.1)-(1.3), we have

$$
\begin{aligned}
x'\tilde{\Sigma}_{\gamma,t}x &= x'\widehat{\Sigma}_t x - 2(x'\mathbf{1})(x'\widehat{\Sigma}_t\boldsymbol{w}_t^*) + \mu(x'\mathbf{1})^2\\
&\ge x'\widehat{\Sigma}_t x - 2|(x'\mathbf{1})(x'\widehat{\Sigma}_t\boldsymbol{w}_t^*)| + \mu(x'\mathbf{1})^2\\
&\ge x'\widehat{\Sigma}_t x - 2|x'\mathbf{1}|(x'\widehat{\Sigma}_t x)^{\frac{1}{2}}(\frac{1}{2}\mu)^{\frac{1}{2}} + \mu(x'\mathbf{1})^2\\
&= (a-b)^2 + b^2,
\end{aligned}
\tag{1.4}
$$

where $a = (x'\widehat{\Sigma}_t x)^{\frac{1}{2}}$, and $b = (\frac{1}{2}\mu)^{\frac{1}{2}}|x'\mathbf{1}|$.

Moreover, $(a-b)^2 + b^2$ is always nonnegative and is zero if and only if $a = b$ and $b = 0$ hold simultaneously. However, $a = (x'\widehat{\Sigma}_t x)^{\frac{1}{2}} > 0$ because $\widehat{\Sigma}_t$ is positive definite. Therefore, for any vector x, $x'\tilde{\Sigma}_{\gamma,t}x > 0$ holds. This indicates the positive definiteness of $\tilde{\Sigma}_{\gamma,t}$. ∎

Proof of Theorem 1 (ii). Firstly, the optimization problem (2.11) with equality constraint could be solved through the Lagrange multiplier method. Construct the Lagrangian

$$
L(\boldsymbol{w}_t, \mu_\gamma) = \boldsymbol{w}_t'\tilde{\Sigma}_{\gamma,t}\boldsymbol{w}_t - \mu_\gamma(\boldsymbol{w}_t'\mathbf{1} - 1),
$$

then the solution $\boldsymbol{w}_t^{opt}$ to this minimization problem should satisfy

$$
\begin{cases}
2\tilde{\Sigma}_{\gamma,t}\boldsymbol{w}_t^{opt} - \mu_\gamma\mathbf{1} = 0,\\
\boldsymbol{w}_t^{opt'}\mathbf{1} - 1 = 0.
\end{cases}
$$

Because $\tilde{\Sigma}_{\gamma,t}$ is invertible, then the solution to this problem is given by

$$
\boldsymbol{w}_t^{opt} = \frac{\tilde{\Sigma}_{\gamma,t}^{-1}\mathbf{1}}{\mathbf{1}'\tilde{\Sigma}_{\gamma,t}^{-1}\mathbf{1}}.
\tag{1.5}
$$

By the Lagrange multiplier method, problem (2.8) is to minimize

$$L(\boldsymbol{w}_t, \mu, \lambda) \;=\; \boldsymbol{w}_t' \widehat{\varSigma}_t \boldsymbol{w}_t - \mu(\boldsymbol{w}_t'\mathbf{1} - 1) - \lambda(\gamma - \|\boldsymbol{w}_t\|_1).$$

Based on the fact that $\mathbf{g}_t^{*'} \boldsymbol{w}_t^* = \|\boldsymbol{w}_t^*\|_1$ and KKT conditions in (2.10), we have

$$
\begin{aligned}
\tilde{\varSigma}_{\gamma,t} \boldsymbol{w}_t^* &= \widehat{\varSigma}_t \boldsymbol{w}_t^* + \frac{1}{2}\lambda g_t^* \mathbf{1}' \boldsymbol{w}_t^* + \frac{1}{2}\lambda \mathbf{1} g_t^{*'} \boldsymbol{w}_t^* \\
&= \widehat{\varSigma}_t \boldsymbol{w}_t^* + \frac{1}{2}\lambda g_t^* + \frac{1}{2}\lambda \|\boldsymbol{w}_t^*\|_1 \mathbf{1} \\
&= \frac{1}{2}(\lambda\gamma + \mu)\mathbf{1}.
\end{aligned}
$$

Hence, the solution to problem (2.8) $\boldsymbol{w}_t^* = \frac{1}{2}(\lambda\gamma + \mu)\tilde{\varSigma}_{\gamma,t}^{-1}\mathbf{1}$. Moreover, because of the constraint $\boldsymbol{w}_t'\mathbf{1} = 1$, solving for $\lambda\gamma + \mu$ yields $\lambda\gamma + \mu = \frac{2}{\mathbf{1}'\tilde{\varSigma}_{\gamma,t}^{-1}\mathbf{1}}$. This fact indicates that

$$\boldsymbol{w}_t^* = \boldsymbol{w}_t^{opt}.$$

Therefore, it implies the equivalence of the partial constrained optimization problem and the (unconstrained) optimization problem with regularized covariance matrix estimator. ∎

Proof of Theorem 2. Firstly, the optimization problem (2.17) with equality constraint could be solved through the Lagrange multiplier method. Construct the Lagrangian

$$L(\boldsymbol{w}_t, \mu_{1\gamma}, \mu_{2\gamma}) = \boldsymbol{w}_t' \tilde{\varSigma}_{\gamma,t} \boldsymbol{w}_t - \mu_{1\gamma}(\boldsymbol{w}_t'\mathbf{1} - 1) - \mu_{2\gamma}(\boldsymbol{w}_t'\boldsymbol{m}_t - b_t),$$

then the solution $\boldsymbol{w}_{b,t}^{opt}$ to this minimization problem should satisfy

$$
\begin{cases}
2\tilde{\varSigma}_{\gamma,t} \boldsymbol{w}_{b,t}^{opt} - \mu_{1\gamma}\mathbf{1} - \mu_{2\gamma}\boldsymbol{m}_t = 0, \\
\boldsymbol{w}_{b,t}^{opt'}\mathbf{1} - 1 = 0, \quad \boldsymbol{w}_{b,t}^{opt'}\boldsymbol{m}_t - b_t = 0.
\end{cases}
$$

Therefore, $\boldsymbol{w}_{b,t}^{opt} = \mu_{1c}\tilde{\varSigma}_{\gamma,t}^{-1}\mathbf{1} + \mu_{2\gamma}\tilde{\varSigma}_{\gamma,t}^{-1}\boldsymbol{m}_t = \tilde{\varSigma}_{\gamma,t}^{-1}(\mathbf{1}, \boldsymbol{m}_t)\begin{pmatrix}\mu_{1\gamma}\\\mu_{2\gamma}\end{pmatrix}$.

The above equations also imply that

$$
\begin{aligned}
1 &= \frac{1}{2}\mu_{1\gamma}\mathbf{1}'\tilde{\varSigma}_{\gamma,t}^{-1}\mathbf{1} + \frac{1}{2}\mu_{2\gamma}\mathbf{1}'\tilde{\varSigma}_{\gamma,t}^{-1}\boldsymbol{m}_t, \\
b_t &= \frac{1}{2}\mu_{1\gamma}m_t'\tilde{\varSigma}_{\gamma,t}^{-1}\mathbf{1} + \frac{1}{2}\mu_{2\gamma}m_t'\tilde{\varSigma}_{\gamma,t}^{-1}\boldsymbol{m}_t,
\end{aligned}
$$

or

$$\begin{pmatrix}1\\b_t\end{pmatrix} = \frac{1}{2}(\mathbf{1}, \boldsymbol{m}_t)'\tilde{\varSigma}_{\gamma,t}^{-1}(\mathbf{1}, \boldsymbol{m}_t)\begin{pmatrix}\mu_{1\gamma}\\\mu_{2\gamma}\end{pmatrix}.$$

Solving for $(\mu_{1\gamma}, \mu_{2\gamma})'$ yields

$$\begin{pmatrix}\mu_{1\gamma}\\\mu_{2\gamma}\end{pmatrix} = 2\left[(\mathbf{1}, \boldsymbol{m}_t)'\tilde{\varSigma}_{\gamma,t}^{-1}(\mathbf{1}, \boldsymbol{m}_t)\right]\begin{pmatrix}1\\b_t\end{pmatrix}.$$

Therefore, the solution to this problem is given by

$$\boldsymbol{w}_{b,t}^{opt} = \tilde{\varSigma}_{\gamma,t}^{-1}(\mathbf{1}, \boldsymbol{m}_t)\left[(\mathbf{1}, \boldsymbol{m}_t)'\tilde{\varSigma}_{\gamma,t}^{-1}(\mathbf{1}, \boldsymbol{m}_t)\right]\begin{pmatrix}1\\b_t\end{pmatrix}. \tag{1.6}$$

By the Lagrange multiplier method, problem (2.12) is to minimize

$$L(\boldsymbol{w}_t, \mu_1, \mu_2, \lambda) \;=\; \boldsymbol{w}_t'\widehat{\varSigma}_t \boldsymbol{w}_t - \mu_1(\boldsymbol{w}_t'\mathbf{1} - 1) - \mu_2(\boldsymbol{w}_t'\boldsymbol{m}_t - b_t) - \lambda(\gamma - \|\boldsymbol{w}_t\|_1).$$

So, the KKT conditions are

$$
\begin{cases}
2\widehat{\Sigma}_t \boldsymbol{w}_t - \mu_1 \mathbf{1} - \mu_2 \boldsymbol{m}_t + \lambda \mathbf{g}_t = 0, \\
\lambda(\gamma - \|\boldsymbol{w}_t\|_1) = 0, \quad \lambda \geq 0, \\
\|\boldsymbol{w}_t\|_1 \leq \gamma, \quad \boldsymbol{w}_t'\mathbf{1} - 1 = 0, \quad \boldsymbol{w}_t'\boldsymbol{m}_t - b_t = 0.
\end{cases}
\tag{1.7}
$$

Based on the fact that $\mathbf{g}_{b,t}^{*\prime}\boldsymbol{w}_{b,t}^* = \|\boldsymbol{w}_{b,t}^*\|_1$ and KKT conditions in (1.7), we have

$$
\begin{aligned}
\tilde{\Sigma}_{\gamma,t}\boldsymbol{w}_{b,t}^* &= \widehat{\Sigma}_t \boldsymbol{w}_{b,t}^* + \frac{1}{2}\lambda g_{b,t}^* \mathbf{1}' \boldsymbol{w}_{b,t}^* + \frac{1}{2}\lambda \mathbf{1} g_{b,t}^{*\prime} \boldsymbol{w}_{b,t}^* \\
&= \widehat{\Sigma}_t \boldsymbol{w}_{b,t}^* + \frac{1}{2}\lambda g_{b,t}^* + \frac{1}{2}\lambda \|\boldsymbol{w}_{b,t}^*\|_1 \mathbf{1} \\
&= \frac{1}{2}(\lambda\gamma + \mu_1)\mathbf{1} + \frac{1}{2}\mu_2 \boldsymbol{m}_t.
\end{aligned}
$$

It then follows that $\boldsymbol{w}_{b,t}^* = \frac{1}{2}\tilde{\Sigma}_{\gamma,t}^{-1}[(\lambda\gamma + \mu_1)\mathbf{1} + \mu_2 \boldsymbol{m}_t]$. The constraints also imply that

$$
\begin{aligned}
1 &= \frac{1}{2}(\lambda\gamma + \mu_1)\mathbf{1}'\tilde{\Sigma}_{\gamma,t}^{-1}\mathbf{1} + \frac{1}{2}\mu_2 \mathbf{1}'\tilde{\Sigma}_{\gamma,t}^{-1}\boldsymbol{m}_t, \\
b_t &= \frac{1}{2}(\lambda\gamma + \mu_1)m_t'\tilde{\Sigma}_{\gamma,t}^{-1}\mathbf{1} + \frac{1}{2}\mu_2 m_t'\tilde{\Sigma}_{\gamma,t}^{-1}\boldsymbol{m}_t,
\end{aligned}
$$

or

$$
\begin{pmatrix} 1 \\ b_t \end{pmatrix} = \frac{1}{2}(\mathbf{1},\boldsymbol{m}_t)'\tilde{\Sigma}_{\gamma,t}^{-1}(\mathbf{1},\boldsymbol{m}_t)\begin{pmatrix} \lambda\gamma + \mu_1 \\ \mu_2 \end{pmatrix}.
$$

Solving for $(\lambda\gamma + \mu_1, \mu_2)'$ yields

$$
\begin{pmatrix} \lambda\gamma + \mu_1 \\ \mu_2 \end{pmatrix} = 2\left[(\mathbf{1},\boldsymbol{m}_t)'\tilde{\Sigma}_{\gamma,t}^{-1}(\mathbf{1},\boldsymbol{m}_t)\right]\begin{pmatrix} 1 \\ b_t \end{pmatrix}.
$$

Hence,

$$
\boldsymbol{w}_{b,t}^* = \tilde{\Sigma}_{\gamma,t}^{-1}(\mathbf{1},\boldsymbol{m}_t)\left[(\mathbf{1},\boldsymbol{m}_t)'\tilde{\Sigma}_{\gamma,t}^{-1}(\mathbf{1},\boldsymbol{m}_t)\right]\begin{pmatrix} 1 \\ b_t \end{pmatrix}.
\tag{1.8}
$$

We can then conclude that $\boldsymbol{w}_{b,t}^{opt} = \boldsymbol{w}_{b,t}^*$. This completes the proof. ∎

References

Ana Avramovic and Phil Mackintosh. Inside the NBBO: Pushing for wider — and narrower! — spreads. *Trading strategy: Market commentary, Credit Suisse Research and Analytics*, 2013.

Sanjoy Basu. The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics*, 12(1):129–156, 1983.

Patrick Behr, Andre Guettler, and Felix Miebs. On portfolio optimization: Imposing the right constraints. *Journal of Banking & Finance*, 37(4):1232–1242, 2013.

Tim Bollerslev, Andrew J. Patton, and Rogier Quaedvlieg. Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207(1):71–91, 2018.

Michael W. Brandt, Pedro Santa-Clara, and Rossen Valkanov. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, 22(9):3411–3447, 2009.

Nicole Branger, Katarína Lučivjanská, and Alex Weissensteiner. Optimal granularity for portfolio choice. *Journal of Empirical Finance*, 50:125–146, 2019.

Mark Britten-Jones. The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671, 1999.

Joshua Brodie, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris. Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272, 2009.

Laurent Callot, Mehmet Caner, A. Özlem Önder, and Esra Ulasan. A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*, 2020. Forthcoming.

Victor DeMiguel, Lorenzo Garlappi, Francisco J. Nogales, and Raman Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812, 2009a.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial studies*, 22(5):1915–1953, 2009b.

Victor DeMiguel, Alberto Martin-Utrera, and Francisco J. Nogales. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034, 2013.

Robert F. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

Robert F. Engle, Olivier Ledoit, and Michael Wolf. Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375, 2019.

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

Guanhao Feng, Stefano Giglio, and Dacheng Xiu. Taming the factor zoo: A test of new factors. *The Journal of Finance*, 2020. Forthcoming.

Lorenzo Garlappi, Raman Uppal, and Tan Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81, 2007.

Jeremiah Green, John R. M. Hand, and X. Frank Zhang. The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies*, 30(12):4389–4436, 2017.

Campbell R. Harvey, Yan Liu, and Heqing Zhu. and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016.

Robert A Haugen and Nardin L Baker. Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3):401–439, 1996.

Kewei Hou, Chen Xue, and Lu Zhang. Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705, 2015.

Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 54(4):1651–1684, 2003.

Raymond Kan and Guofu Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656, 2007.

Oliver Ledoit and Michael Wolf. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859, 2008.

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

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

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

Olivier Ledoit and Michael Wolf. Robust performances hypothesis testing with the variance. *Wilmott Magazine*, 2011(55):86–89, 2011.

Olivier Ledoit and Michael Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060, 2012.

Olivier Ledoit and Michael Wolf. Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139(2):360–384, 2015.

Olivier Ledoit and Michael Wolf. Numerical implementation of the QuEST function. *Computational Statistics & Data Analysis*, 115:199–223, 2017.

Olivier Ledoit, Michael Wolf, and Zhao Zhao. Efficient sorting: A more powerful test for cross-sectional anomalies. *Journal of Financial Econometrics*, 17(4):645–686, 2019.

Jiahan Li. Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business & Economic Statistics*, 33(3):381–392, 2015.

V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483, 1967.

Harry Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

Xiaoling Mei and Francisco J. Nogales. Portfolio selection with proportional transaction costs and predictability. *Journal of Banking & Finance*, 94:131–151, 2018.

Xiaoling Mei, Victor DeMiguel, and Francisco J. Nogales. Multiperiod portfolio optimization with multiple risky assets and general transaction costs. *Journal of Banking & Finance*, 69:108–120, 2016.

Cavit Pakel, Neil Shephard, Kevin Sheppard, and Robert F Engle. Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics*, 2020. Forthcoming.

William F. Sharpe. A simplified model for portfolio analysis. *Management Science*, 9(1):277–293, 1963.

Jun Tu and Guofu Zhou. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215, 2011.

Kevin Webster, Yin Luo, Miguel A. Alvarez, Javed Jussa, Sheng Wang, Gaurav Rohal, Allen Wang, David Elledge, and George Zhao. A portfolio managers guidebook to trade execution: From light rays to dark pools. *Deutsche Bank Markets Research*, Quantitative Strategy(Signal Processing), 2015.