Journal of Financial Econometrics, 2023, Vol. 21, No. 1, 73–105
doi: 10.1093/jjfinec/nbab001
Advance Access Publication Date: 26 February 2021
Article

Risk Reduction and Efficiency Increase in Large Portfolios: Gross-Exposure Constraints and Shrinkage of the Covariance Matrix*

Zhao Zhao[1], Olivier Ledoit[2] and Hui Jiang [3]

[1]School of Economics, Huazhong University of Science and Technology, [2]University of Zurich and AlphaCrest Capital Management and [3]School of Mathematics and Statistics, Huazhong University of Science and Technology

Address correspondence to Hui Jiang, School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, Hubei, China, or e-mail: jianghui@hust.edu.cn.

Received July 5, 2020; revised January 2, 2021; editorial decision January 6, 2021

Abstract

We investigate the effects of constraining gross-exposure 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 versus larger universe of stocks, alternative portfolio formation objectives (global minimum variance versus exposure to profitable factors), and various transaction cost assumptions, we find that a judiciously chosen shrinkage method always outperforms an arbitrarily determined constraint on gross-exposure. We extend the mathematical connection between constraints on the gross-exposure and shrinkage of the covariance matrix from static to dynamic, and provide a new explanation for our finding from the perspective of degrees of freedom. In addition, both simulation and empirical analysis show that the dynamic conditional correlation-nonlinear shrinkage (DCC-NL) estimator results in risk reduction and efficiency increase in large portfolios as long as a small amount of short position is allowed, whereas imposing a constraint on gross-exposure often hurts a DCC-NL portfolio.

* We acknowledge the financial support from the National Natural Science Foundation of China (Grant Nos. 71803055, 11801197 and 72033002), and the Fundamental Research Funds for the Central Universities 2018KFYYXJJ040. We also thank Prof. Fabio Trojani (the Editor); two anonymous referees; and Prof. Michael Wolf, Prof. Yixiao Sun, and Dr. Gianluca De Nard for valuable comments and suggestions. Any errors are of authors.

Key words: DCC; gross-exposure constraint; large portfolios; mean–variance efficient; nonlinear shrinkage; risk reduction

JEL classification: C13, C58, G11

It is well established that using the 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 are very popular. The first one is to constrain (e.g., Jagannathan and Ma, 2003; DeMiguel et al., 2009a; Fan, Zhang, and Yu, 2012; Behr, Guettler, and Miebs, 2013; Li, 2015) or shrink (e.g., Golosnoy and Okhrin, 2007; Frahm and Memmel, 2010; Tu and Zhou, 2011; DeMiguel, Martin-Utrera, and Nogales, 2013; Bodnar, Parolya, and Schmid, 2018) the portfolio weights directly, and the second is to use an improved estimator of covariance matrix to constrain or shrink the portfolio weights indirectly (e.g., Ledoit and Wolf, 2003, 2017a; Fan, Fan, and Lv, 2008, Fan, Liao, and Mincheva, 2013; Engle, Ledoit, and Wolf, 2019; De Nard, Ledoit, and Wolf, 2020).

There is a general perception that constraining the portfolio weights 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. Although 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 constraints on portfolio weights) 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, that is, 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 up to zero). Even though the weights sum up to one, there is some leeway to take on short positions, and an interesting question is how much. The major candidates are the strategy without any constraint on gross-exposure, 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 are allowed). For a more solid analysis, we consider a wide choice for the constraint on short sales, with the gross-exposure constraint parameter continuously increasing from 1 (the "100/0" strategy) to 16 (the "850/750" strategy). Moreover, at the same time, instead of linearly shrinking the unconditional covariance matrix as the previous literature does, 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 constraining portfolio weights and shrinking the covariance matrix from the conventional static framework to dynamic, where multivariate GARCH effects are incorporated. Second, we find that using the judiciously chosen shrunk covariance matrix always outperforms the strategy of imposing an arbitrarily determined gross-exposure constraint. 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 the shrinkage estimator and the gross-exposure constraint, which can provide enlightenments to quantitative investors and analysts. Moreover, we find that the nonlinear (NL) shrinkage of the covariance matrix improves a large portfolio as long as a small amount of short position is allowed, but imposing a constraint on gross-exposure often hurts a covariance-shrunk portfolio.

If constraining weights and shrinking covariances are essentially equivalent, then why is shrinkage still beneficial for portfolios with moderate gross-exposure constraints? The answer is that constraining the gross-exposure has only one degree of freedom, that is, the amount of gross-exposure permitted; whereas shrinkage of the NL 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 gross-exposure constraint applies uniformly across the board.[1] The gross-exposure constraint, therefore, is redundant when the optimal shrinkage is provided by the NL shrinkage method.

Our Monte Carlo simulation and empirical analysis demonstrate the advantage of using the NL shrinkage covariance matrix estimator over imposing the gross-exposure constraints for a wide range of constrained portfolios. Additionally, we find that using the dynamic conditional correlation (DCC) model improves the portfolio performance. Consequently, the DCC-NL shrinkage (DCC-NL) portfolio is always preferred: as long as *some* short positions are allowed, which can be done in many fully invested funds as long as the prime broker allows it, the DCC-NL delivers the best performance. 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) portfolios and the mean–variance efficient (MVE) portfolios exposed to profitable factors 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 MVE portfolio). The first type is a *purer* test of the covariance matrix estimator and the optimization program used, whereas the second one requires, in addition, a good predictive model for expected returns, which is notoriously hard to obtain.

1 Behr, Guettler, and Miebs (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.

During the past decades, on the one hand, various attempts have been made to reduce the estimation errors of the large-dimensional covariance matrix. Imposing the assumption of a factor structure (e.g., Stock and Watson, 2002; Fan, Fan, and Lv, 2008; Bai and Li, 2012; Fan, Liao, and Mincheva, 2013) and shrinking the sample covariance matrix (Ledoit and Wolf, 2003, 2004a, 2004b, 2012, 2015, 2020a) are two popular and effective ways. The shrinkage method can also be used in a factor model to estimate the residual covariance matrix (De Nard, Ledoit, and Wolf, 2020). On the other hand, a variety of methods have been introduced in the literature to deal with the parameter uncertainty about expected returns, including Bayesian approaches (e.g., Pástor, 2000; Wang, 2005; Pástor and Stambaugh, 2009; Tu and Zhou, 2010; Bauder et al., 2020) and some other non-Bayesian strategies (e.g., Goldfarb and Iyengar, 2003; Garlappi, Uppal, and Wang, 2007; Brandt, Santa-Clara, and Valkanov, 2009; Branger, Lučivjanská, and Weissensteiner, 2019). Moreover, hundreds of signals have been proposed to predict expected returns (see Hou, Xue, and Zhang, 2015; Harvey, Liu, and Zhu, 2016; Green, Hand, and Zhang, 2017; and the references therein). Given that the two strands of literature have attracted a large amount of attention, we consider both exercises. Specifically, to construct the MVE portfolios, we use the signal return-on-equity (ROE) as a proxy for the expected return, which has been proven to have statistically significant explanatory power for cross-sectional anomalies (Feng, Giglio, and Xiu, 2020). For robustness check, we consider the widely used signal earnings-to-price (E/P) as an alternative proxy. We also consider different setups for transaction costs.

Our work is most closely related to the literature that attempts to provide a unified analysis framework for imposing constraints on weights and using advanced estimators for the covariance matrix. Jagannathan and Ma (2003) find with the no-short-sale constraint in place, GMV portfolios constructed based on the 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.

DeMiguel et al. (2009a) and Brodie et al. (2009) use the ℓ_2-norm or ℓ_1-norm framework to unify the shrinkage effects in weights and in covariance matrix. The ℓ_2-norm constrained portfolios, related to ridge regression which shrinks all regression coefficients toward zero and does not produce any sparsity, correspond to the shrinkage estimators in Ledoit and Wolf (2004b). The ℓ_1-norm constrained portfolios, related to LASSO regression which tends to give sparse weights, are equivalent to the short-sale-constrained portfolios in Jagannathan and Ma (2003). Yen (2016) imposes both ℓ_1-norm and ℓ_2-norm penalties to study portfolio optimization.

Equivalent unconstrained regression representation for portfolio optimization has also been investigated. Britten-Jones (1999) discusses the portfolio selection problem through an artificial linear regression. On this basis, Fan, Zhang, and Yu (2012) and Li (2015) show that constraining portfolio norms amounts to constraining estimation risks. In addition, Ao, Li, and Zheng (2019) proposes a strategy MAXSER building upon a novel unconstrained regression representation and can simultaneously achieve mean–variance efficiency and risk control. Furthermore, Callot et al. (2020) study the variance, weights, and risk of large portfolios by using nodewise regression to directly estimate the inverse covariance matrix.

Note that shrinkage estimators of the covariance matrix whose performances in portfolio selection have been compared with those constraining on portfolio weights are all linear, with the shrinkage target being the Sharpe (1963)'s single index covariance matrix proposed by Ledoit and Wolf (2003) (Jagannathan and Ma, 2003; DeMiguel et al., 2009a; 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 essentially equivalent to linearly shrinking the sample eigenvalues toward a more centralized set of eigenvalues by a unified shrinkage intensity. Ledoit and Wolf (2012, 2015) extend the linear shrinkage of the sample eigenvalues to the NL transformation, and obtain the NL estimator of the covariance matrix, which has been proven to have better out-of-sample performance (Ledoit and Wolf, 2015, 2017a). Ledoit and Wolf (2020b) summarize that the linear shrinkage is simpler to understand and to implement, but the NL shrinkage is more flexible and powerful. 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 gross-exposure constraint.

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

The key differences between our work and the previous literature are embodied in the following two aspects. First, we adopt the most advanced estimator for covariance matrix, which not only applies the NL shrinkage procedure to account for the individual shrinkage intensity, but also considers the dynamic variation of the covariances using the GARCH model. Second, we focus on the improvement of DCC-NL over DCC when some certain gross-exposure constraints are imposed. On this basis, we obtain new findings that the DCC-NL estimator is always preferred and using the NL shrinkage estimator is superior to imposing a gross-exposure constraint. Compared with imposing a gross-exposure constraint, using the NL shrinkage estimator has a larger degree of freedom, which gives enough room for the benefits of the automatic optimization procedure.

The rest of the paper is organized as follows. Section 1 provides the methodologies, including the NL shrinkage estimator of the covariance matrix, its combination with the DCC model, and our theoretical findings in constructing constrained GMV and MVE portfolios. In Section 2, we use Monte Carlo simulations to verify our theoretical results. In Section 3, we describe our data, report the results for out-of-sample performance of the GMV and the MVE portfolios, and conduct robustness checks. Section 4 concludes.

1 Methodology

1.1 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 toward the grand mean of all sample eigenvalues, since the smallest sample eigenvalues are biased downward and the largest ones upward. Ledoit and Wolf (2003, 2004a, 2004b) propose the linear shrinkage estimators, which are the first-order approximation solutions to a NL optimization problem, as all sample eigenvalues adjust with the same shrinkage intensity. The NL 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 Quantized Eigenvalues Sampling Transform (QuEST) function. By numerically inverting the QuEST function,[2] 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 $Q_{T,N}(\boldsymbol{t}) := (q^1_{T,N}(\boldsymbol{t}), \ldots, q^N_{T,N}(\boldsymbol{t}))'$ 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:

$$\widehat{\boldsymbol{\tau}} := \underset{\mathbf{t}\in[0,+\infty)^N}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N}\left(q^i_{T,N}(\mathbf{t}) - \lambda_i\right)^2 . \tag{1.1}$$

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

$$\hat{\Sigma} := \sum_{i=1}^{N} \widehat{\lambda}_i(\widehat{\boldsymbol{\tau}}) \cdot \boldsymbol{u}_i \boldsymbol{u}_i'. \tag{1.2}$$

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

$$\widehat{\lambda}_i(\widehat{\boldsymbol{\tau}}) \approx \boldsymbol{u}_i' \, \Sigma \boldsymbol{u}_i, \tag{1.3}$$

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). Equations (1.2) and (1.3) 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 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 $D_t := diag(\sigma^{1/2}_{11t} \ldots \sigma^{1/2}_{NNt})$

2 See Ledoit and Wolf (2017b) for the detailed implementation.

denote the volatility matrix, $Q_t := (q_{ijt})$ the pseudo-correlation matrix, and $P_t := (\rho_{ijt})$ the correlation matrix, satisfying

$$P_t := diag(q_{11t}^{-1/2} \dots q_{NNt}^{-1/2}) Q_t diag(q_{11t}^{-1/2} \dots q_{NNt}^{-1/2}). \tag{1.4}$$

The DCC model is defined as

$$\Sigma_t = D_t P_t D_t. \tag{1.5}$$

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

$$\sigma_{ii,t}^2 = \sigma_{ii,0}^2 (1 - \alpha_i - \beta_i) + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{ii,t-1}^2, \tag{1.6}$$

and the pseudo-correlation matrix Q_t is specified as

$$Q_t = \overline{Q}(1 - \alpha - \beta) + \alpha s_{t-1} s_{t-1}' + \beta Q_{t-1}, \tag{1.7}$$

where α_i, β_i, α, and β are non-negative scalars satisfying $\alpha_i + \beta_i < 1$ for every $i \in \{1, 2, \dots, N\}$ 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 return at time t, and $\overline{Q}$ is the long-run covariance matrix of s_t.

By combining the NL shrinkage estimator of $\overline{Q}$ with the DCC model, Engle, Ledoit, and Wolf (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 estimation 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.

Even though our paper uses (in part) the DCC-NL method of Engle, Ledoit, and Wolf (2019), our paper differs from theirs because they focus on the unconstrained case, whereas we explore the interaction of the DCC-NL (among others) with gross exposure constraints. This is more practically relevant because many investors face leverage constraints. At the start of the investigation, we cannot know a priori whether the benefits of DCC-NL observed by Engle, Ledoit, and Wolf (2019) carry over to the constrained case because applying a gross book exposure constraint may already induce a sufficient of (implicit) shrinkage to nullify the benefits of explicitly shrinking via DCC-NL or another shrinkage method. This is the question that we need to answer.

1.2 Constructing GMV Portfolios with Gross-Exposure Constraints

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

$$\begin{gathered} \min_{w_t} \boldsymbol{w}_t' \hat{\Sigma}_t \boldsymbol{w}_t \\ \text{subject to } \boldsymbol{w}_t' 1 = 1 \text{ and } \sum_{i=1}^{N} |w_{i,t}| \leq \gamma. \end{gathered} \tag{1.8}$$

The constraint $\sum_{i=1}^{N}|w_{i,t}| \le \gamma$ could be expressed as $||\boldsymbol{w}_t||_1 \le \gamma$. Note that $\gamma \ge 1$, 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 $\gamma = \infty$, the gross-exposure is unconstrained.

Define the Lagrangian as

$$L(\boldsymbol{w}_t, \mu, \lambda) = \boldsymbol{w}_t'\hat{\Sigma}_t\boldsymbol{w}_t - \mu(\boldsymbol{w}_t'1 - 1) - \lambda(\gamma - ||\boldsymbol{w}_t||_1), \tag{1.9}$$

and let $\mathbf{g}_t$ be the subgradient[3] vector of $||\boldsymbol{w}_t||_1$. Then, for $w_{i,t} \neq 0$, the i-th element of $\mathbf{g}_t$ is unique, that is, $g_{i,t} = \text{sign}(w_{i,t})$; 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 gross-exposure-constrained optimization problem (1.8) are

$$\begin{cases} 2\hat{\Sigma}_t\boldsymbol{w}_t - \mu 1 + \lambda\mathbf{g}_t = 0, \\ \lambda(\gamma - ||\boldsymbol{w}_t||_1) = 0, \ \lambda \ge 0, \\ ||\boldsymbol{w}_t||_1 \le \gamma, \ \boldsymbol{w}_t'1 - 1 = 0, \end{cases} \tag{1.10}$$

where $\mathbf{1}$ is the column vector of ones, and λ and μ are Lagrange multipliers. Denote a solution to Equation (1.10) by $\boldsymbol{w}_t^*$. The following result shows that constructing the gross-exposure-constrained minimum variance portfolio from the DCC estimator is equivalent to constructing a (unconstrained) minimum variance portfolio from a shrunk version of the DCC estimator.

Theorem 1.

i. *Let* $\tilde{\Sigma}_{\gamma,t} = \hat{\Sigma}_t + \frac{1}{2}\lambda(\mathbf{g}_t^*1' + 1\mathbf{g}_t^{*\prime})$, *where* $\mathbf{g}_t^*$ *is the subgradient at* $\boldsymbol{w}_t^*$, *and* λ *is the Lagrange multiplier defined inEquation (1.10). Then,* $\tilde{\Sigma}_{\gamma,t}$ *is positive definite if* $\hat{\Sigma}_t$ *is a positive definite DCC covariance matrix estimator.*
ii. *The partial constrained portfolio optimization problem (1.8) is equivalent to the optimization problem*

$$\min_{\mathbf{w}_t'1=1} \mathbf{w}_t'\tilde{\Sigma}_{\gamma,t}\mathbf{w}_t \tag{1.11}$$

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

It is noteworthy that the optimization problem (1.8) is solved by using the quadratic programming algorithm (Tibshirani, 1996). Even though we do not compute the theoretical regularized matrix $\tilde{\Sigma}_{\gamma,t}$ in empirical analysis, the equivalence established in Theorem 1 could explain the relation between the two kinds of optimization problems. The Lagrange multiplier λ could be viewed as the parameter that controls the amount of shrinkage, which illustrates the relation between $\tilde{\Sigma}_{\gamma,t}$ and the well-known shrinkage estimator of covariance matrix.

3 We say that a vector $\mathbf{g}$ is a subgradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $w_0 \in \mathbf{dom}\, f$ if for all $w \in \mathbf{dom}\, f$, $f(w) \ge f(w_0) + \mathbf{g}^\top(w - w_0)$. If f is convex and differentiable, then its gradient at w_0 is a subgradient. But a subgradient can exist even when f is not differentiable at w_0. A vector $\mathbf{g}$ is a subgradient of f at w_0 if the affine function (of $\boldsymbol{w}$) $f(w_0) + \mathbf{g}^\top(w - w_0)$ is a global underestimator of f (Boyd, Boyd, and Vandenberghe, 2004).

It can easily be seen that the approach of imposing constraints on gross-exposure 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. In contrast, the NL shrinkage has N degrees of freedom $\widehat{\lambda_1}(\widehat{\boldsymbol{\tau}}), \ldots, \widehat{\lambda_N}(\widehat{\boldsymbol{\tau}})$, 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.

This analysis provides a mathematical justification why the NL and DCC-NL can improve gross-exposure-constrained strategies *in spite* of the well-known tendency for gross exposure constraints to apply *some* shrinkage implicitly. Thus, the purpose of the empirical analysis in Section 3.2.1 will be to verify that the usefulness of NL and DCC-NL—already established in the unconstrained case by Ledoit and Wolf (2015) and Engle, Ledoit, and Wolf (2019), respectively—carries over to the more realistic case of exposure-constrained portfolios.

1.3 Constructing MVE Portfolios with Gross-Exposure Constraints

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

$$\min_{\boldsymbol{w}_t} \boldsymbol{w}_t' \hat{\Sigma}_t \boldsymbol{w}_t \tag{1.12}$$

$$\text{subject to } \boldsymbol{w}_t' 1 = 1 \ , \tag{1.13}$$

$$\boldsymbol{w}_t' \boldsymbol{m}_t = b_t \ \text{ and} \tag{1.14}$$

$$\sum_{i=1}^{N} |w_{i,t}| \leq \gamma, \tag{1.15}$$

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,

$$b_t = \boldsymbol{w}Q_t' \boldsymbol{m}_t \ , \tag{1.16}$$

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, $wQ_t^{(1)} = \cdots = wQ_t^{(d)} := 1/d$ and $wQ_t^{(d+1)} = \cdots = wQ_t^{(N)} := 0$, 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 (1.12) as $\boldsymbol{w}_{b,t}^*$, then we could obtain the following theorem similar to Theorem 1.

Theorem 2. *The partial constrained portfolio optimization problem (1.12) is equivalent to the optimization problem*

$$\begin{array}{c}\min_{\mathbf{w}_t} \mathbf{w}_t'\tilde{\Sigma}_{\gamma,t}\mathbf{w}_t \\ \text{subject to } \boldsymbol{w}_t'1 = 1 \text{ and } \boldsymbol{w}_t'\boldsymbol{m}_t = b_t.\end{array} \tag{1.17}$$

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

As in the GMV case of Section 1.2, we can see that the shrinkage implicit in gross-book-exposure constraints has only one degree of freedom, so NL shrinkage enjoys a theoretical advantage with its N degrees of freedom. This is even more so for DCC-NL, since it overlays a time-varying component that is missing in gross-book-exposure constraints. This provides theoretical justification for the empirical analysis of Section 3.2.2 that examines whether the results of Ledoit and Wolf (2015) and Engle, Ledoit, and Wolf (2019), for NL and DCC-NL, respectively, carry over from the unconstrained MVE portfolio to the exposure-constrained MVE portfolio.

2 Monte Carlo Simulations

In the last section, we demonstrate that imposing the gross-exposure constraint on a portfolio is equivalent to using the shrinkage estimator when the time-varying structure of the covariance matrix is captured by the DCC model. In this section, we use Monte Carlo simulations to quantify the finite sample performance of portfolios constructed using different covariance matrix estimators and with varying levels of gross-exposure constraint. In particular, we want to study and compare the usefulness of the NL shrinkage estimation, the DCC model, and the gross-exposure constraint in portfolio improvement and address the following questions, among others. Is it better to use the NL shrinkage covariance matrix estimator or the optimal gross-exposure constraint? Is it better to use a dynamic model (with DCC) or a static model (without DCC)? Does the NL shrinkage improve over a pure portfolio with a moderate gross-exposure constraint (for both the dynamic model and the static model)? Last but not least, how do the NL shrinkage, the DCC model, and the gross-exposure constraint affect the risk approximations?

2.1 Data Generating Process

To generate realistic simulations that match the empirical data, we first estimate the unconditional covariance matrix from the most liquid stocks ($N = 500$, 1000) in the CRSP database based on the NL shrinkage method using 5 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 $r_t = \Sigma_t^{1/2} z_t$ with disturbance terms z_t drawn from a multivariate standard normal distribution or a multivariate "Student" t-distribution with five degrees of freedom. The conditional covariance matrix Σ_t is generated from Equation (1.5) to Equation (1.7), with parameters $\alpha_i = 0.05$ and $\beta_i = 0.90$ for all individual stocks $i = 1, \ldots, N$ in Equation (1.6) and parameters $\alpha = 0.05$ and $\beta = 0.93$ in Equation (1.7). For each simulation, we thereby generate a $T \times N \times N$ time-varying covariance matrix and correspondingly a $T \times N$ matrix of simulated returns, where the time length T is

1250 and the portfolio size N is either 500 or 1000. We repeat each simulation for 100 times.

2.2 Portfolio Improvement

We construct GMV portfolios [Equation (1.8)] based on four different covariance matrix estimators, including the sample covariance matrix (S), the NL shrinkage estimator (NL) proposed by Ledoit and Wolf (2015), the covariance matrix estimator based on DCC model (Engle, 2002), and the DCC-NL estimator proposed by Engle, Ledoit, and Wolf (2019). For each covariance matrix estimator, we allow a wide choice of the gross-exposure parameter γ, ranging from 1 to 16, and we also entertain the situation where no constraint on gross-exposure is imposed ($\gamma = \infty$).

In Table 1, we report the annualized actual risks of the empirical portfolio obtained by Equation (2.1), the standard errors of weights and the total short positions of the empirical portfolios constructed based on the four covariance matrix estimators, either without any constraint on gross-exposure or with the optimal constraint on gross-exposure (+constraint*). All the results shown are for a typical simulated dataset, which has the median oracle risk among 100 simulations.

We can see that purely using a NL shrinkage estimator of covariance matrix uniformly dominates purely imposing an optimal constraint on gross-exposure, and using the DCC model helps improve the performance in all cases. For example, when $N = 1000$ and the disturbance terms z_t are generated from a multivariate standard normal distribution, imposing gross-exposure constraint reduces the annualized actual risk of the empirical portfolio from 6.63% to 4.11% at best, whereas using the NL covariance matrix estimator reduces the risk to 3.92%, and DCC-NL reduces the risk to 3.69%. We also find that additionally imposing an optimal constraint on the gross-exposure has no effect on the NL portfolio, but it improves the performance of a pure DCC-NL portfolio.

Remark 3.1 (Determinants of the optimal constraint). The optimal constraint (with gross-exposure parameter γ^*), which minimizes actual risk among portfolios with different degrees of gross-exposure constraint, is unknown in practice. According to Equation (2.1), the optimal constraint parameter γ^* is determined by both the real covariance matrix Σ_t and the estimated weights vector $\widehat{\boldsymbol{w}}_t$. Suppose the estimated weights vector is close to the oracle one when the gross-exposure parameter γ is around the oracle optimal parameter $\gamma^{*\mathrm{orc}}$. Then, γ^* should be close to the oracle optimal parameter $\gamma^{*\mathrm{orc}} = \sum_{i=1}^N |w_{i,t}^*|$, where $w_{i,t}^*$ denotes the i-th element of the optimal weights vector $w_t^* = \frac{\Sigma_t^{-1}1}{1'\Sigma_t^{-1}1}$. Let $\Sigma_t^{-1} = (\breve{\sigma}_{ij,t})_{1 \le i,j \le N}$, then $1'\Sigma_t^{-1}1 = \sum_{i=1}^N \sum_{j=1}^N \breve{\sigma}_{ij,t}$, and $\gamma^{*\mathrm{orc}} = \frac{\sum_{i=1}^N |\sum_{j=1}^N \breve{\sigma}_{ij,t}|}{\sum_{i=1}^N \sum_{j=1}^N \breve{\sigma}_{ij,t}}$. Given the positive definiteness of the inverse covariance matrix, $\sum_{i=1}^N \sum_{j=1}^N \breve{\sigma}_{ij,t}$ is positive. Therefore, if many of the row sums $\sum_{j=1}^N \breve{\sigma}_{ij,t}(i = 1, \ldots, N)$ are negatively large, then both $\gamma^{*\mathrm{orc}}$ and γ^* will be large. For a special case, if none of the row sums is negative, then $\gamma^{*\mathrm{orc}} = 1$ and γ^* will be close to 1. □

For a more comprehensive comparison, we present the annualized actual risks for the four covariance matrix estimators with a continuously changing gross-exposure constraint parameter γ in Figures 1 and 2, for simulated data with $z_t \sim N(0, 1)$ and $z_t \sim t(5)$, respectively. It is clear from the figures that using the DCC-NL covariance matrix achieves the

Table 1 Actual risk of portfolio and standard deviation and total short position of weights

	$z_t \sim N(0,1)$				$z_t \sim t(5)$			
	Γ	R_{act}	StdW	ShortW	γ	R_{act}	StdW	ShortW
Panel A: 500 stocks contained in the portfolio								
S	∞	5.53	2.35	403.15	∞	5.54	2.36	407.13
S + constraint*	5.5	5.10	1.72	225.00	6.0	5.14	1.80	250.00
NL	∞	5.01	1.58	258.62	∞	5.01	1.59	265.09
NL + constraint*	5.9	5.01	1.53	245.00	5.9	5.01	1.54	245.00
DCC	∞	5.11	2.33	397.52	∞	5.07	2.33	396.53
DCC + constraint*	4.6	4.45	1.57	180.00	4.5	4.40	1.56	175.00
DCC-NL	∞	4.40	1.68	270.62	∞	4.37	1.70	275.47
DCC-NL + constraint*	4.9	4.32	1.48	195.00	4.8	4.27	1.48	190.00
Panel B: 1000 stocks contained in the portfolio								
S	∞	6.63	1.92	701.15	∞	6.48	1.91	698.72
S + constraint*	5.0	4.11	0.87	200.00	5.1	4.04	0.88	205.00
NL	∞	3.92	0.74	245.30	∞	3.84	0.75	250.86
NL + constraint*	5.9	3.92	0.74	244.44	5.6	3.84	0.73	229.99
DCC	∞	6.43	1.84	665.61	∞	6.24	1.85	673.41
DCC + constraint*	4.4	3.91	0.82	169.98	4.4	3.84	0.82	169.98
DCC-NL	∞	3.69	0.79	252.52	∞	3.65	0.80	256.07
DCC-NL + constraint*	5.2	3.66	0.74	209.97	5.2	3.61	0.75	209.97

Notes: This table shows the simulation results for the empirical GMV portfolios constructed based on different covariance matrix estimators, facing no gross-exposure constraint or with an optimal gross-exposure constraint (+constraint*). All the results presented are the median values among 100 simulations. Panels A and B show results for portfolios with 500 and 1000 stocks, respectively. The left panel and the right panel show results for simulated data, where the disturbance terms are drawn from a multivariate standard normal distribution and a multivariate *t*-distribution with five degrees of freedom, respectively. 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 NL shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle, Ledoit, and Wolf, 2019). The sum of the absolute weights should not exceed the gross-exposure parameter γ. $\gamma = \infty$ means there is no constraint on the gross-exposure of the portfolio, and the optimal γ corresponds to the gross-exposure constraint where the portfolio has the minimum annualized actual risk (R_{act}). The standard deviation of weights (StdW) and the total short position of weights (ShortW) of the empirical portfolios are also reported. All the figures shown are in percentage.

minimum actual risks in all cases with different gross-exposure constraints. The gains from NL shrinkage increase as the constraint γ becomes less binding and the portfolio size N becomes larger. The former is explained by Jagannathan and Ma (2003) that imposing constraints on portfolio weights has a shrinkage-like effect, and thus it hurts the performance of the shrinkage estimator of covariance matrix. The latter is consistent with the finding of Ledoit and Wolf (2017a) that the amount of improvement is more pronounced in large dimensions. By contrast, the gains from DCC decrease as the constraint γ becomes less binding and the portfolio size N becomes larger. This is because without an effective constraint, the large errors generated in estimating the model impair the performance of DCC, and the large portfolio size exacerbates the problem.

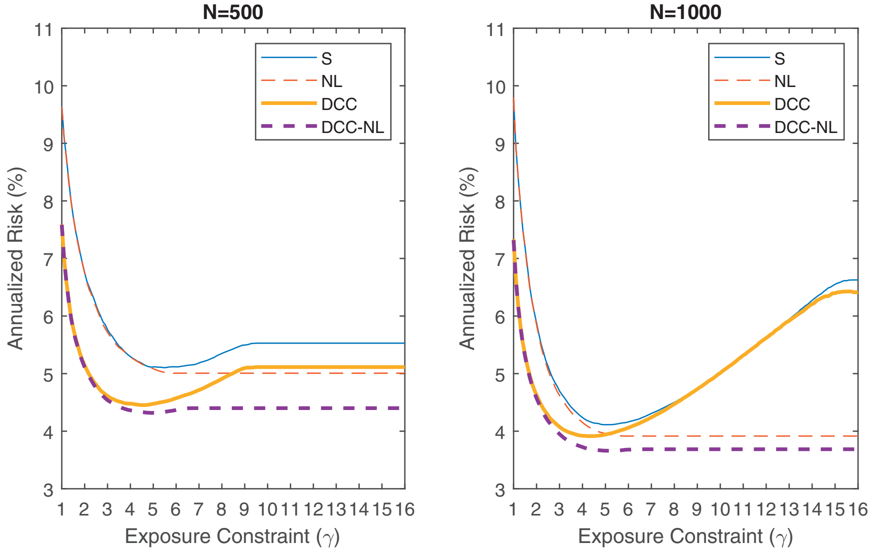

Figure 1 Comparisons of the median annualized actual risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of gross-exposure constraints (the intensity of the constraint declines with the increase of parameter γ). The disturbance terms of the simulated data are drawn from a multivariate standard normal distribution.

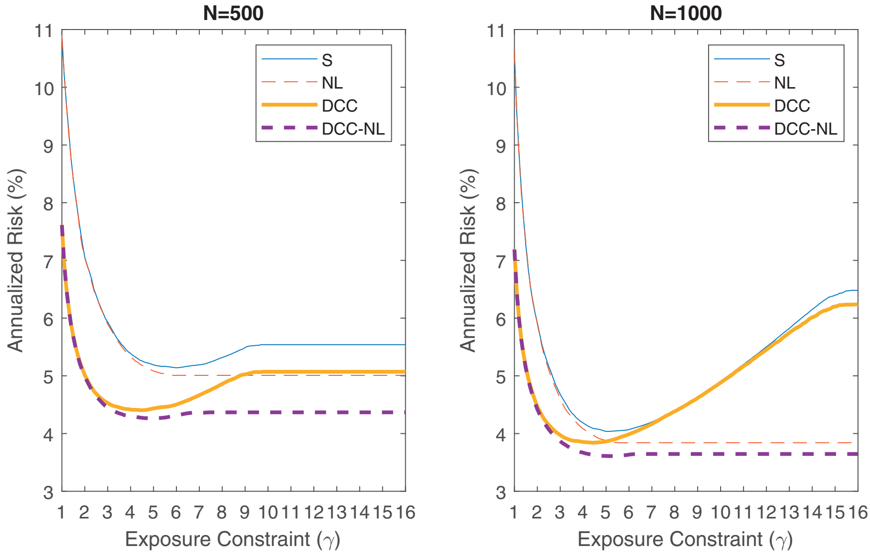

Figure 2 Comparisons of the median annualized actual risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of gross-exposure constraints (the intensity of the constraint declines with the increase of parameter γ). The disturbance terms of the simulated data are drawn from a multivariate *t*-distribution with five degrees of freedom.

Remark 3.2 (The performance of DCC when $T < N$). Due to completeness, we also considered the special case when the number of observations (T) is smaller than the number of stocks (N). It is well-known that when $T < N$, the sample covariance matrix is not of full rank, so its inverse will not exist. As the traditional DCC estimator relies on the sample covariance matrix, its inverse will not exist either. In this case, the advantage of using NL shrinkage estimators is self-evident. To evaluate the performance of DCC, we compare the results of NL and DCC-NL. The data generating process is the same as described in Section 2.1 except that we set $T = 500$ and $N = 1000$. In unreported results, we find that the benefit of using DCC model is weakened by the large ratio of N/T. In particular, when the gross-exposure is tight (γ is small), DCC-NL still delivers smaller actual risks, but NL performs better if no effective gross-exposure constraint is imposed. This finding indicates that in the estimation of the DCC model the large-dimensional small sample size problem ($N/T = 2$) causes large approximation errors, which can be reduced by imposing tight gross-exposure constraints. In this special case, it is advisable to use the pure NL method if no external gross-exposure constraint is imposed, but DCC-NL is still preferred if a common external gross-exposure (such as the "130/30" or the "150/50" requirement) is imposed. □

We also show the 10th, 50th, and 90th percentiles of R_{act} among the 100 simulations in Figure 3 for $N = 500$ and in Figure 4 for $N = 1000$, both for simulated data with $z_t \sim N(0, 1)$.[4] We can see that the sampling variation is always small. Moreover, the three percentiles of R_{act} for DCC-NL estimator are more closer than those of the other three estimators, indicating that the performance of DCC-NL estimator is more stable in portfolio selection.

The numbers in the columns StdW and ShortW of Table 1 indicate that both the shrinkage estimator and the gross-exposure constraint largely reduce the standard deviations of weights and the total short positions. By comparison, the shrinkage estimator reduces the standard deviations of weights more remarkably and the gross-exposure constraint has a more distinct effect on reducing the total short positions.

To sum up, for varying gross-exposure constraints, it is always the DCC-NL estimator that yields the lowest risks. Imposing moderate constraints on gross-exposure effectively reduces the risks but is uniformly dominated by using the DCC-NL covariance matrix estimator.

2.3 Risk Approximations

Following Fan, Zhang, and Yu (2012), besides R_{act}, we also computer the oracle risk (denoted by R_{orc}) of the theoretical optimal portfolio and the empirical risk (denoted by R_{emp}) based on the estimated weights and estimated covariance matrix to analyze the risk approximations. In particular, the three annualized risks are defined as follows:

$$R_{\text{act}} = \sqrt{252}\, \widehat{\boldsymbol{w}}_t' \Sigma_t \widehat{\boldsymbol{w}}_t. \tag{2.1}$$

4 Since the results for simulated data with $z_t \sim t(5)$ are very similar with those for simulated data with $z_t \sim N(0, 1)$, we do not report the details here for simplicity.

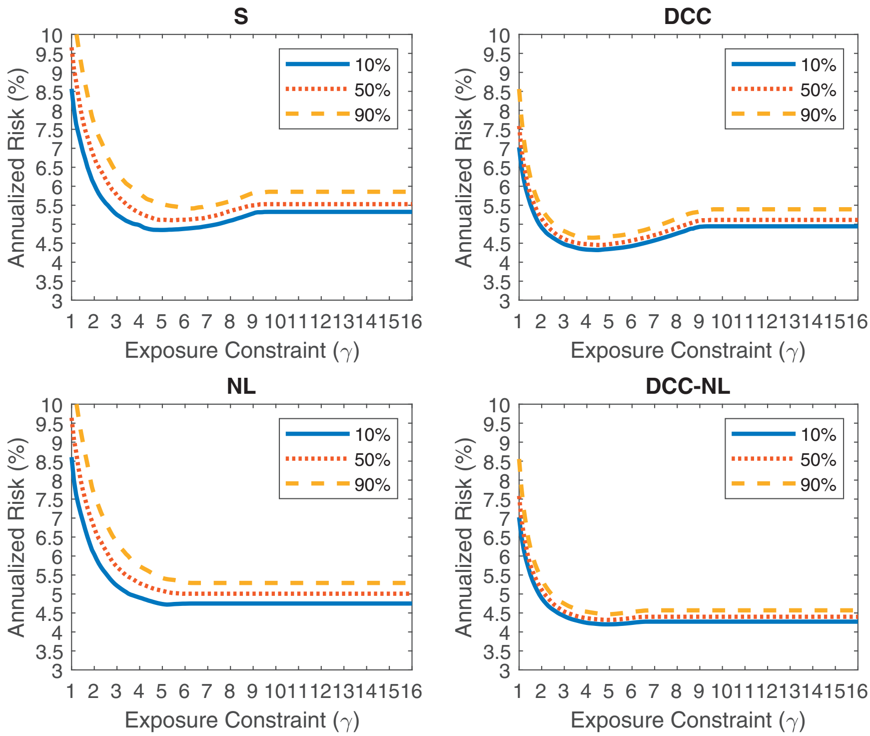

Figure 3 Comparisons of the 10%, 50%, and 90% quantiles of the annualized actual risks of the 100 simulated GMV portfolios constructed. $N = 500$, and the disturbance terms are drawn from a multivariate standard normal distribution.

$$R_{\text{orc}} = \sqrt{252}\ \boldsymbol{w}_t' \Sigma_t \boldsymbol{w}_t. \tag{2.2}$$

$$R_{\text{emp}} = \sqrt{252}\ \widehat{\boldsymbol{w}}_t' \hat{\Sigma}_t \widehat{\boldsymbol{w}}_t. \tag{2.3}$$

Note that only the empirical risk is known, and the difference between R_{emp} and R_{act} reflects the estimation error in the covariance matrix.

Figures 5 and 6 depict all three risks for $N = 500$ and $N = 1000$, respectively, both for simulated data with $z_t \sim N(0, 1)$. The curve of the oracle risk shared by the four graphs in Figure 5 (or Figure 6) indicates that the theoretical risk decreases quickly with the increase of the gross-exposure parameter γ before γ reaches 2, when the constraint forms the 150/50 strategy, which involves 150% long positions and 50% short positions. In fact, based on the true covariance matrix, the GMV portfolios have $\gamma^{*\text{orc}} = 5.758$ for $N = 500$ and $\gamma^{*\text{orc}} = 5.846$ for $N = 1000$, indicating oracle total short positions of 237.9% for $N = 500$ and 242.3% for $N = 1000$. When $\gamma > \gamma^{*\text{orc}}$, the oracle risk remains constant, and relaxing the constraint reduces the empirical risk but increases the actual risk. The increase of the actual risk is especially distinct when $N = 1000$ and S or DCC covariance matrix estimator is used.

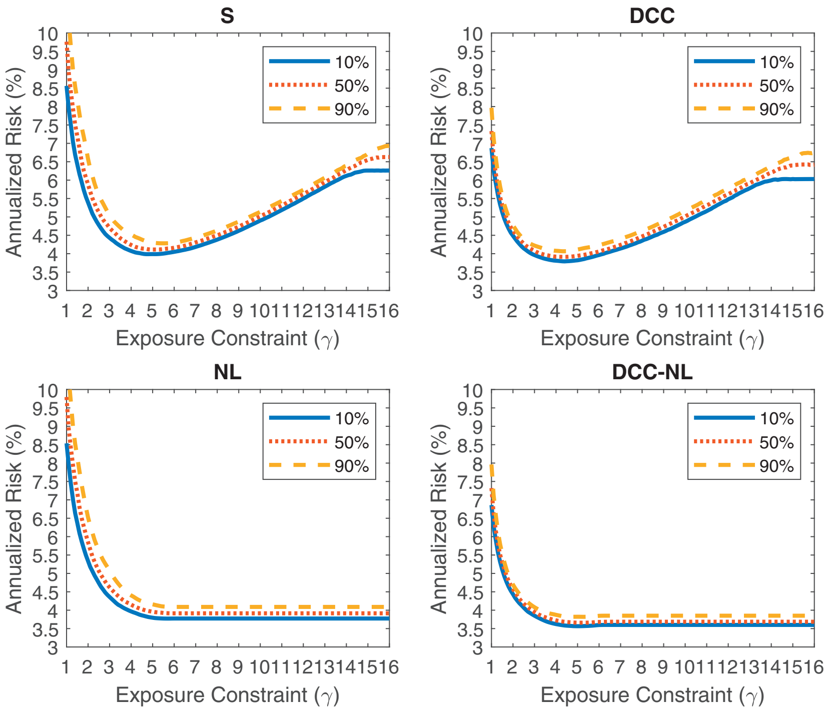

Figure 4 This is similar to Figure 3 except $N = 1000$.

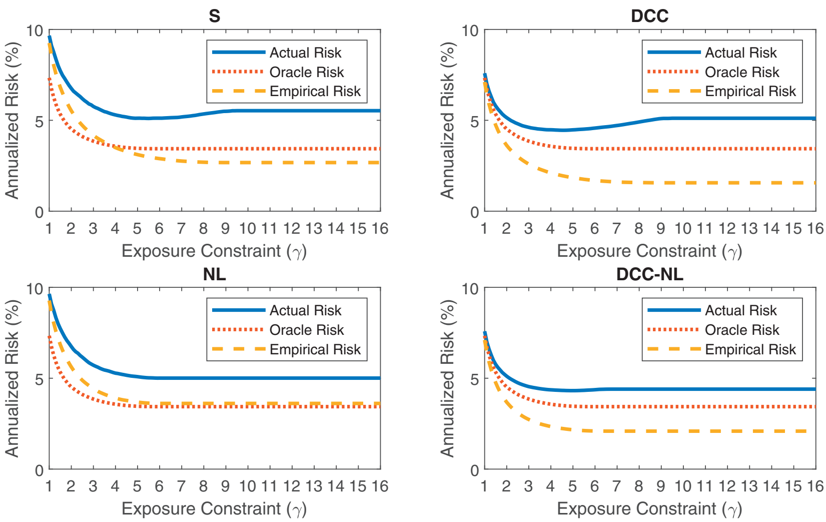

Figure 5 Comparisons of the median annualized oracle risks, median annualized actual risks, and median annualized empirical risks of the 100 simulated GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of gross-exposure constraints (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.

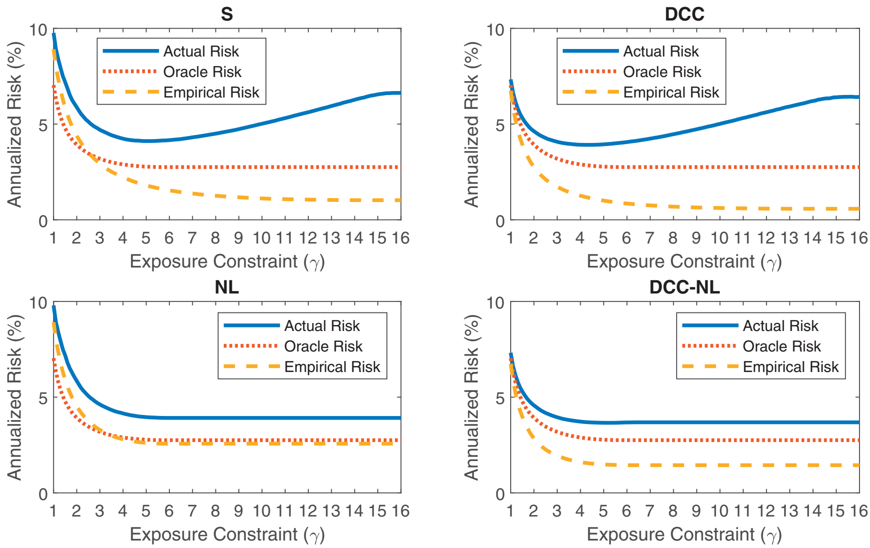

Figure 6 This is similar to Figure 5 except $N = 1000$.

The three risks are high but nearly the same when $\gamma = 1$. With the increase of γ, the difference between the actual risk and the empirical risk increases, suggesting that the estimation of the covariance matrix becomes harder. If no-shrinkage-involved covariance matrix estimator is used for constructing a large portfolio with $N = 1000$, the curves of actual risks start upward sloping when $\gamma > 4$, making both the actual–empirical gap and the actual–oracle gap increase dramatically. The difference between the actual and the oracle risks shows very similar properties to the actual–empirical gap, except that the former is much smaller than the latter when DCC model is used, indicating the important role of DCC model in estimating the optimal weights.

Overall, the NL shrinkage reduces approximation errors of the covariance matrix and improves the weights allocations when a relatively loose constraint is imposed, whereas the DCC model improves portfolio allocations and thereby reduces the actual risks. By contrast, approximation errors reduce with the reinforcement of the gross-exposure constraint, but a tight constraint impairs the portfolio performance as it results in too conservative allocations.

2.4 Combining Shrinkage with Gross-Exposure Constraints

As a final lesson, there are circumstances where the gross-exposure 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, Table 1 and Figures 1 and 2 show that there is still incremental benefit to using DCC-NL, even *after* the gross-exposure 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 GMV portfolio. Note, however, that the gains from shrinkage become monotonically weaker as the gross exposure constraint γ becomes more binding. This is because the gross-exposure constraint applies a "brute-force" one-size-fits-all *over*-shrinkage that leaves little room for the benefits of a locally adaptive optimal NL shrinkage formula to express themselves.

3 Empirical Results

3.1 Data

We examine the effects of the gross-exposure constraint, the shrinkage estimation of covariance matrix, and the use of DCC model to capture the heteroscedasticity on the out-of-sample performance of the GMV portfolio and the MVE portfolio. We use the same portfolio-construction rules as in Ledoit, Wolf, and Zhao (2019), except that we also impose different levels of 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.[5] The daily return data we use, which cover the period from January 1, 1980 to December 31, 2018, are from the Center for Research in Security Prices (CRSP) database. The out-of-sample period is from January 8, 1986 to December 31, 2018. We update the portfolios every 21 consecutive trading days, and thus form 396 investment dates from January 8, 1986 to December 31, 2018. At every investment date h, the covariance matrix is estimated based on the most recent 1250 daily returns (roughly equals 5 years).

For both the GMV portfolio and the MVE portfolio, we consider two different portfolio sizes $N = 500$ and 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.

We also consider the four covariance matrix estimators: the sample covariance matrix (S), the NL shrinkage estimator (NL) (Ledoit and Wolf, 2015), the covariance matrix estimator based on DCC model (Engle, 2002), and the DCC-NL estimator (Engle, Ledoit, and Wolf, 2019). To consider the effects of the gross-exposure 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.

3.2 Main Results

3.2.1 Results for GMV portfolios

Table 2 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

5 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.

Table 2 Out-of-sample performance, characteristics of weights, and average turnover of the GMV portfolio

	$\hat{\Sigma}$	AvR	StdR	IR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
$\gamma = \infty$	S	10.23	10.78	0.95	−6.23	10.27	1.73	255.14	6.50
	NL	10.63	9.75	1.09	−2.85	4.90	1.00	140.21	2.35
	DCC	13.22	10.44	1.27	−4.25	16.29	1.61	179.22	4.11
	DCC-NL	12.94	9.55***	1.35	−2.24	14.86	1.27	109.56	2.04
$\gamma = 2$	S	10.78	9.99	1.08	−3.43	9.84	1.02	50.00	3.19
	NL	10.69	9.86	1.08	−2.75	5.69	0.80	50.00	1.39
	DCC	12.22	9.82	1.24	−2.72	17.97	1.30	49.97	2.88
	DCC-NL	12.38	9.52***	1.30	−2.05	16.85	1.23	49.41	1.36
$\gamma = 1.6$	S	10.71	10.11	1.06	−3.12	10.64	0.99	30.00	2.80
	NL	10.77	10.06	1.07	−2.73	6.39	0.78	30.00	1.39
	DCC	12.02	9.78	1.23	−2.42	19.29	1.32	30.00	2.76
	DCC-NL	12.24	9.59***	1.28	−1.94	18.50	1.28	29.94	1.29
$\gamma = 1$	S	10.67	11.34	0.94	0.00	12.99	1.02	0.00	2.36
	NL	10.97	11.34	0.97	0.00	8.20	0.80	0.00	0.36
	DCC	11.13	10.19	1.09	0.00	24.83	1.52	0.00	1.29
	DCC-NL	11.24	10.17	1.10	0.00	24.74	1.52	0.00	0.17
Panel B: 1000 stocks contained in the portfolio									
$\gamma = \infty$	S	10.16	13.54	0.75	−7.26	9.78	1.69	557.00	12.62
	NL	10.64	8.81	1.21	−1.44	2.44	0.49	142.76	3.26
	DCC	10.28	10.51	0.98	−4.44	19.60	1.36	335.99	7.42
	DCC-NL	11.26	8.16***	1.38	−1.15	16.22	0.79	97.29	2.73
$\gamma = 2$	S	10.78	9.27	1.16	−2.39	8.41	0.62	50.00	3.56
	NL	10.67	9.13	1.17	−1.64	3.30	0.41	50.00	1.69
	DCC	10.71	8.52	1.26	−1.77	21.98	0.95	49.99	3.30
	DCC-NL	11.24	8.11***	1.39	−1.13	18.71	0.83	49.60	1.66
$\gamma = 1.6$	S	10.95	9.37	1.17	−2.07	9.15	0.62	30.00	2.96
	NL	10.92	9.41	1.16	−1.76	3.80	0.42	30.00	1.65
	DCC	10.90	8.35	1.31	−1.55	23.70	0.99	30.00	2.96
	DCC-NL	11.23	8.10***	1.39	−1.09	21.15	0.90	29.98	1.48
$\gamma = 1$	S	12.02	10.94	1.10	0.00	11.75	0.67	0.00	2.47
	NL	12.04	10.98	1.10	0.00	5.52	0.46	0.00	0.55
	DCC	9.96	8.62	1.16	0.00	31.73	1.24	0.00	1.42
	DCC-NL	9.66	8.53***	1.13	0.00	31.01	1.22	0.00	0.22

Notes: This table shows the out-of-sample results for the GMV portfolios constructed based on different covariance matrix estimators and facing various degrees of gross-exposure 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 NL shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle, Ledoit, and Wolf, 2019). γ is a gross-exposure parameter indicating the supremum of the sum of the absolute weights. Therefore, $\gamma = \infty, 2, 1.6, 1$ stands for an increasing restriction with the total short position not exceeding $\infty, 50\%, 30\%, 0$ of the total investment. We hold the portfolios for 21 days and record their daily returns. We report their out-of-sample AvR, annualized standard deviations (StdR), and IRs. 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 IRs. 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: ***, **, and * denote significance at the 0.01, 0.05, and 0.1 levels, respectively.

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 (IRs), which is the ratio of AvR to StdR.

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

As the gross-exposure constraint becomes more binding, the gains from shrinkage decline. Nevertheless, the DCC-NL estimator is always preferred: it delivers the smallest out-of-sample standard deviation even when an appropriate gross-exposure 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 gross-exposure constraints. The results show that the outperformance is always significant at the 0.01 level, except when no short position is allowed.

On the other hand, the moderate constraints with $\gamma = 2$ and $\gamma = 1.6$ 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 the 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 with the no-constraint strategy and the no-short-sale strategy, respectively. However, the effect of imposing a gross-exposure constraint is limited and becomes insignificant if the NL shrinkage estimator is used.

As expected, the combined effect of using DCC and imposing an appropriate gross-exposure constraint is not as good as that of using the DCC-NL estimator. Specifically, when $N = 1000$, the GMV portfolio constructed using the DCC estimator with the constraint of $\gamma = 1.6$ has an annualized out-of-sample standard deviation of 8.35%, reducing the 13.54% from using the sample covariance matrix by over 5 percentage points, but using the DCC-NL estimator without any gross-exposure constraint generates a lower standard deviation of 8.16%.

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 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 gross-exposure constraint is imposed; imposing the 30–50% constraint on gross-exposure also improves the portfolio performance, but is not as effective as using the DCC-NL estimator. Thus, investors should use the DCC-NL estimator to improve the portfolio performance rather than imposing an external gross-exposure constraint.

3.2.2 Results for MVE portfolios

Table 3 presents results for the MVE portfolios constructed based on the signal ROE, which is a well-known profitability factor that can indicate the growth potential of a firm

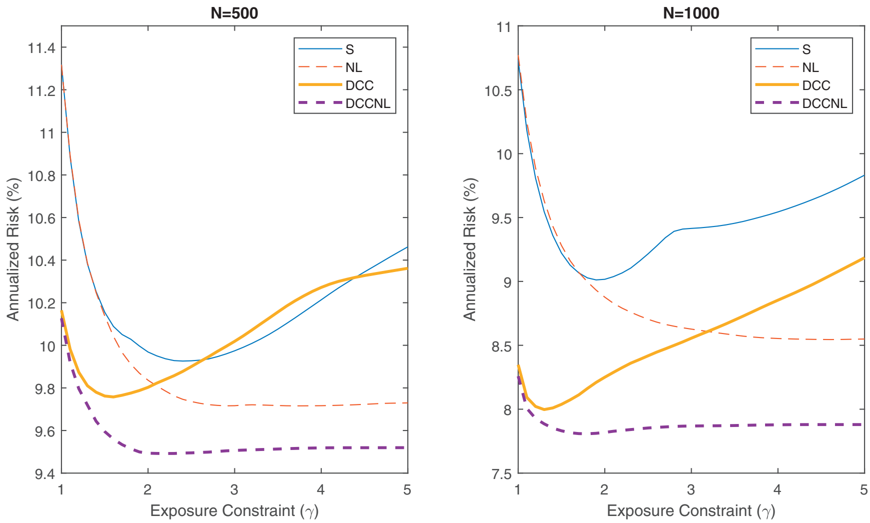

Figure 7 Comparisons of the out-of-sample risks of the GMV portfolios constructed based on different covariance matrix estimators (S, NL, DCC, DCC-NL) and facing various degrees of gross-exposure constraints (the intensity of the constraint declines with the increase of parameter γ).

(Haugen and Baker, 1996) and has been proven to have statistically significant explanatory power for cross-sectional anomalies (Hou, Xue, and Zhang, 2015; Feng, Giglio, and Xiu, 2020). We calculate ROE by the income before extraordinary divided by 1-quarter-lagged book equity. The data are from the merged CRSP/Compustat database.

Judging by the IRs and the corresponding significant test,[6] the DCC-NL estimator still performs the best among all covariance matrix estimators and its outperformance relative to DCC is statistically significant when the gross-exposure constraint is not too tight. When $N = 1000$ and no constraint is imposed on weights, the MVE portfolio constructed using the DCC-NL estimator has an IR 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. In addition, directly using the NL shrinkage estimator excels imposing the 30% or 50% short-sale constraint, both of which are superior to imposing the no-short-sale constraint, whereas 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 NL shrinkage technique, it is difficult to achieve the optimal shrinkage level for the gross-exposure constraint with one degree of freedom.

6 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 IR is significant.

Table 3 Out-of-sample performance, characteristics of weights, and average turnover of the MVE portfolio constructed based on the ROE signal

	$\hat{\Sigma}$	AvR	StdR	IR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
$\gamma=\infty$	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
	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
$\gamma=2$	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
	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
$\gamma=1.6$	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
	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
$\gamma=1$	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
	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									
$\gamma=\infty$	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
	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
$\gamma=2$	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
	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
$\gamma=1.6$	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
	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
$\gamma=1$	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
	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 out-of-sample results for the MVE portfolios constructed based on the signal ROE, using different covariance matrix estimators and facing various degrees of gross-exposure 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 NL shrinkage estimator (NL) (Ledoit and Wolf, 2015), the DCC estimator (Engle, 2002), and the DCC-NL estimator (Engle, Ledoit, and Wolf, 2019). γ is a gross-exposure parameter indicating the supremum of the sum of the absolute weights. Therefore, $\gamma=\infty, 2, 1.6, 1$ stands for an increasing restriction with the total short position not exceeding $\infty, 50\%, 30\%, 0$ of the total investment. We hold the portfolios for 21 days and record their daily returns. We report their out-of-sample AvR, annualized standard deviations (StdR), and IRs. 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 IRs. Panels A and 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: ***, **, and * denote significance at the 0.01, 0.05, and 0.1 levels, respectively.

3.3 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 January 8, 1986 to December 31, 2018 for the four characteristics of portfolio weights in Tables 2 and 3.

We find that portfolios constructed based on the sample covariance matrix have the smallest minimum weight and the largest total short position, whereas 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 gross-exposure 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 335.99% if the DCC estimator is used, and this number even comes up to 557.00 if the sample covariance matrix is used. For MVE portfolios, the corresponding short positions are even larger, with 361.70% for the DCC estimator and 571.36% for the sample covariance matrix.

Similar to imposing the gross-exposure constraints, using the NL shrinkage estimators largely reduces the total short position and the turnover of portfolios. This should not be surprising considering how the NL shrinkage method works in improving the estimation precision of covariance matrix.

3.4 Robustness Checks

3.4.1 Alternative predictors

For robustness check, we consider the MVE portfolios with an alternative factor exposure. Instead of using the signal ROE, now we focus on a factor that not only indicates growth potential but also reflects price level. We follow Basu (1983) and use E/P, measured as income before extraordinary divided by the market capitalization.

Table 4 presents the results for the MVE 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 the constraint on gross-exposure is loose enough. When $N = 1000$ and no constraint on gross-exposure is imposed, the NL estimator and the DCC-NL estimator increase the IR based on the sample covariance matrix by more than 70% and 90%, respectively. The gains from NL shrinkage become weaker as the intensity of the gross-exposure constraint increases and the size of the portfolio decreases. 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 S or DCC is used. Third, the effect of using the NL estimator is more remarkable than that of imposing the 30% or 50% short-sale constraint. When $N = 1000$, the use of NL estimator and the 50% short-sale constraint raise the IR of the portfolio constructed from 1.03 to 1.74 and 1.62, respectively.

Table 4 Out-of-sample performance, characteristics of weights, and average turnover of the MVE portfolio constructed based on the E/P signal

	$\hat{\Sigma}$	AvR	StdR	IR	MinW	MaxW	StdW	ShortW	AvT
Panel A: 500 stocks contained in the portfolio									
$\gamma = \infty$	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
	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
$\gamma=2$	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
	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
$\gamma = 1.6$	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
	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
$\gamma=1$	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
	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									
$\gamma = \infty$	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
	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
$\gamma=2$	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
	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
$\gamma = 1.6$	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
	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
$\gamma=1$	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
	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 is similar to Table 3 except that the MVE portfolios are constructed based on the E/P signal.

3.4.2 Transaction costs

Transaction costs are important issues in practical implementations (Mei, DeMiguel, and Nogales, 2016; Mei and Nogales, 2018). In Table 5, we present results for the MVE portfolios constructed based on the ROE signal, when the transaction costs are 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 costs.

Table 5 Out-of-sample performance of the MVE portfolio constructed based on the ROE signal with transaction costs considered

		Spread = 3 basis-points			Spread = 5 basis-points		
	$\hat{\Sigma}$	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 out-of-sample results for the MVE portfolios constructed based on the signal ROE, using different covariance matrix estimators and facing various degrees of gross-exposure constraints when transaction costs are considered. We report their out-of-sample AvR, annualized standard deviations (StdR), and IRs. AvR and StdR are shown in percentage. Panels A and B show results for portfolios with 500 and 1000 stocks, respectively. The left and right panels show results under the assumptions of three basis-points and five 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: ***, **, and * denote significance at the 0.01, 0.05, and 0.1 levels, respectively.

Unsurprisingly, the IR becomes lower with the increase of transaction costs. Features suggested by the pattern of IRs 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 gross-exposure constraint is imposed. When $N = 500$ and the bid–ask spread is three basis-points, using DCC-NL estimator increases the IR of using S estimator from 0.94 to 1.23, if no constraint is imposed on gross-exposure. When constraints are imposed, the DCC-NL estimator still helps increase the IR. Moreover, owing to the parsimony of the shrinkage method in turnover, the advantage of the DCC-NL estimator over the DCC (S) estimator becomes more significant and robust than when the transaction costs are ignored.

3.5 Combining Shrinkage with Gross-Exposure Constraint in Practice

In the process of researching the comparison of gross-exposure constraint versus 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 gross-exposure-constrained portfolio, upgrading from the sample covariance matrix to the DCC-NL estimator (while preserving the gross-exposure constraint) yields benefits almost as across-the-board as in the Monte-Carlo simulations analyzed in Section 2.4. The pattern identified earlier still holds: shrinkage has more room for improvement if gross-exposure is less binding of a constraint.

With respect to a pure DCC-NL portfolio, gradually tightening the gross-exposure 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:

- GMV portfolio, 500-stock universe.
- GMV portfolio, 1000-stock universe.
- ROE-optimized portfolio, 5-bp transaction cost, and 1000-stock universe.

In most other cases, moving from an absence of gross-exposure 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 short-sale does not start to hurt (much), provided we have a good NL shrinkage estimator of the conditional covariance matrix such as DCC-NL.

4 Conclusion

Constraints on portfolio weights 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 prevalent in practice, that is, the so-called "130/30" strategy and "150/50" strategy. Previous studies argue that using an advanced estimator for the covariance matrix is unnecessary or even harmful in constructing GMV or MVE portfolios if some certain constraints on weights are in place as the two different approaches reduce portfolio risks in a similar way. This paper finds that the DCC-NL estimator, which considers both the dynamics and the estimation precision of the covariance matrix, is always preferred for a wide range of gross-exposure constraint parameters.

We extend the mathematical connection between imposing the gross-exposure constraint and using the shrinkage estimator of the covariance matrix to a dynamic framework. Despite the mathematical equivalence, the NL shrinkage method has at least two advantages compared with the gross-exposure constraint: first, it has N degrees of freedom, leaving enough room for the benefits of the optimization; second, it uses an automatic procedure to achieve optimization and does not need any exogenous constraint parameter. Thus, using the NL shrinkage estimator outperforms imposing a gross-exposure constraint. Besides the above finding, we also demonstrate through Monte Carlo simulations that upgrading from the sample covariance matrix to the DCC-NL estimator is beneficial, even after the gross-exposure constraint has already been imposed. 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, and the introduction of an appropriate shrinkage, which reduces estimation errors.

Based on daily return data from stocks traded on the NYSE, AMEX, and NASDAQ, we construct GMV portfolios and MVE portfolios exposed to return predictive signals for portfolio sizes N =500 and 1000. The empirical results show that with respect to a gross-exposure-constrained portfolio, using the DCC-NL estimator yields the best out-of-sample performance in all cases, and the benefits of the shrinkage are significant if the gross-exposure constraint is moderate and thus there is enough room for the benefits of the shrinkage formula. In contrast, gradually tightening the gross-exposure constraint often hurts a pure DCC-NL portfolio when γ is smaller than 2, corresponding to the constraint that allows 50% short positions at most. Though imposing an appropriate gross-exposure constraint has a similar effect as using the NL shrinkage estimator in reducing risks, the latter always performs better. Taking into account the dynamic nature and upgrading to the DCC-NL estimator improve the performance even further. In addition, both the NL shrinkage and the gross-exposure constraint help reduce the standard deviation of weights and the turnover of portfolios.

In our main study, we use the ROE as a proxy for the expected return to construct the MVE portfolio. As robustness checks, we consider the E/P as an alternative proxy and also take the effect of transaction costs into account by assuming the bid–ask spread to be three or five basis-points. Our findings turn out to be robust to both of these changes.

Appendix A

A. Theoretical Justifications

Proof of Theorem 1. (i) Note that the matrix $\tilde{\Sigma}_{\gamma,t}$ is obviously symmetric and the solution to problem (1.8) is denoted by $\boldsymbol{w}_t^*$. For any vector x,

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

Based on the KKT conditions in Equation (1.10), $2\hat{\Sigma}_t\boldsymbol{w}_t^* - \mu 1 + \lambda\mathbf{g}_t^* = 0$. Therefore,

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

Note that

$$|(x'1)\left(x'\hat{\Sigma}_t \boldsymbol{w}_t^*\right)| = |(x'1)\left(x'\hat{\Sigma}_t^{\frac{1}{2}}\right)\left(\hat{\Sigma}_t^{\frac{1}{2}} \boldsymbol{w}_t^*\right)| \le |(x'1)(x\prime\hat{\Sigma}_t x)^{\frac{1}{2}}(\boldsymbol{w}_t^{*\prime}\hat{\Sigma}_t \boldsymbol{w}_t^*)^{\frac{1}{2}}|,$$

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

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

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

Hence,

$$|(x'1)\left(x'\hat{\Sigma}_t \boldsymbol{w}_t^*\right)| \le |x'1|(x\prime\hat{\Sigma}_t x)^{\frac{1}{2}}\left(\frac{1}{2}\mu\right)^{\frac{1}{2}}. \tag{A.3}$$

Combining (A.1)–(A.3), we have

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

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

Moreover, $(a-b)^2 + b^2$ is always nonnegative and is zero if and only if $a = b$ and $b = 0$ hold simultaneously. However, $a = (x\prime\hat{\Sigma}_t x)^{\frac{1}{2}} > 0$ 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) First, the optimization problem (1.11) with equality constraint could be solved through the Lagrange multiplier method. Construct the Lagrangian

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

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

$$\begin{cases} 2\tilde{\Sigma}_{\gamma,t}\boldsymbol{w}_t^{\text{opt}} - \mu_\gamma 1 = 0, \\ \boldsymbol{w}_t^{\text{opt}\prime}1 - 1 = 0. \end{cases}$$

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

$$\boldsymbol{w}_t^{\text{opt}} = \frac{\tilde{\Sigma}_{\gamma,t}^{-1}1}{1'\tilde{\Sigma}_{\gamma,t}^{-1}1}. \tag{A.5}$$

By the Lagrange multiplier method, problem (1.8) is to minimize

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

Based on the fact that $\mathbf{g}_t^{*\prime}\boldsymbol{w}_t^* = ||\boldsymbol{w}_t^*||_1$ and KKT conditions in Equation (1.10), we have

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

Hence, the solution to problem (1.8) $\boldsymbol{w}_t^* = \frac{1}{2}(\lambda\gamma+\mu)\tilde{\Sigma}_{\gamma,t}^{-1}1$. Moreover, because of the constraint $\boldsymbol{w}_t'1 = 1$, solving for $\lambda\gamma+\mu$ yields $\lambda\gamma+\mu = \frac{2}{1'\tilde{\Sigma}_{\gamma,t}^{-1}1}$. This fact indicates that

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

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. First, the optimization problem (1.17) with equality constraint could be solved through the Lagrange multiplier method. Construct the Lagrangian

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

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

$$\begin{cases} 2\tilde{\Sigma}_{\gamma,t}\boldsymbol{w}_{b,t}^{\text{opt}} - \mu_{1\gamma}1 - \mu_{2\gamma}\boldsymbol{m}_t = 0, \\ \boldsymbol{w}_{b,t}^{\text{opt}\prime}1 - 1 = 0,\ \boldsymbol{w}_{b,t}^{\text{opt}\prime}\boldsymbol{m}_t - b_t = 0. \end{cases}$$

Therefore, $\boldsymbol{w}_{b,t}^{\text{opt}} = \mu_{1c}\tilde{\Sigma}_{\gamma,t}^{-1}1 + \mu_{2\gamma}\tilde{\Sigma}_{\gamma,t}^{-1}\boldsymbol{m}_t = \tilde{\Sigma}_{\gamma,t}^{-1}\left(1,\boldsymbol{m}_t\right)\binom{\mu_{1\gamma}}{\mu_{2\gamma}}$.

The above equations also imply that

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

or

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

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

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

Therefore, the solution to this problem is given by

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

By the Lagrange multiplier method, problem (1.12) is to minimize

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

So, the KKT conditions are

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

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

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

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

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

or

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

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

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

Hence,

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

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

References

Ao, M., Y. Li, and X. Zheng. 2019. Approaching Mean–Variance Efficiency for Large Portfolios. *The Review of Financial Studies* 32: 2890–2919.

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

Bai, J., and K. Li. 2012. Statistical Analysis of Factor Models of High Dimension. *The Annals of Statistics* 40: 436–465.

Basu, S. 1983. The Relationship between Earnings' Yield, Market Value and Return for NYSE Common Stocks: Further Evidence. *Journal of Financial Economics* 12: 129–156.

Bauder, D., T. Bodnar, N. Parolya, and W. Schmid. 2020. Bayesian Mean–Variance Analysis: Optimal Portfolio Selection under Parameter Uncertainty. *Quantitative Finance* 1–22.

Behr, P., A. Guettler, and F. Miebs. 2013. On Portfolio Optimization: Imposing the Right Constraints. *Journal of Banking & Finance* 37: 1232–1242.

Bodnar, T., N. Parolya, and W. Schmid. 2018. Estimation of the Global Minimum Variance Portfolio in High Dimensions. *European Journal of Operational Research* 266: 371–390.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg. 2018. Modeling and Forecasting (un)Reliable Realized Covariances for More Reliable Financial Decisions. *Journal of Econometrics* 207: 71–91.

Boyd, S., S. P. Boyd, and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Brandt, M. W., P. Santa-Clara, and R. Valkanov. 2009. Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns. *Review of Financial Studies* 22: 3411–3447.

Branger, N., K. Lučivjanská, and A. Weissensteiner. 2019. Optimal Granularity for Portfolio Choice. *Journal of Empirical Finance* 50: 125–146.

Britten-Jones, M. 1999. The Sampling Error in Estimates of Mean–Variance Efficient Portfolio Weights. *The Journal of Finance* 54: 655–671.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris. 2009. Sparse and Stable Markowitz Portfolios. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12267–12272.

Callot, L., M. Caner, A. Ö. Önder, and E. Ulasan. 2020. A Nodewise Regression Approach to Estimating Large Portfolios. *Journal of Business & Economic Statistics*, forthcoming.

De Nard, G., O. Ledoit, and M. Wolf. 2020. Factor Models for Portfolio Selection in Large Dimensions: The Good, the Better and the Ugly. *Journal of Financial Econometrics*, forthcoming (https://doi.org/10.1093/jjfinec/nby033).

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal. 2009a. A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management Science* 55: 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009b. Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies* 22: 1915–1953.

DeMiguel, V., A. Martin-Utrera, and F. J. Nogales. 2013. Size Matters: Optimal Calibration of Shrinkage Estimators for Portfolio Selection. *Journal of Banking & Finance* 37: 3018–3034.

Engle, R. F. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 20: 339–350.

Engle, R. F., O. Ledoit, and M. Wolf. 2019. Large Dynamic Covariance Matrices. *Journal of Business & Economic Statistics* 37: 363–375.

Fan, J., Y. Fan, and J. Lv. 2008. High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics* 147: 186–197.

Fan, J., Y. Liao, and M. Mincheva. 2013. Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 75.

Fan, J., J. Zhang, and K. Yu. 2012. Vast Portfolio Selection with Gross-Exposure Constraints. *Journal of the American Statistical Association* 107: 592–606.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance*, 75: 1327–1370.

Frahm, G., and C. Memmel. 2010. Dominating Estimators for Minimum-Variance Portfolios. *Journal of Econometrics* 159: 289–302.

Garlappi, L., R. Uppal, and T. Wang. 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *Review of Financial Studies* 20: 41–81.

Goldfarb, D., and G. Iyengar. 2003. Robust Portfolio Selection Problems. *Mathematics of Operations Research* 28: 1–38.

Golosnoy, V., and Y. Okhrin. 2007. Multivariate Shrinkage for Optimal Portfolio Weights. *The European Journal of Finance* 13: 441–458.

Green, J., J. R. M. Hand, and X. F. Zhang. 2017. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies* 30: 4389–4436.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. …. and the Cross-Section of Expected Returns. *Review of Financial Studies* 29: 5–68.

Haugen, R. A., and N. L. Baker. 1996. Commonality in the Determinants of Expected Stock Returns. *Journal of Financial Economics* 41: 401–439.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting Anomalies: An Investment Approach. *Review of Financial Studies* 28: 650–705.

Jagannathan, R., and T. Ma. 2003. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance* 54: 1651–1684.

Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10: 603–621.

Ledoit, O., and M. Wolf. 2004a. Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management* 30: 110–119.

Ledoit, O., and M. Wolf. 2004b. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* 88: 365–411.

Ledoit, O., and M. Wolf. 2008. Robust Performance Hypothesis Testing with the Sharpe Ratio. *Journal of Empirical Finance* 15: 850–859.

Ledoit, O., and M. Wolf. 2011. Robust Performances Hypothesis Testing with the Variance. *Wilmott Magazine* 2011: 86–89.

Ledoit, O., and M. Wolf. 2012. Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices. *The Annals of Statistics* 40: 1024–1060.

Ledoit, O., and M. Wolf. 2015. Spectrum Estimation: A Unified Framework for Covariance Matrix Estimation and PCA in Large Dimensions. *Journal of Multivariate Analysis* 139: 360–384.

Ledoit, O., and M. Wolf. 2017a. Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *The Review of Financial Studies* 30: 4349–4388.

Ledoit, O., and M. Wolf. 2017b. Numerical Implementation of the QuEST Function. *Computational Statistics & Data Analysis* 115: 199–223.

Ledoit, O., and M. Wolf. 2020a. Analytical Nonlinear Shrinkage of Large-Dimensional Covariance Matrices. *Annals of Statistics*, 48: 3043–3065.

Ledoit, O., and M. Wolf. 2020b. The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics*, forthcoming (https://doi.org/10.1093/jjfinec/nbaa007).

Ledoit, O., M. Wolf, and Z. Zhao. 2019. Efficient Sorting: A More Powerful Test for Cross-Sectional Anomalies. *Journal of Financial Econometrics* 17: 645–686.

Li, J. 2015. Sparse and Stable Portfolio Selection with Parameter Uncertainty. *Journal of Business & Economic Statistics* 33: 381–392.

Marčenko, V. A., and L. A. Pastur. 1967. Distribution of Eigenvalues for Some Sets of Random Matrices. *Mathematics of the USSR-Sbornik* 1: 457–483.

Markowitz, H. 1952. Portfolio Selection. *Journal of Finance* 7: 77–91.

Mei, X., V. DeMiguel, and F. J. Nogales. 2016. Multiperiod Portfolio Optimization with Multiple Risky Assets and General Transaction Costs. *Journal of Banking & Finance* 69: 108–120.

Mei, X., and F. J. Nogales. 2018. Portfolio Selection with Proportional Transaction Costs and Predictability. *Journal of Banking & Finance* 94: 131–151.

Pakel, C., N. Shephard, K. Sheppard, and R. F. Engle. 2020. Fitting Vast Dimensional Time-Varying Covariance Models. *Journal of Business & Economic Statistics*, forthcoming.

Pástor, L. 2000. Portfolio Selection and Asset Pricing Models. *The Journal of Finance* 55: 179–223.

Pástor, L., and R. F. Stambaugh. 2009. Predictive Systems: Living with Imperfect Predictors. *The Journal of Finance* 64: 1583–1628.

Sharpe, W. F. 1963. A Simplified Model for Portfolio Analysis. *Management Science* 9: 277–293.

Stock, J. H., and M. W. Watson. 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97: 1167–1179.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso: A Retrospective. *Journal of the Royal Statistical Society* 58: 267–288.

Tu, J., and G. Zhou. 2010. Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty. *Journal of Financial and Quantitative Analysis* 45: 959–986.

Tu, J., and G. Zhou. 2011. Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies. *Journal of Financial Economics* 99: 204–215.

Wang, Z. 2005. A Shrinkage Approach to Model Uncertainty and Asset Allocation. *Review of Financial Studies* 18: 673–705.

Webster, K., Y. Luo, M. A. Alvarez, J. Jussa, S. Wang, G. Rohal, A. Wang, D. Elledge, and G. Zhao. 2015. A Portfolio Manager's Guidebook to Trade Execution: From Light Rays to Dark Pools. *Deutsche Bank Markets Research*. Quantitative Strategy: Signal Processing, Deutsche Bank Markets Research, New York, USA.

Yen, Y.-M. 2016. Sparse Weighted-Norm Minimum Variance Portfolios. *Review of Finance* 20: 1259–1287.