

Zhao, Zhao; Ledoit, Olivier; Jiang, Hui

Working Paper

Risk reduction and efficiency increase in large portfolios: Leverage and shrinkage

Working Paper, No. 328

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Zhao, Zhao; Ledoit, Olivier; Jiang, Hui (2019) : Risk reduction and efficiency increase in large portfolios: Leverage and shrinkage, Working Paper, No. 328, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-172206>

This Version is available at:

<https://hdl.handle.net/10419/201533>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich** ^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 328

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

Zhao Zhao, Olivier Ledoit and Hui Jiang

July 2019

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

Two basic solutions have been proposed to fix the well-documented incompatibility of the sample covariance matrix with Markowitz mean-variance portfolio optimization: first, restrict leverage so much that no short sales are allowed; or, second, linearly shrink the sample covariance matrix towards a parsimonious target. Mathematically, there is a deep connection between the two approaches, and empirically they display similar performances. Recent developments have turned the choice between no-short-sales and linear shrinkage into a false 'either-or' dichotomy. What if, instead of 0% leverage we considered fully-invested, long-short 130/30 portfolios, or even 150/50, given that prime brokers, fund regulators and investors have started to allow it? And instead of linearly shrinking the unconditional covariance matrix, what if we allowed for 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 empirical evidence finds that, indeed, these new developments enable us to have 'the best of both worlds' by combining some appropriate leverage constraint with a judiciously chosen shrinkage method. The overall winner is a 150/50 investment strategy where the covariance matrix estimator is a combination of DCC (Dynamic Conditional Correlation — a well-known Multivariate GARCH model) — with NL (Non-Linear shrinkage, a substantial upgrade upon linear shrinkage technology); although 130/30 DCC-NL comes a close second. This is true both in the 'pure' case of estimating the Global Minimum Variance portfolio, and also for textbook-style construction of Markowitz mean-variance efficient portfolio.

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

JEL CLASSIFICATION NOS: C13, C58, G11.

*Corresponding Author.

1 Introduction

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 “Global Minimum Variance” Portfolio); or (ii) are efficient in terms of risk-return trade-off (the “Tangency” Portfolio). The first type is a *pure* test of the covariance matrix estimator and the optimization program used because the second one requires, in addition, a good predictive model for expected returns — or abnormal expected returns —, which is notoriously hard to obtain. Given that the second problem is more common in basic Finance textbooks, we consider both exercises.

Here we only 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 100/0 (no short sales at all), 130/30, and 150/50 (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).

[Jagannathan and Ma \(2003\)](#) explain why imposing the wrong constraints helps reducing the risks. They claim that large negative weights imply high sample covariances between stocks, while the high sample covariances are likely to be caused by upward-biased estimation errors. Thus, to some extent imposing the nonnegative constraints on portfolio weights reduces the sampling error. This mechanism is analogous to shrinkage estimation of the covariance matrix. Comparing effects of the two mechanisms, [Jagannathan and Ma \(2003\)](#) show that with the no-short-sale constraint in place, the sample covariance matrix performs as well as the shrinkage estimator. Moreover, they observe that if the shrinkage estimator is used in constructing portfolios, imposing the no-short-sale constraint would hurt the out-of-sample performance. [Demiguel et al. \(2009\)](#) use the norm framework to unify the shrinkage effects in weights and in the covariance matrix. Further, [Fan et al. \(2012\)](#) extend the no-short-sale constraint to varying degrees of leverage constraints and provide mathematical insights. Building on their work, we provide both theoretical insights and empirical evidence for the advantage of the 130/30 and 150/50 strategies, and the outperformance of the shrinkage covariance matrix estimators.

The shrinkage estimator used in [Jagannathan and Ma \(2003\)](#) and [Demiguel et al. \(2009\)](#) is a weighted average of the sample covariance matrix and the [Sharpe \(1963\)](#) one-factor model-based estimator, proposed by [Ledoit and Wolf \(2003\)](#). The estimation is substantially equivalent to linearly shrinking the sample eigenvalues towards a more centralized set of eigenvalues by a unified shrinkage intensity, while retaining the sample eigenvectors. [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 remains the same, it is necessary to compare the effect of using the nonlinear

shrinkage estimator with that of imposing varying degrees of leverage constraints.

Further, [Engle et al. \(2017\)](#) use the nonlinear shrinkage estimator to replace the sample covariance matrix in the “correlation targeting” maximum likelihood estimation of the dynamic conditional correlation (DCC) model, and propose the DCC-NL estimator, which turns out to perform better than the DCC estimator, the NL estimator and other conventional estimators for the covariance matrix. Since the DCC model works in capturing the conditional heteroscedasticity, which is totally different to the shrinkage mechanism, we conjecture that the use of DCC model would also help improving the out-of-sample performance of portfolios with leverage constraints.

The two main original contribution of this paper are as follows. First, we demonstrate that loosening the leverage constrain from zero leverage (100/0) to 130/30 or even 150/50 delivers better performance, both in terms of constructing a global minimum variance portfolio and a mean-variance efficient portfolio. Second, we show that upgrading from first-generation linear shrinkage technology for the estimation of the covariance matrix to a 1000-dimensional Multivariate GARCH model with built-in nonlinear shrinkage is independently and additively beneficial. In short, we find that as long as there is *some* leverage — which can be 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.

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, how we construct the global minimum variance portfolios and the Markowitz mean-variance efficient portfolios with the momentum signal, and our extension compared to the research of [Jagannathan and Ma \(2003\)](#). In Section 3, we use Monte Carlo simulations to verify our theoretical results. Section 4.1 describes our data. In Section 4, we report the results for out-of-sample performance of the global minimum variance portfolios and the Markowitz 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 \(2012, 2015\)](#), however, allow the

sample eigenvalues to adjust with heterogeneous shrinkage intensities, and thus 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 $(\lambda_1, \dots, \lambda_N)$ denote a set of eigenvalues of the $N \times N$ sample covariance matrix S , sorted in descending order, and $(\mathbf{u}_1, \dots, \mathbf{u}_N)$ be the corresponding eigenvectors. Let $Q_{T,N}(\mathbf{t}) := (q_{T,N}^1(\mathbf{t}), \dots, q_{T,N}^N(\mathbf{t}))'$ denote the QuEST function, which turns the set of population eigenvalues $\mathbf{t} := (t_1, \dots, t_N)$ 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:

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

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

$$\hat{\Sigma} := \sum_{i=1}^N \hat{\lambda}_i(\hat{\boldsymbol{\tau}}) \cdot \mathbf{u}_i \mathbf{u}_i', \quad (2.2)$$

where $\hat{\lambda}_i(\hat{\boldsymbol{\tau}})$ for $i = 1, \dots, N$ denote the shrunk eigenvalues based on $\hat{\boldsymbol{\tau}}$.

On the other hand, to capture the volatility-clustering feature of asset returns, [Engle \(2002\)](#) use 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 $t = 1, \dots, T$. Let $D_t := \operatorname{diag}(\sigma_{11t}^{1/2} \dots \sigma_{NNt}^{1/2})$ denote the volatility matrix, $Q_t := (q_{ijt})$ the pseudo-correlation matrix, and $P_t := (\rho_{ijt})$ the correlation matrix, satisfying

$$P_t := \operatorname{diag}(q_{11t}^{-1/2} \dots q_{NNt}^{-1/2}) Q_t \operatorname{diag}(q_{11t}^{-1/2} \dots q_{NNt}^{-1/2}). \quad (2.3)$$

The DCC model is defined as

$$\Sigma_t = D_t P_t D_t. \quad (2.4)$$

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

$$\sigma_{ii,t}^2 = \sigma_{ii,0}^2 (1 - \alpha_i - \beta_i) + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{ii,t-1}^2, \quad (2.5)$$

and the pseudo-correlation matrix Q_t is specified as

$$Q_t = \bar{Q}(1 - \alpha - \beta) + \alpha s_{t-1} s_{t-1}' + \beta Q_{t-1}, \quad (2.6)$$

where α_i , β_i , α , and β are non-negative scalars satisfying $\alpha_i + \beta_i < 1$ for every $i \in \{1, 2, \dots, N\}$ 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. \(2017\)](#) 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., 2017](#)) 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 heteroscedasticity 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 Portfolio with Gross-Exposure Constraints

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

$$\begin{aligned} \min_{\mathbf{w}_t} \quad & \mathbf{w}_t' \hat{\Sigma}_t \mathbf{w}_t & (2.7) \\ \text{subject to} \quad & \mathbf{w}_t' \mathbf{1} = 1 \text{ and } \sum_{i=1}^N |w_{i,t}| \leq \gamma. \end{aligned}$$

The constraint $\sum_{i=1}^N |w_{i,t}| \leq \gamma$ could be expressed as $\|\mathbf{w}_t\|_1 \leq \gamma$. 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

$$L(\mathbf{w}_t, \mu, \lambda) = \mathbf{w}_t' \hat{\Sigma}_t \mathbf{w}_t - \mu(\mathbf{w}_t' \mathbf{1} - 1) - \lambda(\gamma - \|\mathbf{w}_t\|_1), \quad (2.8)$$

and let \mathbf{g}_t be the subgradient vector of $\|\mathbf{w}_t\|_1$. Then, for $w_{i,t} \neq 0$, the i -th element of \mathbf{g}_t is unique, i.e., $g_{i,t} = \text{sign}(w_{i,t})$; 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 partial constrained optimization problem (2.7) are

$$\begin{cases} 2\hat{\Sigma}_t \mathbf{w}_t - \mu \mathbf{1} + \lambda \mathbf{g}_t = 0, \\ \lambda(\gamma - \|\mathbf{w}_t\|_1) = 0, \quad \lambda \geq 0, \\ \|\mathbf{w}_t\|_1 \leq \gamma, \quad \mathbf{w}_t' \mathbf{1} - 1 = 0, \end{cases} \quad (2.9)$$

where $\mathbf{1}$ is the column vector of ones, λ and μ are Lagrange multipliers. Denote a solution to (2.9) as \mathbf{w}_t^* . The following result shows that constructing the partial constrained minimum variance portfolio from the DCC estimator $\hat{\Sigma}_t$ is equivalent to constructing a (unconstrained) minimum variance portfolio from a shrunk version of $\hat{\Sigma}_t$.

Theorem 1. (i) Let $\tilde{\Sigma}_{\gamma,t} = \hat{\Sigma}_t + \frac{1}{2}\lambda(\mathbf{g}_t^* \mathbf{1}' + \mathbf{1g}_t^{*\prime})$, where \mathbf{g}_t^* is the subgradient at \mathbf{w}_t^* , and λ is the Lagrange multiplier defined in (2.9). Then $\tilde{\Sigma}_{\gamma,t}$ is positive definite if $\hat{\Sigma}_t$ is positive definite DCC covariance matrix estimator.

(ii) The partial constrained portfolio optimization problem (2.7) is equivalent to the optimization problem

$$\min_{\mathbf{w}'_t \mathbf{1} = 1} \mathbf{w}'_t \tilde{\Sigma}_{\gamma,t} \mathbf{w}_t \quad (2.10)$$

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

2.3 Constructing Markowitz Mean-Variance Efficient Portfolios

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

$$\min_{\mathbf{w}_t} \mathbf{w}'_t \hat{\Sigma}_t \mathbf{w}_t \quad (2.11)$$

$$\text{subject to } \mathbf{w}'_t \mathbf{1} = 1, \quad (2.12)$$

$$\mathbf{w}'_t \mathbf{m}_t = b_t \text{ and} \quad (2.13)$$

$$\sum_{i=1}^N |w_{i,t}| \leq \gamma, \quad (2.14)$$

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

$$b_t = \mathbf{w}Q'_t \mathbf{m}_t \quad (2.15)$$

where $\mathbf{w}Q_t$ is the weight vector of quantile-based portfolios. Let $\{(1), (2), \dots, (N)\}$ be the permutation of $\{1, 2, \dots, N\}$ that results in descending order of scores for the signal \mathbf{m}_t . Then, $wQ_t^{(1)} = \dots = wQ_t^{(d)} := 1/d$ and $wQ_t^{(d+1)} = \dots = wQ_t^{(N)} := 0$, 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 (2.11) as $\mathbf{w}_{b,t}^*$, then we could obtain the following theorem similar to Theorem 1.

Theorem 2. The partial constrained portfolio optimization problem (2.11) is equivalent to the optimization problem

$$\min_{\mathbf{w}_t} \mathbf{w}'_t \tilde{\Sigma}_{\gamma,t} \mathbf{w}_t \quad (2.16)$$

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

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

2.4 Comparison with the Study of [Jagannathan and Ma \(2003\)](#)

[Jagannathan and Ma \(2003\)](#) explain why imposing the wrong constraints on large portfolios helps reduce the risk. The underlying mathematics connecting shrinkage to short-sales constraints are sophisticated and unexpected, yet profound and compelling. While some people may have read this connection as confirmation that “shrinkage is not necessary”, their paper could equally be interpreted as saying that shrinkage works because *we* (meaning the asset management and finance professors who impose no-short-sales constraints) have been effectively shrinking all along without even knowing it. So why not get the best of both worlds?

First, [Jagannathan and Ma \(2003\)](#) only consider the constraint of nonnegative weights. To bridge the gap between the extreme no-short-sale constraint and no constraint, we study the effects of different degrees of leverage constraints on risk reduction by considering a wide range of gross-exposure constraints parameter γ . Specifically, we find valid empirical evidence supporting the 130/30 and 150/50 strategies.

Second, they compare the gain from imposing the nonnegative constraint with that from using the single-index linear shrinkage estimator proposed by [Ledoit and Wolf \(2003\)](#), and find the comparable effects supporting the conjecture that imposing the nonnegative constraint has a shrinkage-like effect. Given that [Ledoit and Wolf \(2012, 2015\)](#) extend the linear shrinkage to the nonlinear shrinkage, and [Engle et al. \(2017\)](#) advocate the use of the nonlinear shrinkage estimator in estimating the DCC model, both of which turn out to have improved out-of-sample performance, we focus on whether the gains from the nonlinear shrinkage estimator and the DCC-NL estimator exceed that from imposing various constraints on the portfolio weights.

Third, they construct the global minimum variance (GMV) portfolios but do not consider the Markowitz mean-variance efficient portfolios, as they believe that the sample mean is such an imprecise estimator of the population mean that based on which the mean-variance efficient portfolios constructed always perform poorly out of sample. To construct better Markowitz mean-variance efficient portfolios, we use the well-known momentum signal of [Jegadeesh and Titman \(1993\)](#) as a proxy for the population mean. For a given stock and a given investment date, the momentum signal m_{it} equals to the geometric average of the previous 12 monthly returns on the stock but excluding the most recent month (the signal is actually the geometric average return over 11 months, thus we denote it as 11-MM).

Finally, they construct portfolios with 500 stocks in their empirical research without considering the change of portfolio size. Actually, they find that the portfolio size and the sample size matter in determining the influence of no-short-sale constraints on the performance in their simulation study, while the portfolio size is set to be no greater than 300. To study whether and how much the effects of the shrinkage and the leverage constraints on risk reduction and efficiency increase would be influenced by the portfolio size in practice, we consider two cases in both the Monte Carlo simulation studies and the empirical study: $N = 500, 1000$.

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 emphasize that one of our most important improvements compared to [Jagannathan and Ma \(2003\)](#) is that 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. Since the essence of both the constraint on leverage and the shrinkage in covariance matrix estimation is the trade-off between the reduction in sampling error and the increase in specification error, the true covariance matrix matters in determining the performance. We therefore generate Monte Carlo data that can simulate the real data the best. The simulated data are generated as following:

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 (2.6), and with parameters $\alpha_i = 0.05$ and $\beta_i = 0.90$ for all individual stocks $i = 1, \dots, N$ for (2.5). 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 six different estimators for the covariance matrix, including the sample covariance matrix (S), the single-index linear shrinkage estimator (L) proposed by [Ledoit and Wolf \(2003\)](#), 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-L and DCC-NL estimator proposed by [Engle et al. \(2017\)](#). 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 (2.7).

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.

$$R_{orc} = \mathbf{w}'_t \Sigma_t \mathbf{w}_t \quad (3.1)$$

$$R_{emp} = \hat{\mathbf{w}}'_t \hat{\Sigma}_t \hat{\mathbf{w}}_t \quad (3.2)$$

$$R_{act} = \hat{\mathbf{w}}'_t \Sigma_t \hat{\mathbf{w}}_t \quad (3.3)$$

Figures 1-4 show R_{orc} , R_{emp} , and R_{act} for six 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 or the wrong constraints on weights, whereas the difference between the actual and the empirical risks is because of the estimation errors in the covariance matrix estimators.

The curve of the oracle risk shared by six 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 confirming the positive effects of the shrinkage and the gross-exposure constraints on reducing risks. However, in comparison with the consistent positive effects of the shrinkage, the constraints help reducing portfolio risks only under 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, closely followed by DCC-L estimator. 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 L, NL, DCC-L, 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 reducing 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 (DCC-L) 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 constraints with parameter $\gamma = 2$ or $\gamma = 1.6$ help reducing 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 constraints 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.

4 Empirical Results

4.1 Data

We examine the effects of the leverage constraint, the shrinkage in covariance matrix estimation, 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

data and the same portfolio-construction rules as in [Ledoit et al. \(2018\)](#), but we impose leverage 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 0.95¹. The daily return data we use, which covers the period from 01/01/1980 to 12/31/2015, are from the Center for Research in Security Prices (CRSP) database. The out-of-sample period is from 01/08/1986 to 12/31/2015.

For both the minimum variance portfolio and the efficient mean-variance portfolio, we consider two different portfolio sizes $N = 500, 1000$. The covariance matrix estimators we consider here include the sample covariance matrix (S), the nonlinear shrinkage estimator (NL) ([Ledoit and Wolf, 2012, 2015](#)), the covariance matrix estimator based on DCC model ([Engle, 2002](#)), and the DCC-NL estimator ([Engle et al., 2017](#)). 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.

Following [Ledoit et al. \(2018\)](#), we update the portfolios every 21 consecutive trading days, and thus form 360 investment dates from 01/01/1980 to 12/31/2015. At any investment date, the covariance matrix is estimated based on the most recent 1250 daily returns (roughly equals five years).

4.2 Main Results

Table 3 presents the out-of-sample performance of the GMV portfolios with gross-exposure parameter $\gamma = \infty, 2, 1.6, \text{ and } 1$. On the one hand, judging by the out-of-sample standard deviations of the GMV portfolios, which are shown in the fourth columns of Table 3, DCC-NL performs the best among four covariance matrix estimators considered. The outperformance is most obvious when $N = 1000$ and no leverage constraint is imposed: it reduces the out-of-sample standard deviation by 5.08 percentage points compared to the sample covariance matrix. As the leverage constraint becomes tighter, the relative outperformance of the NL estimator declines, while that of the DCC improves. The decline in the outperformance of NL estimator is consistent with the observations in [Jagannathan and Ma \(2003\)](#). According to them, imposing the nonnegative constraints on portfolio weights has a shrinkage-like effect, and thus it hurts the performance of the NL estimator.

On the other hand, the moderate constraints with $\gamma = 2$, or $\gamma = 1.6$ consistently outperform the extreme no-short-sale constraint with $\gamma = 1$ as well as the no constraint with $\gamma = \infty$, but the degrees of outperformance vary with the change of covariance matrix estimators. 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 4.57 and 2.07 percentage points compared to the no-constraint strategy and the no-short-sale strategy, respectively, while the numbers reduce

¹The 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.

to 0.37 and 0.54, respectively, if the DCC-NL estimator is used. Nevertheless, the DCC-NL estimator is always preferred: it leads to the smallest out-of-sample standard deviation even when an appropriate leverage constraint is imposed.

Interestingly, the effect of using DCC-NL is comparable to 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 $\gamma = 2$ have annualized out-of-sample standard deviations of 8.28 percent and 8.18 percent, respectively, both reducing the 13.36 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 of 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 $\gamma = 2$ (the 150/50 strategy) for portfolios with 500 stocks and is $\gamma = 1.5$ (the 130/30 strategy) for portfolios with 1000 stocks, where in both cases the DCC-NL estimator is suggested.

To sum up, the DCC-NL estimator achieves the best performance in all cases no matter whether a leverage constraint is imposed, and imposing the 30% to 50% constraint on leverage also improves the performance. The out-of-sample standard deviations of all GMV portfolios decrease in N , except the cases where neither the shrinkage is used nor the leverage constraint is imposed. When $N = 1000$, the DCC-NL estimator is used for covariance matrix, and the 30% leverage constraint is imposed, the portfolio has the smallest out-of-sample standard deviation among all GMV portfolios constructed.

Table 4 presents results for the Markowitz portfolios constructed based on the 11-MM momentum signal (Jegadeesh and Titman, 1993). Those portfolios have consistently larger out-of-sample Sharpe ratios than the corresponding global minimum variance portfolios, thus supporting the effectiveness of the momentum signal as a proxy for the population mean.

Judging by the Sharpe ratios, the 130/30 and 150/50 strategies still outperform the strategy without any constraint on weights or with the extreme constraint that no short-sale is allowed, and the DCC-NL estimator still performs the best among all covariance matrix estimators considering the robustness of performance. When $N = 1000$ and no constraint imposed on weights, the Markowitz portfolio constructed using the DCC-NL estimator has a Sharpe ratio of 1.12, almost double that from using the sample covariance matrix. Furthermore, it is not difficult to see that the effect of directly using the nonlinear shrinkage estimator approximates 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, whose performance can be imitated by using the DCC estimator and imposing an appropriate leverage constraint at the same time. These findings again support our conjecture that imposing appropriate 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.

4.3 Portfolio Weights

During 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 360 investment periods from 01/01/1980 to 12/31/2015 for the four characteristics of portfolio weights in Tables 3-4.

We find that portfolios constructed based on the sample covariance matrix turn out to have the smallest minimum weight and the largest total short position, while portfolios constructed based on the DCC estimator always have the largest maximum weight, and portfolios constructed based on the NL estimator always have 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 method is used in the estimation of the covariance matrix. For example, when $N = 1000$, the total short position of GMV portfolio reaches 327.64 percent if the DCC estimator is used, and this number even comes up to 525.34 if the sample covariance matrix is used. For Markowitz portfolios, the corresponding short positions are even larger, with 425.38 percent for the DCC estimator and 594.48 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 momentum signal. Instead of using the 11-months momentum, now we follow [Gettleman and Marks \(2006\)](#) and focus on the change in the 6-months momentum (denoted by $\Delta 6\text{-MM}$), where the 6-months momentum is the geometric average return reversal over the previous seven months but excluding the most recent month ([Jegadeesh and Titman, 1993](#)).

Table 5 presents the results for the Markowitz mean-variance efficient portfolios constructed based on the $\Delta 6\text{-MM}$ momentum signal, from which we draw similar conclusions to our main findings aforementioned. First, the DCC-NL estimator performs the best among all estimators we considered. Second, the moderate constraints with $\gamma = 2$ and $\gamma = 1.6$ lead to much better out-of-sample performance than no constraint or the extreme constraint that no short-sale is allowed. Third, the effects of using NL estimator and imposing the 30% or 50% leverage constraint are very close. When $N = 1000$, both the use of NL estimator and the imposing of the 50% leverage constraint raise the Sharpe ratio of the portfolio constructed from 0.7 to 1. 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: both the NL estimator and the DCC-NL estimator

increase the Sharpe ratio by more than 40 percent compared to the sample covariance matrix.

4.4.2 Transaction Cost

Transaction cost is an important issue in practical implementations. In Table 6, we present results of the Markowitz mean-variance efficient portfolios constructed based on the 11-MM momentum signal, when the transaction cost is considered. Referring to studies with practical background (e.g., [Credit Suisse Trading Strategy \(2013\)](#); [Deutsche Bank Market Research \(2015\)](#)) and literature from academic views (e.g., [Mei et al. \(2016\)](#); [Mei and Nogales \(2018\)](#)), we set the bid-ask spread to be three or five basis-points to embody the transaction cost.

Unsurprisingly, the Sharpe ratio becomes lower with the increase of transaction costs. Assuming a bid-ask spread of three basis-points, the Sharpe ratio reaches 1 only when $N = 1000$, 30% or 50% short-sale constraint is imposed, and DCC or DCC-NL estimator is used at the same time. If the bid-ask spread is set to five basis-points, the maximum value of Sharpe ratio is 0.93, which only occurs when $N = 1000$, 30% short-sale constraint is imposed, and DCC-NL estimator is used at the same time.

Features suggested by the pattern of Sharpe 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, and the 30% ($\gamma = 1.6$) and 50% ($\gamma = 2$) short-sale constraints are appropriate. The advantage of the DCC-NL estimator is most remarkable when no leverage constraint is imposed. For example, if the bid-ask spread is 3 basis-points, when $N = 500$ and no constraints imposed, using the DCC-NL estimator increases the Sharpe ratio from using S estimator by nearly 50% to 0.83. When appropriate constraints are imposed on portfolio weights, the DCC-NL estimator still helps increasing the Sharpe 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.

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. [Jagannathan and Ma \(2003\)](#) 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 their research by considering 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 single-index linear shrinkage estimator considered in [Jagannathan and Ma \(2003\)](#).

We provide theoretical insights into the similar effects of imposing the gross-exposure

constraint and using the shrinkage covariance matrix estimator on reducing risks of portfolios constructed. Moreover, we demonstrate through Monte Carlo simulations and empirical studies that portfolios constructed based on the DCC-NL estimator are well diversified, and the 30% to 50% leverage constraint is appropriate, which turns out to be better than the no-short-sale constraint as well as the no constraint. The good out-of-sample performance of the DCC-NL estimator are 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. The leverage constraint also has a shrinkage-like effect, but it is unlikely to hurt the performance of the shrinkage estimators as long as the degree of the constraint is moderate.

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 momentum signals with portfolio sizes $N = 500$ and 1000 . The empirical results show that though imposing an appropriate leverage constraint can achieve a similar effect as using the nonlinear shrinkage estimator, it always fall behind using the DCC-NL estimator, which considers both reducing the sampling errors and capturing the conditional heteroscedasticity. The effects of DCC and NL both increase in portfolio size, but the latter decreases as the leverage constraint becomes tighter. Moreover, we find that using DCC-NL estimator improves the out-of-sample performance even when the 30% or 50% leverage constraint is imposed. Overall, using the NL estimator, the DCC model, and an appropriate leverage constraint all help improving the out-of-sample performance, but only NL and leverage constraints help reducing the standard deviation of weights and the turnover of portfolios.

In our main study, we use the geometric average of the previous 12 monthly returns on the stock but excluding the most recent month as a proxy for the mean return to construct the mean-variance efficient portfolio. For robustness check, we consider the change in the 6 months momentum 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.

A Theoretical Justifications

Proof of Theorem 1 (i). Note that the matrix $\tilde{\Sigma}_{\gamma,t}$ is obviously symmetric and the solution to problem (2.7) is denoted by \mathbf{w}_t^* . For any vector x ,

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

Based on the KKT conditions in (2.9), $2\hat{\Sigma}_t \mathbf{w}_t^* - \mu \mathbf{1} + \lambda \mathbf{g}_t^* = 0$. Therefore,

$$\lambda (x' \mathbf{g}_t^*) (\mathbf{1}' x) = -2(x' \mathbf{1}) (x' \hat{\Sigma}_t \mathbf{w}_t^*) + \mu (x' \mathbf{1})^2. \quad (\text{A.2})$$

Note that

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

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

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

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

Hence,

$$|(x' \mathbf{1}) (x' \hat{\Sigma}_t \mathbf{w}_t^*)| \leq |x' \mathbf{1}| (x' \hat{\Sigma}_t x)^{\frac{1}{2}} \left(\frac{1}{2} \mu\right)^{\frac{1}{2}}. \quad (\text{A.3})$$

Combining (A.1)-(A.3), we have

$$\begin{aligned} x' \tilde{\Sigma}_{\gamma,t} x &= x' \hat{\Sigma}_t x - 2(x' \mathbf{1}) (x' \hat{\Sigma}_t \mathbf{w}_t^*) + \mu (x' \mathbf{1})^2 \\ &\geq x' \hat{\Sigma}_t x - 2|(x' \mathbf{1}) (x' \hat{\Sigma}_t \mathbf{w}_t^*)| + \mu (x' \mathbf{1})^2 \\ &\geq x' \hat{\Sigma}_t x - 2|x' \mathbf{1}| (x' \hat{\Sigma}_t x)^{\frac{1}{2}} \left(\frac{1}{2} \mu\right)^{\frac{1}{2}} + \mu (x' \mathbf{1})^2 \\ &= (a - b)^2 + b^2, \end{aligned} \quad (\text{A.4})$$

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

Moreover, $(a - b)^2 + b^2$ is always nonnegative and is zero if and only if $a = b$ and $b = 0$ hold simultaneously. However, $a = (x' \hat{\Sigma}_t x)^{\frac{1}{2}} > 0$ because $\hat{\Sigma}_t$ 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 (2.10) with equality constraint could be solved through the Lagrange multiplier method. Construct the Lagrangian

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

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

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

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

$$\mathbf{w}_t^{opt} = \frac{\tilde{\Sigma}_{\gamma,t}^{-1}\mathbf{1}}{\mathbf{1}'\tilde{\Sigma}_{\gamma,t}^{-1}\mathbf{1}}. \quad (\text{A.5})$$

By the Lagrange multiplier method, problem (2.7) is to minimize

$$L(\mathbf{w}_t, \mu, \lambda) = \mathbf{w}_t' \hat{\Sigma}_t \mathbf{w}_t - \mu(\mathbf{w}_t' \mathbf{1} - 1) - \lambda(c - \|\mathbf{w}_t\|_1).$$

Based on the fact that $\mathbf{g}_t^{*\prime} \mathbf{w}_t^* = \|\mathbf{w}_t^*\|_1$ and KKT conditions in (2.9), we have

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

Hence, the solution to problem (2.7) $\mathbf{w}_t^* = \frac{1}{2} (\lambda \gamma + \mu) \tilde{\Sigma}_{\gamma,t}^{-1} \mathbf{1}$. Moreover, because of the constraint $\mathbf{w}_t' \mathbf{1} = 1$, solving for $\lambda \gamma + \mu$ yields $\lambda \gamma + \mu = \frac{2}{\mathbf{1}' \tilde{\Sigma}_{\gamma,t}^{-1} \mathbf{1}}$. This fact indicates that

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

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 (2.16) with equality constraint could be solved through the Lagrange multiplier method. Construct the Lagrangian

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

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

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

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

The above equations also imply that

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

or

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

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

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

Therefore, the solution to this problem is given by

$$\mathbf{w}_{b,t}^{opt} = \tilde{\Sigma}_{\gamma,t}^{-1} (\mathbf{1}, \mathbf{m}_t) \left[(\mathbf{1}, \mathbf{m}_t)' \tilde{\Sigma}_{\gamma,t}^{-1} (\mathbf{1}, \mathbf{m}_t) \right] \begin{pmatrix} 1 \\ b_t \end{pmatrix}. \quad (\text{A.6})$$

By the Lagrange multiplier method, problem (2.11) is to minimize

$$L(\mathbf{w}_t, \mu_1, \mu_2, \lambda) = \mathbf{w}_t' \hat{\Sigma}_t \mathbf{w}_t - \mu_1 (\mathbf{w}_t' \mathbf{1} - 1) - \mu_2 (\mathbf{w}_t' \mathbf{m}_t - b_t) - \lambda (c - \|\mathbf{w}_t\|_1).$$

So, the KKT conditions are

$$\begin{cases} 2\hat{\Sigma}_t \mathbf{w}_t - \mu_1 \mathbf{1} - \mu_2 \mathbf{m}_t + \lambda \mathbf{g}_t = 0, \\ \lambda (c - \|\mathbf{w}_t\|_1) = 0, \quad \lambda \geq 0, \\ \|\mathbf{w}_t\|_1 \leq \gamma, \quad \mathbf{w}_t' \mathbf{1} - 1 = 0, \quad \mathbf{w}_t' \mathbf{m}_t - b_t = 0. \end{cases} \quad (\text{A.7})$$

Based on the fact that $\mathbf{g}_{b,t}^* \mathbf{w}_{b,t}^* = \|\mathbf{w}_{b,t}^*\|_1$ and KKT conditions in (A.7), we have

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

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

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

or

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

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

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

Hence,

$$\mathbf{w}_{b,t}^* = \tilde{\Sigma}_{\gamma,t}^{-1} (\mathbf{1}, \mathbf{m}_t) \left[(\mathbf{1}, \mathbf{m}_t)' \tilde{\Sigma}_{\gamma,t}^{-1} (\mathbf{1}, \mathbf{m}_t) \right] \begin{pmatrix} 1 \\ b_t \end{pmatrix}. \quad (\text{A.8})$$

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

B Tables and Figures

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

	$\hat{\Sigma}$	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
	L	4.98	1.21	172.41	3.88	0.59	170.19
	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-L	4.43	1.13	157.73	3.73	0.54	153.92
	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
	L	6.74	0.80	47.95	5.83	0.43	48.81
	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-L	5.10	0.72	45.91	4.57	0.40	47.76
	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
	L	7.47	0.77	29.35	6.80	0.42	29.60
	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-L	5.59	0.69	28.35	5.10	0.39	29.17
	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
	L	9.67	0.78	0.00	9.75	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-L	7.58	0.62	0.00	7.31	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 linear shrinkage estimator (L) (Ledoit and Wolf, 2003), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2012, 2015), the DCC estimator (Engle, 2002), the DCC-L estimator and the DCC-NL estimator (Engle et al., 2017). $\gamma = \infty, 2, 1.6, 1$ 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 Portfolio Weights.

		Panel A: $N = 500$			Panel B: $N = 1000$		
$\hat{\Sigma}$		StdR	StdW	ShortW	StdR	StdW	ShortW
$\gamma = \infty$	S	5.54	1.42	216.01	6.48	0.74	235.76
	L	4.97	1.18	169.42	3.81	0.58	168.62
	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-L	4.41	1.09	151.78	3.71	0.54	153.05
	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
	L	7.05	0.77	46.75	5.90	0.41	48.30
	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-L	4.98	0.70	44.14	4.42	0.38	46.79
	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
	L	8.13	0.74	28.77	6.88	0.41	29.32
	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-L	5.48	0.66	27.55	4.94	0.37	28.59
	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
	L	10.92	0.73	0.00	10.62	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-L	7.61	0.59	0.00	7.18	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 linear shrinkage estimator (L) (Ledoit and Wolf, 2003), the nonlinear shrinkage estimator (NL) (Ledoit and Wolf, 2012, 2015), the DCC estimator (Engle, 2002), the DCC-L estimator and the DCC-NL estimator (Engle et al., 2017). $\gamma = \infty, 2, 1.6, 1$ 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 Average Return, Standard Deviation, Sharpe Ratio, four Characteristics of Portfolio Weights and Average Turnover of the Global Minimum Variance Portfolios.

	$\hat{\Sigma}$	AvR	StdR	SR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
$\gamma = \infty$	S	6.74	10.82	0.62	-5.90	9.86	1.64	239.93	5.70
	NL	7.24	9.82	0.74	-2.71	4.77	0.95	131.52	2.21
	DCC	10.25	10.57	0.97	-4.05	15.52	1.55	172.78	3.97
	DCC-NL	9.51	9.66	0.98	-2.11	13.79	1.21	105.09	1.99
$\gamma = 2$	S	7.38	9.88	0.75	-3.27	9.76	1.01	50.00	2.93
	NL	7.34	9.86	0.74	-2.67	5.61	0.79	50.00	1.15
	DCC	9.18	9.82	0.93	-2.56	16.79	1.25	49.97	2.61
	DCC-NL	9.05	9.55	0.95	-1.92	15.45	1.17	49.74	1.26
$\gamma = 1.6$	S	7.31	10.09	0.72	-3.05	10.53	0.98	30.00	2.77
	NL	7.44	10.09	0.74	-2.72	6.30	0.77	30.00	1.17
	DCC	9.02	9.79	0.92	-2.27	18.03	1.27	30.00	2.49
	DCC-NL	8.97	9.64	0.93	-1.82	17.02	1.22	30.00	1.18
$\gamma = 1$	S	7.73	11.48	0.67	0.00	12.75	1.02	0.00	1.54
	NL	8.04	11.48	0.70	0.00	8.00	0.80	0.00	0.36
	DCC	8.62	10.36	0.83	0.00	23.48	1.47	0.00	1.27
	DCC-NL	8.43	10.32	0.82	0.00	23.18	1.46	0.00	0.20
Panel B: 1000 stocks contained in the portfolio									
$\gamma = \infty$	S	7.37	13.36	0.55	-6.85	9.19	1.59	525.34	11.41
	NL	7.38	8.80	0.84	-1.34	2.37	0.46	132.93	3.09
	DCC	8.20	10.62	0.77	-4.28	17.83	1.31	327.64	7.23
	DCC-NL	8.27	8.28	1.00	-1.11	14.31	0.73	94.70	2.68
$\gamma = 2$	S	7.64	8.79	0.87	-2.21	8.03	0.61	50.00	2.93
	NL	7.32	8.87	0.83	-1.57	3.18	0.40	50.00	1.42
	DCC	8.56	8.18	1.05	-1.67	19.63	0.88	49.99	2.75
	DCC-NL	8.16	7.91	1.03	-1.08	16.45	0.76	49.71	1.52
$\gamma = 1.6$	S	7.67	9.10	0.84	-2.00	8.70	0.60	30.00	2.75
	NL	7.60	9.19	0.83	-1.72	3.66	0.41	30.00	1.41
	DCC	8.47	8.08	1.05	-1.46	21.23	0.91	30.00	2.61
	DCC-NL	8.15	7.91	1.03	-1.04	18.76	0.83	29.99	1.35
$\gamma = 1$	S	9.28	10.86	0.85	0.00	11.29	0.66	0.00	1.56
	NL	9.21	10.91	0.84	0.00	5.35	0.45	0.00	0.55
	DCC	7.79	8.53	0.91	0.00	29.03	1.16	0.00	1.39
	DCC-NL	7.47	8.45	0.88	0.00	28.29	1.14	0.00	0.23

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, 2012, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2017). $\gamma = \infty, 2, 1.6, 1$ 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 Sharpe ratios (SR). 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 Sharpe ratios. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively.

Table 4: Out-of-sample Average Return, Standard Deviation, Sharpe Ratio, four Characteristics of Portfolio Weights and Average Turnover of the Markowitz Portfolios constructed based on the Momentum Signal (11-MM).

	$\hat{\Sigma}$	AvR	StdR	SR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
$\gamma = \infty$	S	9.65	12.77	0.76	-6.43	10.29	1.79	269.74	6.20
	NL	10.94	11.81	0.93	-2.98	4.92	1.05	152.84	2.56
	DCC	12.96	12.44	1.04	-4.96	16.62	1.77	214.38	4.59
	DCC-NL	12.65	11.43	1.11	-2.68	14.68	1.36	135.40	2.44
$\gamma = 2$	S	11.45	12.68	0.90	-3.59	9.97	1.06	50.00	2.90
	NL	11.85	12.64	0.94	-3.16	6.07	0.86	50.00	1.11
	DCC	12.90	12.27	1.05	-3.03	18.57	1.37	50.00	2.49
	DCC-NL	12.86	12.01	1.07	-2.52	17.55	1.31	49.95	1.20
$\gamma = 1.6$	S	11.46	13.49	0.85	-3.56	10.78	1.05	30.00	2.73
	NL	11.66	13.49	0.86	-3.43	6.97	0.87	30.00	1.12
	DCC	12.79	12.82	1.00	-2.86	19.86	1.39	30.00	2.35
	DCC-NL	12.57	12.67	0.99	-2.53	19.30	1.35	30.00	1.14
$\gamma = 1$	S	12.37	17.52	0.71	0.00	14.12	1.13	0.00	1.47
	NL	12.77	17.53	0.73	0.00	10.78	0.98	0.00	0.26
	DCC	12.17	16.11	0.76	0.00	22.46	1.45	0.00	1.12
	DCC-NL	11.84	16.08	0.74	0.00	22.45	1.46	0.00	0.18
Panel B: 1000 stocks contained in the portfolio									
$\gamma = \infty$	S	9.28	15.41	0.60	-7.73	9.89	1.78	594.48	12.75
	NL	10.57	10.39	1.02	-1.49	2.42	0.51	150.55	3.44
	DCC	10.41	13.12	0.79	-5.65	20.63	1.60	425.38	9.15
	DCC-NL	10.75	9.63	1.12	-1.44	15.65	0.82	122.08	3.14
$\gamma = 2$	S	11.23	10.99	1.02	-2.40	8.38	0.63	50.00	2.89
	NL	11.06	11.03	1.00	-1.92	3.47	0.44	50.00	1.35
	DCC	12.13	9.90	1.23	-1.92	21.82	0.97	50.00	2.65
	DCC-NL	11.51	9.63	1.20	-1.42	19.27	0.87	49.95	1.42
$\gamma = 1.6$	S	11.73	11.82	0.99	-2.35	9.11	0.63	30.00	2.70
	NL	11.62	11.90	0.98	-2.19	4.15	0.46	30.00	1.33
	DCC	12.37	10.26	1.21	-1.72	23.42	1.00	30.00	2.50
	DCC-NL	11.73	10.08	1.16	-1.44	21.93	0.95	30.00	1.26
$\gamma = 1$	S	13.07	16.12	0.81	0.00	11.82	0.70	0.00	1.49
	NL	12.86	16.11	0.80	0.00	7.24	0.55	0.00	0.41
	DCC	12.25	13.52	0.91	0.00	24.91	1.05	0.00	1.24
	DCC-NL	11.94	13.41	0.89	0.00	25.01	1.06	0.00	0.21

Notes: This table shows the empirical results for the Markowitz Portfolios constructed based on the Momentum Signal (11-MM), 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, 2012, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2017). $\gamma = \infty, 2, 1.6, 1$ 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 Sharpe ratios (SR). 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 Sharpe ratios. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively.

Table 5: Out-of-sample Average Return, Standard Deviation, Sharpe Ratio, four Characteristics of Portfolio Weights and Average Turnover of the Markowitz Portfolios constructed based on an Alternative Momentum Signal ($\Delta 6$ -MM).

	$\hat{\Sigma}$	AvR	StdR	SR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
$\gamma = \infty$	S	8.55	11.60	0.74	-6.33	10.25	1.76	262.57	6.31
	NL	9.08	10.54	0.86	-2.93	4.91	1.04	149.60	2.49
	DCC	12.01	11.81	1.02	-4.83	15.66	1.70	208.15	4.48
	DCC-NL	11.82	10.70	1.10	-2.72	13.86	1.32	133.18	2.39
$\gamma = 2$	S	9.58	11.31	0.85	-3.71	10.24	1.08	50.00	3.28
	NL	9.80	11.27	0.87	-3.13	6.17	0.87	50.00	1.12
	DCC	11.53	11.34	1.02	-3.10	17.09	1.32	50.00	2.49
	DCC-NL	11.82	11.07	1.07	-2.63	16.21	1.26	49.94	1.20
$\gamma = 1.6$	S	9.50	12.12	0.78	-3.46	11.14	1.07	30.00	3.10
	NL	10.00	12.10	0.83	-3.27	7.11	0.87	30.00	1.13
	DCC	11.37	11.86	0.96	-2.95	18.31	1.33	30.00	2.34
	DCC-NL	11.56	11.70	0.99	-2.67	17.75	1.30	30.00	1.13
$\gamma = 1$	S	9.76	15.98	0.61	0.00	15.12	1.18	0.00	1.87
	NL	10.03	16.00	0.63	0.00	11.23	1.01	0.00	0.28
	DCC	11.59	15.12	0.77	0.00	21.19	1.41	0.00	1.07
	DCC-NL	11.23	15.08	0.75	0.00	21.20	1.41	0.00	0.17
Panel B: 1000 stocks contained in the portfolio									
$\gamma = \infty$	S	10.28	14.73	0.70	-7.39	9.58	1.70	566.20	12.33
	NL	9.48	9.51	1.00	-1.49	2.42	0.50	149.18	3.37
	DCC	10.33	13.32	0.78	-5.84	19.13	1.55	424.24	9.11
	DCC-NL	10.14	9.39	1.08	-1.56	14.79	0.79	124.47	3.18
$\gamma = 2$	S	10.19	10.12	1.01	-2.47	8.34	0.64	50.00	3.27
	NL	9.64	10.15	0.95	-1.92	3.53	0.45	50.00	1.35
	DCC	10.34	9.78	1.06	-2.03	20.47	0.93	50.00	2.65
	DCC-NL	10.28	9.40	1.09	-1.57	18.23	0.85	49.94	1.41
$\gamma = 1.6$	S	10.56	10.96	0.96	-2.35	9.15	0.64	30.00	3.10
	NL	10.42	11.04	0.94	-2.15	4.24	0.46	30.00	1.33
	DCC	10.46	10.12	1.03	-1.89	21.94	0.96	30.00	2.49
	DCC-NL	10.24	9.89	1.03	-1.64	20.51	0.91	30.00	1.25
$\gamma = 1$	S	11.60	14.92	0.78	0.00	12.71	0.73	0.00	1.87
	NL	11.67	15.01	0.78	0.00	7.74	0.56	0.00	0.42
	DCC	10.63	13.35	0.80	0.00	22.84	1.00	0.00	1.20
	DCC-NL	10.55	13.28	0.79	0.00	22.71	1.00	0.00	0.20

Notes: This table shows the empirical results for the Markowitz Portfolios constructed based on an alternative Momentum Signal ($\Delta 6$ -MM), 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, 2012, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2017). $\gamma = \infty, 2, 1.6, 1$ 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 Sharpe ratios (SR). 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 Sharpe ratios. Panel A and Panel B show results for portfolios with 500 and 1000 stocks, respectively.

Table 6: Out-of-sample Average Return, Standard Deviation, and Sharpe Ratio of the Markowitz Portfolios constructed based on the Momentum Signal (11-MM) with Transaction Costs.

	$\hat{\Sigma}$	Spread = 3 basis-points			Spread = 5 basis-points		
		AvR	StdR	SR	AvR	StdR	SR
Panel A: 500 stocks contained in the portfolio							
$\gamma = \infty$	S	7.42	12.77	0.58	5.93	12.79	0.46
	NL	7.78	11.82	0.66	5.68	11.87	0.48
	DCC	9.08	12.47	0.73	6.49	12.54	0.52
	DCC-NL	9.54	11.45	0.83	7.47	11.50	0.65
$\gamma = 2$	S	10.40	12.67	0.82	9.70	12.67	0.77
	NL	10.40	12.63	0.82	9.44	12.64	0.75
	DCC	10.96	12.27	0.89	9.67	12.28	0.79
	DCC-NL	11.38	12.01	0.95	10.39	12.01	0.87
$\gamma = 1.6$	S	10.48	13.48	0.78	9.82	13.48	0.73
	NL	10.27	13.49	0.76	9.35	13.49	0.69
	DCC	10.96	12.82	0.86	9.74	12.83	0.76
	DCC-NL	11.18	12.66	0.88	10.25	12.67	0.81
$\gamma = 1$	S	11.84	17.52	0.68	11.48	17.51	0.66
	NL	12.15	17.53	0.69	11.73	17.52	0.67
	DCC	11.23	16.11	0.70	10.61	16.10	0.66
	DCC-NL	11.24	16.07	0.70	10.85	16.07	0.67
Panel B: 1000 stocks contained in the portfolio							
$\gamma = \infty$	S	4.68	15.44	0.30	1.62	15.53	0.10
	NL	4.74	10.49	0.45	0.85	10.70	0.08
	DCC	2.53	13.27	0.19	-2.71	13.57	-0.20
	DCC-NL	5.02	9.72	0.52	1.21	9.94	0.12
$\gamma = 2$	S	10.19	10.98	0.93	9.49	10.99	0.86
	NL	9.54	11.03	0.86	8.52	11.04	0.77
	DCC	10.14	9.91	1.02	8.81	9.93	0.89
	DCC-NL	9.96	9.63	1.03	8.93	9.64	0.93
$\gamma = 1.6$	S	10.75	11.82	0.91	10.10	11.82	0.86
	NL	10.17	11.90	0.85	9.20	11.91	0.77
	DCC	10.49	10.26	1.02	9.24	10.27	0.90
	DCC-NL	10.30	10.08	1.02	9.35	10.08	0.93
$\gamma = 1$	S	12.54	16.12	0.78	12.18	16.11	0.76
	NL	12.18	16.10	0.76	11.72	16.10	0.73
	DCC	11.26	13.51	0.83	10.61	13.51	0.79
	DCC-NL	11.33	13.40	0.85	10.92	13.40	0.81

Notes: This table shows the empirical results for the Markowitz Portfolios constructed based on the Momentum Signal (11-MM), 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, 2012, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle et al., 2017). $\gamma = \infty, 2, 1.6, 1$ 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 Sharpe ratios (SR). 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.

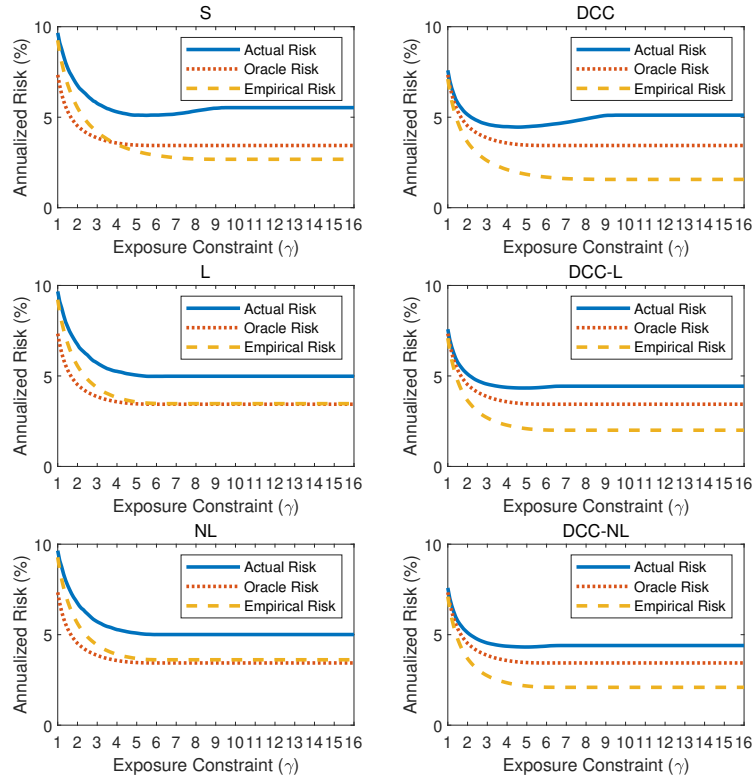


Figure 1: Comparisons of the oracle theoretical optimal risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, L, NL, DCC, DCC-L, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ). $N = 500$, and the disturbance terms of the simulated data are drawn from a multivariate standard normal distribution.

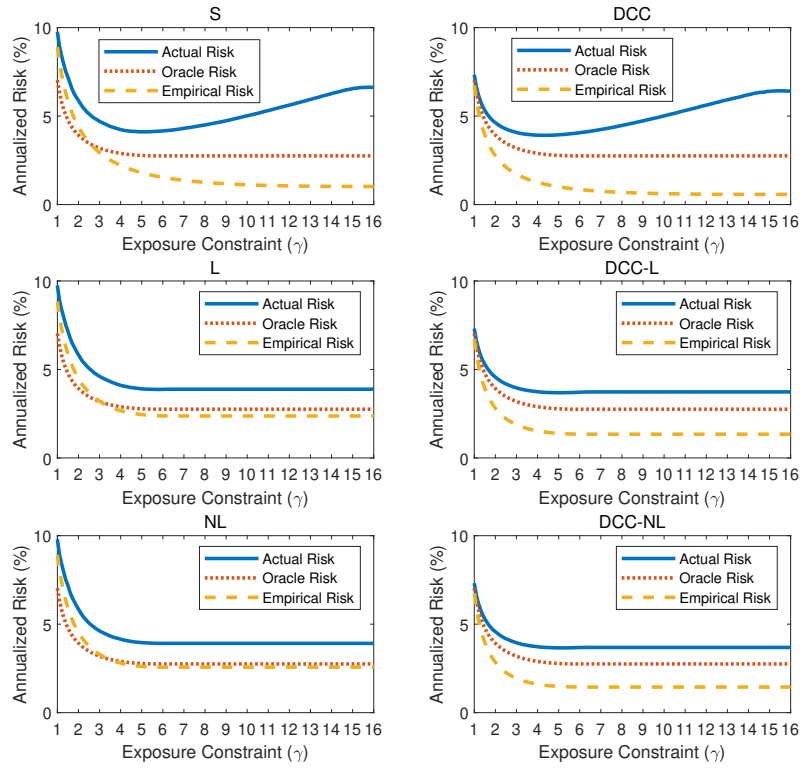


Figure 2: Comparisons of the oracle theoretical optimal risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, L, NL, DCC, DCC-L, DCC-NL) and facing various degrees of leverage constraints (the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

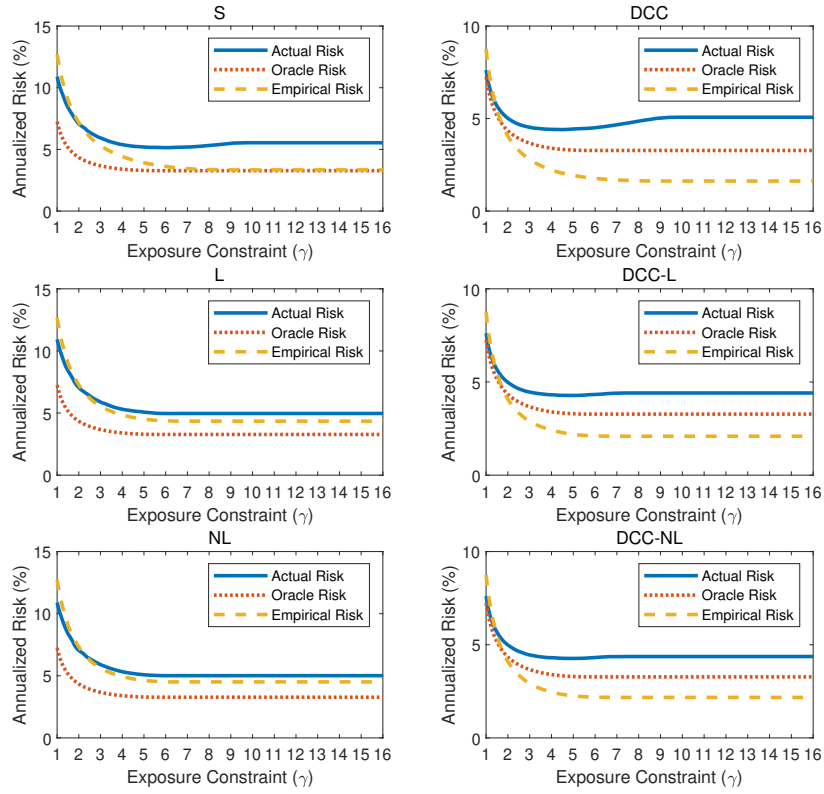


Figure 3: Comparisons of the oracle theoretical optimal risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, L, NL, DCC, DCC-L, DCC-NL) and facing various degrees of leverage constraints (the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

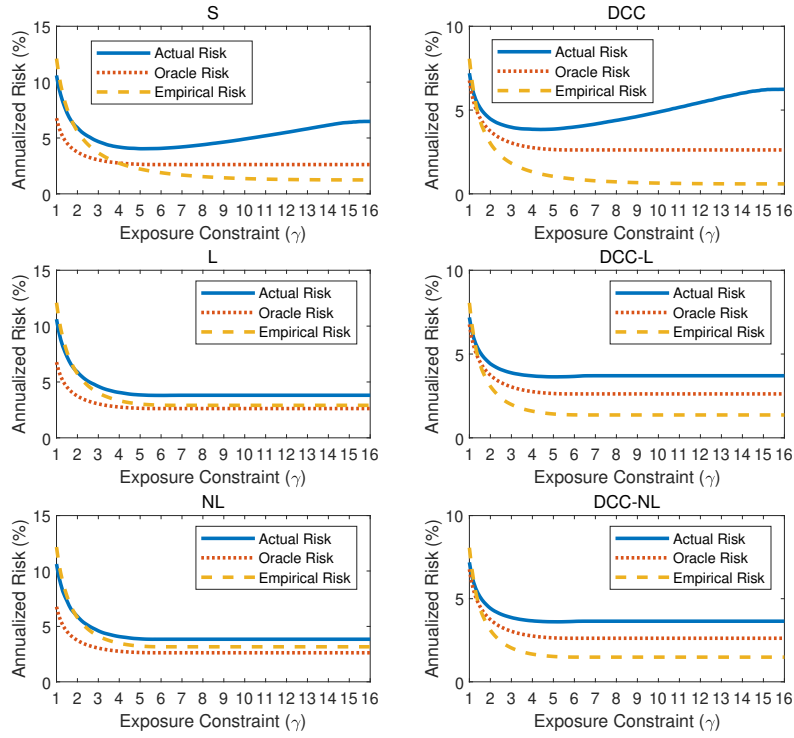


Figure 4: Comparisons of the oracle theoretical optimal risks, the average actual risks, and the average empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, L, NL, DCC, DCC-L, DCC-NL) and facing various degrees of leverage constraints (the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

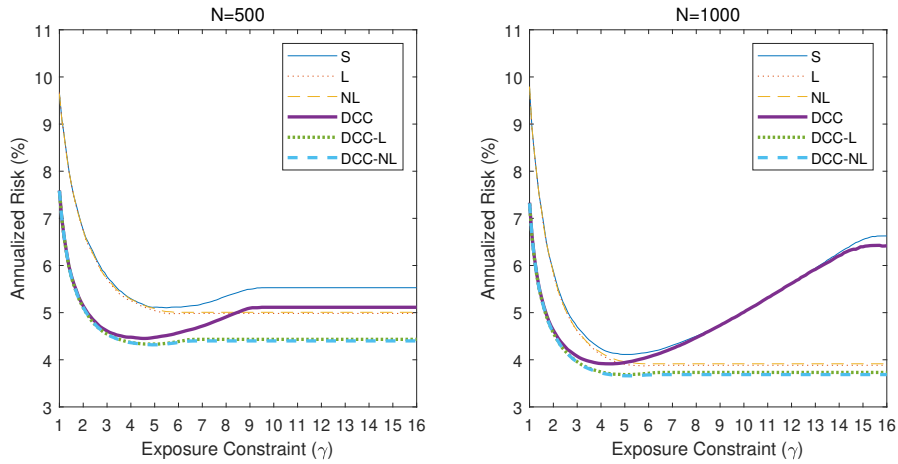


Figure 5: Comparisons of the average actual risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, L, NL, DCC, DCC-L, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter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

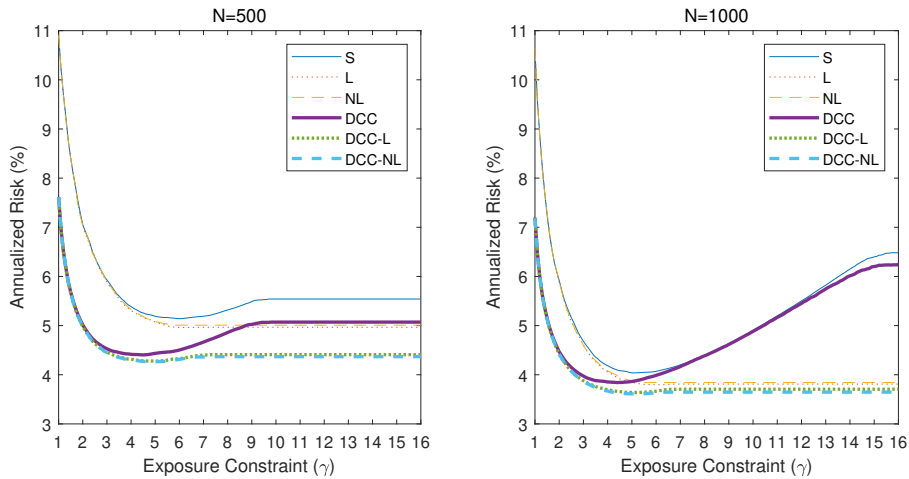


Figure 6: Comparisons of the average actual risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, L, NL, DCC, DCC-L, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter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

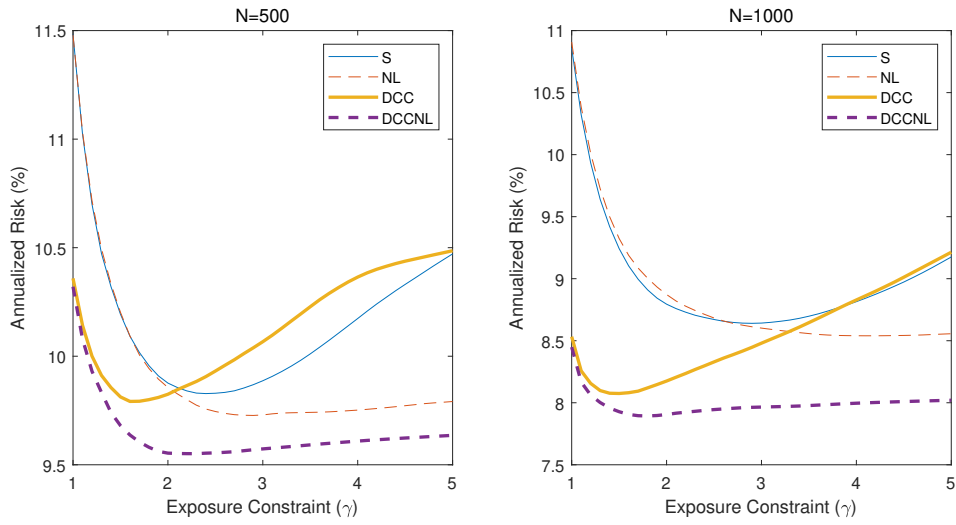


Figure 7: Comparisons of the out-of-sample risks of the GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of leverage constraints (the intensity of the constraint declines with the increase of parameter γ).

References

- Credit Suisse Trading Strategy (2013). Inside the NBBO: Pushing for wider — and narrower! — spreads. Published by Ana Avramovic and Phil Mackintosh on May 15, 2013.
- Demiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.
- Deutsche Bank Market Research (2015). A portfolio manager’s guidebook to trade execution. Published by US Quantitative Strategist Kevin Webster on July 8, 2015.
- Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.
- Engle, R. F., Ledoit, O., and Wolf, M. (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*. doi: 0.1080/07350015.2017.1345683.
- Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.
- Gettleman, E. and Marks, J. M. (2006). Acceleration strategies. *SSRN Working Paper Series*.
- Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 54(4):1651–1684.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.
- Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.
- Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4):110–119.
- Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.
- Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.
- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139(2):360–384.
- Ledoit, O. and Wolf, M. (2017). Numerical implementation of the QuEST function. *Computational Statistics & Data Analysis*, 115:199–223.

- Ledoit, O., Wolf, M., and Zhao, Z. (2018). Efficient sorting: A more powerful test for cross-sectional anomalies. *Journal of Financial Econometrics*.
- Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.
- Mei, X., DeMiguel, V., and Nogales, F. J. (2016). Multiperiod portfolio optimization with multiple risky assets and general transaction costs. *Journal of Banking & Finance*, 69:108–120.
- Mei, X. and Nogales, F. J. (2018). Portfolio selection with proportional transaction costs and predictability. *Journal of Banking & Finance*, 94:131–151.
- Pakel, C., Shephard, N., Sheppard, K., and Engle, R. F. (2017). Fitting vast dimensional time-varying covariance models. Working Paper FIN-08-009, NYU. Available at <https://ssrn.com/abstract=1354497>.
- Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science*, 9(1):277–293.