JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 50, No. 6, Dec. 2015, pp. 1415–1441

doi:10.1017/S0022109015000526

Improving Mean Variance Optimization through Sparse Hedging Restrictions

Shingo Goto and Yan Xu*

Abstract

In portfolio risk minimization, the inverse covariance matrix prescribes the hedge trades in which a stock is hedged by all the other stocks in the portfolio. In practice with finite samples, however, multicollinearity makes the hedge trades too unstable and unreliable. By shrinking trade sizes and reducing the number of stocks in each hedge trade, we propose a "sparse" estimator of the inverse covariance matrix. Comparing favorably with other methods (equal weighting, shrunk covariance matrix, industry factor model, nonnegativity constraints), a portfolio formed on the proposed estimator achieves significant out-of-sample risk reduction and improves certainty equivalent returns after transaction costs.

I. Introduction

In mean variance portfolio optimization, originally developed by Markowitz (1952), the inverse covariance matrix of stock returns,[1] Σ^{-1}, plays a pivotal role in transforming return forecasts μ into optimized portfolio weights by $w \propto \Sigma^{-1}\mu$. Therefore, we take the liberty of calling the inverse covariance matrix, $\Psi \equiv \Sigma^{-1}$, the mean variance optimizer.[2] The mean variance optimizer also transforms a naïve diversification policy (equal weighting across N assets) into the global minimum variance portfolio $w_{\text{GMV}} \propto \Sigma^{-1}\mathbf{1}_N$.[3] In essence, the mean

*Goto (corresponding author), shingo.goto@moore.sc.edu, University of South Carolina, Moore School of Business, Columbia, SC 29208; Xu, yanxuj@hku.hk, University of Hong Kong, Faculty of Business and Economics, Hong Kong. We thank Hendrik Bessembinder (the editor), Bruno Gerard, Aurobindo Ghosh, Yufeng Han, Po-Hsuan Hsu, Jingzhi (Jay) Huang, Yong Jin, Chris Kirby, Steven Mann, Zhan Shi, Walter Torous, Jun Tu, Raman Uppal, Rossen Valkanov, Hao Wang, Masahiro Watanabe, Hong Yan, Shu Yan, Frank Yu, Yuzhao Zhang, Guofu Zhou, and especially an anonymous referee for helpful comments and discussions. We thank conference participants at the 2010 European Finance Association Meeting, 2010 Financial Management Association Meeting, and 2012 China International Conference on Finance. We also thank seminar participants at City University of Hong Kong, Nanyang Technological University, Research Affiliates, LLC, Singapore Management University, the University of Hong Kong, the University of Rhode Island, and the University of South Carolina. This research also benefits greatly from discussions with a number of investment professionals. We are especially grateful to Miguel Alvarez, Nick Baturin, Denis Chaves, Ernie Chow, Julian Douglass, Jonathan Howe, Jason Hsu, Dawn Jia, Vitali Kalesnik, Xiaowei Li, Scott Shaffer, Omid Shakernia, and Li Wang for helpful discussions.

[1]*Stocks* in this paper can mean any risky assets. We assume that a riskless asset exists for both borrowing and lending. Unless otherwise stated, the term *stock return* in this paper refers to excess return, which is calculated by subtracting the return of the riskless asset from the total return.

[2]The inverse covariance matrix is also called the precision matrix or concentration matrix (Dempster (1972)).

[3]$\mathbf{1}_N$ denotes an N-vector of 1s. $\propto$ means "proportional to."

variance optimizer achieves risk reduction over the naïve diversification rule by i) allocating larger weights to stocks with lower risk, and ii) utilizing hedging relations among stocks under consideration for the portfolio.

Although simple and forceful in theory, implementing mean variance optimization turns out to be surprisingly difficult in practice. A sizable literature (see Brandt (2009) for a review) documents that mean variance portfolio optimization tends to erode, rather than enhance, the gains from naïve diversification policies such as equal weighting (e.g., Jobson and Korkie (1981b), DeMiguel, Garlappi, and Uppal (2009)). Michaud (1989) even calls the portfolio optimization an "error maximization." This problem is especially acute in situations in which the sample covariance matrix contains large estimation errors, that is, when the number of available historical return observations per stock (T) is not much larger than the number of stocks in the portfolio (N), or when the portfolio contains highly correlated stocks. A telling example is given by Best and Grauer (1991), who consider randomly sampled portfolios of 10 to 100 stocks ($N = 10, \ldots, 100$) with 120 monthly returns ($T = 120$). They show that the sensitivity of optimized portfolio weights to perturbations in inputs can be on the order of 4,000 times greater in the 100-stock universe than in the 10-stock universe.[4] Kan and Zhou (2007) also demonstrate that when N/T is not so small, estimation errors in return forecasts and covariances produce large interactive effects, making the optimal weighting solutions unstable and unreliable.[5]

The goal of this paper is to suggest an improvement of the mean variance optimizer, $\Psi \equiv \Sigma^{-1}$, to ensure a superior out-of-sample portfolio risk reduction even when the covariance matrix and its inverse are not precisely estimated. The starting point is the insight of Stevens (1998) that the inverse covariance matrix $\Psi \equiv \Sigma^{-1}$ reveals hedging trades among stocks. Specifically, the ith row (or column) of Ψ is proportional to the stock's minimum variance *hedge portfolio*. The hedge portfolio consists of a long position in the ith stock, and a short position in the "tracking portfolio" of the other $N - 1$ stocks to hedge the ith stock.

The off-diagonal elements of the inverse covariance matrix represent the scaled ordinary least squares (OLS) regression coefficients of each stock on the remaining $N - 1$ stocks in the portfolio. However, in a portfolio of 100 stocks, for example, hedging each specific stock with the other 99 stocks is not necessarily optimal in small samples commonly available because multicollinearity subjects the coefficients to large estimation errors. From this perspective, we can enhance the robustness of the mean variance optimization in the same way that we address multicollinearity problems in regressions. Our strategy to get around the multicollinearity problem is twofold: i) to shrink coefficients in each hedge regression

[4]The sensitivity depends on the largest and smallest eigenvalues of Σ, and in particular on the condition number (the largest eigenvalue over the smallest eigenvalue). See Best and Grauer (1991) for exact bounds. A large condition number means that the matrix is ill-conditioned and difficult to invert. Using simulations, Ledoit and Wolf (2004a) show that the condition number of the sample covariance matrix is always larger than that of the true covariance matrix, and ceteris paribus, the condition number tends to increase with N/T.

[5]Black and Litterman (1992), Chan, Karceski, and Lakonishok (1999), Jagannathan and Ma (2003), and Ledoit and Wolf (2004a), (2004b) provide similar perspectives.

toward 0 to curtail extreme hedge positions, and ii) to drop redundant stocks in each hedge portfolio (subset selection). With an effective use of subset selection, the proposed mean variance optimizer, $\Psi \equiv \Sigma^{-1}$, becomes sparse, meaning that some of its off-diagonal elements are set to 0.

To promote shrinkage and subset selection (sparsity), we estimate Ψ by quasi-maximum likelihood with a constraint on the sum of the absolute values (i.e., l_1 norm) of its off-diagonal elements, similarly to Yuan and Lin (2007) and Rothman, Bickel, Levina, and Zhu (2008). We employ the "graphical lasso" (glasso) algorithm of Friedman, Hastie, and Tibshirani (2008) to solve the estimation problem.[6] Intuitively, this estimation methodology works as follows: The constraint imposes a penalty on the overall size of the hedge trades, thus shrinking each trade size toward 0. Meanwhile, different stocks compete with each other to enter the hedge portfolios. The methodology turns off the jth stock in the ith hedged portfolio if the marginal gain from retaining the stock does not meet a threshold. In effect, this approach restricts the (i,j)th and (j,i)th elements of the inverse covariance matrix to be exactly 0, while shrinking other elements toward 0 (the so-called "soft thresholding"). In this way, we encourage shrinkage and sparsity simultaneously to obtain the proposed mean variance optimizer (the glasso estimator of the inverse covariance matrix), $\hat{\Psi}$.

Aiming for risk reduction, we examine the out-of-sample performance of the proposed mean variance optimizer in a few data sets with N/T ratios ranging between 0.4 and 1.23.[7] Our data sets include i) samples of well-diversified portfolios (publicly available by courtesy of Ken French on his Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)) to facilitate comparison with other optimizers, and ii) 100 samples of 100 individual stocks randomly selected from the New York Stock Exchange (NYSE)/American Stock Exchange (AMEX) universe. The gains from shrinkage and subset selection promoted in our approach should be more evident for portfolios of individual stocks, among which the hedging relations are much noisier and less reliable than among those of already-diversified portfolios.

Empirically, out-of-sample portfolio risk reduction from the proposed mean variance optimizer $\hat{\Psi}$ is substantial, especially for the random portfolios of individual stocks. The global minimum variance (GMV) portfolio formed on the proposed optimizer $\hat{\Psi}$ achieves significant out-of-sample risk reduction over any of the competing models considered here.[8] The reduction in the out-of-sample portfolio risk is also accompanied by a high and stable level of Sharpe ratios and certainty equivalent returns adjusted for turnover/transaction costs.

[6]Friedman et al. (2008) make their glasso package available at http://cran.r-project.org. We estimate Ψ by setting "penalize.diagonal=FALSE" (i.e., restricting the off-diagonal elements of $\hat{\Psi}$ but not the diagonal elements) in the code.

[7]Because we are interested in situations in which portfolio optimization becomes very sensitive to estimation errors, our data sets contain more test assets than those studied by DeMiguel, Garlappi, Nogales, and Uppal (2009) and Tu and Zhou (2010), (2011).

[8]They include the naïve equal-weighted portfolio, the GMV portfolio formed on the sample covariance matrix, the GMV portfolio with no-short-sale constraint (Jagannathan and Ma (2003)), the GMV portfolio formed with the covariance matrix shrinking toward a constant correlation matrix by Ledoit and Wolf (2004b), and the GMV portfolio formed on an industry factor model (Chan et al. (1999)).

The rest of this paper is structured as follows: Section II elaborates on the framework proposed by Stevens (1998) to motivate our approach. Section III proposes an improved estimator of Ψ. After setting up the out-of-sample portfolio analysis in Section IV, Section V presents evidence on its out-of-sample performance. Section VI concludes.

II. The Role of Hedging in Portfolio Risk Minimization

In order to motivate our approach, we first discuss the role of hedge portfolios in mean variance portfolio optimization. Stevens (1998) shows that the inverse covariance matrix Σ^{-1} reveals the optimal hedging relations among stocks. Specifically, the ith row (or column) of Σ^{-1} is proportional to the ith stock's hedge portfolio. The ith hedge portfolio consists of taking i) a long position in the ith stock and ii) a short position in a portfolio of the other $N-1$ stocks that tracks the ith stock return to minimize the variance of the tracking error without any constraints. Each tracking portfolio can be estimated from the following regression:

$$r_{i,t} = \alpha_i + \sum_{k=1,k\neq i}^{N} \beta_{i|k} r_{k,t} + \varepsilon_{i,t}, \tag{1}$$

where $r_{i,t}$ denotes the ith stock return in period t; $\varepsilon_{i,t}$ is the unhedgeable component of $r_{i,t}$, whose variance is denoted by $v_i = \text{var}(\varepsilon_{i,t})$; and v_i is a measure of the ith stock's unhedgeable risk. The objective of each hedge regression is to minimize v_i, and consequently, we can view regression (1) as an OLS estimation problem. Stevens (1998) calls this a "regression hedge." When $\beta_{i|j}$ is different from 0 in population, it implies that the jth stock provides a marginal contribution to the hedge of the ith stock beyond the contributions of the other $N-2$ stocks in the portfolio.

Let us denote the inverse of the $N \times N$ covariance matrix Σ by $\Sigma^{-1} = \Psi = [\psi_{ij}]$, where ψ_{ij} denotes the (i,j)th element of Ψ. Then, Stevens (1998) establishes the following identity between Ψ and hedge regression (1):[9]

$$\psi_{ij} = \begin{cases} -\dfrac{\beta_{i|j}}{v_i} & \text{if } i \neq j \\ \dfrac{1}{v_i} & \text{if } i = j \end{cases}. \tag{2}$$

Remember that the hedge coefficient $\beta_{i|j}$ represents the ability of the jth stock to hedge the ith stock beyond the effects of the other $N-2$ stocks. We can view ψ_{ij} as a measure of marginal hedgeability between the ith and jth stocks conditional on all other stocks in the portfolio.[10] If the ith and jth stocks are uncorrelated

[9]For the symmetry $\psi_{ij} = \psi_{ji}$, see Stevens ((1998), fn. 3).

[10]Much earlier in the statistics literature, Dempster (1972) shows that ψ_{ij} represents the conditional dependence between the ith variable and the jth variable given all other variables in the system.

with each other after controlling for all other stocks, then $\beta_{i|j} = 0$, and hence $\psi_{ij} = 0$ must hold.

We can gain more useful insights about the inverse covariance matrix by looking at its ith row:

$$\Psi(i,\cdot) = \frac{1}{v_i}\left[-\beta_{i|1},\ldots,-\beta_{i|i-1},1,-\beta_{i|i+1},\ldots,-\beta_{i|N}\right]. \tag{3}$$

The identity in equation (2) and the expression for the hedge portfolio holdings in equation (3) form the basis for subsequent analysis. We can regard equation (3) as a vector of stock holdings in the ith stock's hedge portfolio. It involves a unit long position in the ith stock and a short position of the hedge portfolio constructed from regression (1). The hedge portfolio is a long–short portfolio whose holdings do not necessarily sum to a prefixed value (e.g., 1). However, holdings of each asset are scaled by $1/v_i$, meaning that the optimized portfolio takes a larger (smaller) position in a stock when its unhedgeable risk is smaller (larger). Thus the mean variance optimizer Ψ provides a detailed prescription about the hedge trades we need to make to minimize the portfolio risk.

III. An Improved Mean Variance Optimizer

A. A Sparse Inverse Covariance Matrix Estimator

The framework laid out in the preceding section is useful to understand the sources of potential large estimation errors in mean variance portfolio optimization. As shown earlier, each hedge regression in equation (1) contains a constant and $N-1$ stock returns as regressors. However, many of these stock returns are highly correlated. Furthermore, in many practical applications, the number of available historical returns to estimate the regression is not much larger than the number of stocks. Therefore, estimation of the hedge regression with practical sample size subjects the estimation to multicollinearity.[11] The undesirable consequences of multicollinearity are that the estimated hedge coefficients ($\hat{\beta}$'s) have such large estimation errors that the estimates are unstable from sample to sample, and hence too unreliable to be useful. By identity equation (2), this also implies that the off-diagonal elements of $\hat{\boldsymbol{\Psi}}$ are also susceptible to large estimation errors.

We conquer the multicollinearity by penalizing the l_1 norm of the parameters that need to be estimated.[12] Specifically, we consider the least absolute shrinkage

[11]Please see Judge, Griffiths, Hill, Lütkepohl, and Lee ((1980), chap. 12), for a detailed discussion of multicollinearity.

[12]An alternative solution to mitigate the multicollinearity problem is to increase the number of historical returns (T). For example, Jagannathan and Ma (2003) report a significant gain from using daily returns, instead of monthly returns, to estimate the sample covariance matrix. Still, in practice, many investment professionals (e.g., funds of funds) make portfolio decisions and recommendations using data updated monthly rather than daily. There are also investors who are constrained by the availability of historical returns.

and selection operator, or the "lasso" (Tibshirani (1996)) estimation problem for the hedge regression in equation (1):[13]

$$\text{(4)} \qquad \hat{\beta}_{i|k}^{\text{lasso}} = \arg\min_{\beta} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \sum_{k=1,k\neq i}^{N} \beta_{i|k} r_{k,t} \right)^2 + \gamma \sum_{k=1,k\neq i}^{N} \left| \beta_{i|k} \right| \right\}.$$

If the regressors were orthonormal, the lasso coefficient solving equation (4) would have the following relationship with the OLS coefficient $\hat{\beta}_{i|k}^{\text{OLS}}$:

$$\hat{\beta}_{i|k}^{\text{lasso}} = \text{sign}\left(\hat{\beta}_{i|k}^{\text{OLS}}\right)\left(\left|\hat{\beta}_{i|k}^{\text{OLS}}\right| - \gamma\right)_+ ; \; k = 1, \ldots, N, \; k \neq i,$$

where $(x)_+ = \max(x, 0)$, and the penalty parameter γ becomes the soft threshold. When the magnitude of an OLS point estimate is below the soft threshold, lasso solution $\hat{\beta}_{i|k}$ is set to 0. When the absolute value of an OLS point estimate is above the soft threshold, lasso solution $\hat{\beta}_{i|k}$ is shrunk by the magnitude of the soft threshold toward 0 but it never crosses 0 or alternates the sign. In this way, the lasso solution achieves i) shrinkage and ii) selection of stocks in each hedge regression.[14]

In practice, the regressors in hedge regression (1) are hardly orthonormal, as the regressors are usually correlated with each other. In this case, we can obtain $\hat{\beta}_{i|k}^{\text{lasso}}$ iteratively. Let us denote by $\tilde{\beta}_{i|k}^{(\gamma)}$ the current estimate for $\beta_{i|k}$ at penalty parameter γ. Then, Friedman, Hastie, Höfling, and Tibshirani (2007) show that, starting from $\hat{\beta}_{i|k}^{\text{OLS}}$, repeated iteration of the following (cycling through each regressor in turn until convergence) yields the lasso estimate;[15] that is, $\tilde{\beta}_{i|k}^{(\gamma)}$ converges to a unique $\hat{\beta}_{i|k}^{\text{lasso}}$.

$$\text{(5)} \qquad \tilde{\beta}_{i|k}^{(\gamma)} = \mathbf{S}\left(\tilde{\beta}_{i|k}^{(\gamma)} + \sum_{t=1}^{T} \left\{ r_{k,t} \times \left(r_{i,t} - \sum_{j=1,j\neq i}^{N} \tilde{\beta}_{i|j}^{(\gamma)} r_{j,t} \right) \right\}, \gamma \right); \quad k = 1, \ldots, N, \; k \neq i,$$

where $\mathbf{S}(b, \gamma) = \text{sign}(b)(|b| - \gamma)_+$ is the soft-thresholding operator.

When $N > T$, each hedge regression has an infinite number of least-squares solutions, because its design matrix is rank deficient. In this case, however, lasso has only one unique solution as long as the regressors are continuous random variables, regardless of which specific numbers N and T are (Tibshirani (2013)). In practice we start the iteration from a least-squares estimate obtained with the Moore–Penrose pseudoinverse (via singular value decomposition), but other starting values will lead to the same solution.

Following Stevens's (1998) framework, one can apply lasso to each hedge regression to estimate each row (or column) of the inverse covariance matrix to

[13]To simplify the exposition, we assume here that the returns are standardized to have mean 0 and unit variance.

[14]Other applications of lasso in finance include that set forth by Huang and Shi (2011), who use it to predict excess bond returns from a large number of macroeconomic factors.

[15]For a given value of γ, the lasso estimation problem is a convex optimization problem for which any local optimum is a global optimum.

achieve shrinkage and variable selection. In fact, Meinshausen and Bühlmann (2006) pursue this methodology rigorously in a different context. Although they show that this methodology correctly estimates the nonzero elements of an inverse covariance matrix asymptotically under certain conditions, this approach suffers from an important drawback in practical applications. The row-by-row (or column-by-column) lasso estimation does not restrict the inverse covariance matrix to be positive definite or symmetric. To avoid this problem, the N hedge regressions should be estimated jointly as a group, rather than separately.

Therefore, we follow Yuan and Lin (2007), Banerjee, El Ghaoui, and d'Aspremont (2008), and Friedman et al. (2008) to estimate all elements of the inverse covariance matrix by quasi-maximum likelihood (QML) once for all, with a penalty on the l_1 norm of its off-diagonal elements. Specifically, we estimate the inverse covariance matrix (the mean variance optimizer) Σ^{-1} as Ψ that maximizes

$$\max_{\Psi=[\psi_{ij}]} \frac{T}{2}\ln(\det(\Psi)) - \frac{T}{2}\mathrm{trace}(\hat{S}\Psi) - \rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} |\psi_{ij}|, \tag{6}$$

where det and trace indicate matrix determinant and trace, respectively, and $\hat{S}$ is the sample covariance matrix. The regularization parameter $\rho \geq 0$ denotes the penalty on the l_1 norm of the off-diagonal elements of Ψ, $\sum_{i=1,i\neq j}^{N}\sum_{j=1,j\neq i}^{N}|\psi_{ij}|$. The estimator, $\hat{\Psi}_\rho$, depends on the regularization parameter ρ, as signified by the subscript. We discuss the choice of ρ when we evaluate out-of-sample performance. A larger value of ρ promotes more sparsity of $\hat{\Psi}_\rho$, whereas $\rho = 0$ makes the solution identical to the unconstrained QML solution.[16]

Friedman et al. (2008) demonstrate that QML problem (6) is equivalent to a N-coupled lasso problem.[17] We solve QML estimation problem (6) using their graphical lasso (glasso) algorithm.[18] The penalized QML solution, or the glasso estimator of the inverse covariance matrix, entails the soft-thresholding effect and promotes the sparsity of the mean variance optimizer. This may introduce potential misspecification biases (omitted regressors in hedge regressions), but it leads

[16]The "penalized" QML estimation in problem (6) is closely related to the model selection problem considered by Akaike (1974) and Schwarz (1978). They propose information criteria (AIC and BIC) to trade off between model misspecification bias and estimation errors in selecting from competing models. For example, according to AIC, the optimal inverse covariance matrix Ψ solves $\hat{\Psi}_{\mathrm{AIC}} = \arg\max(T/2)\ln(\det(\Psi)) - (T/2)\mathrm{trace}(\hat{S}\Psi) - \mathrm{card}(\Psi)$, where $\mathrm{card}(\Psi)$ is the number of nonzero elements of Ψ, whereas Schwarz's BIC uses $(\ln(T/2))\mathrm{card}(\Psi)$ instead of $\mathrm{card}(\Psi)$. The information criteria penalize the number of parameters $\mathrm{card}(\Psi)$ to discourage model complexity and overfitting. Unfortunately, when N is not small, the penalties on $\mathrm{card}(\Psi)$ make our estimation problem infeasible because Ψ can be sparse in $2^{n(n-1)/2}$ different ways. However, by replacing $\mathrm{card}(\Psi)$ or $(\ln(T/2))\mathrm{card}(\Psi)$ with the l_1 norm $\rho\sum_{i=1}^{N}\sum_{j\neq i}^{N}|\psi_{ij}|$, we can combine model selection with parameter estimation within a feasible convex optimization problem.

[17]Although we can view QML problem (6) as a way to solve the joint lasso estimation of N hedge regressions, we can also interpret it as a method of "covariance-regularization" (Witten and Tibshirani (2009)) to estimate N hedge regressions.

[18]This algorithm sweeps over each row $i = 1, 2, \ldots, N,\ 1, 2, \ldots, N,\ \ldots$, to solve individual lasso problems holding all the others fixed, which then proceeds to the next step until all coefficients converge. In the Appendix, we provide details of how the estimation problem can be mapped to the system of hedge portfolios introduced in Section II.

to lower estimation errors. Needless to say, it does not imply the sparsity of the covariance matrix $\Sigma \equiv \Psi^{-1}$ because all stock returns are still correlated with each other. As a matter of fact, Σ is hardly sparse when Ψ is sparse.

B. Relation to the Literature

The proposed mean variance optimizer $\hat{\Psi}$, or the glasso estimator of the inverse covariance matrix, belongs to a wide class of shrinkage estimators. As such, it shares the same objective as a few existing approaches, in that it shrinks the estimator from the unbiased estimator (the sample-based estimator) in the direction that reduces estimation errors.

For example, Jagannathan and Ma (2003) provide a rationale that "wrong constraints" (those leading to misspecifications) can help improve the out-of-sample performance of optimized portfolios through estimation error reduction. They further show that imposing no-short-sale constraints is equivalent to shrinking the estimated covariances. Chan et al. (1999) use a low-dimensional-factor structure on the covariance matrix to improve the optimized portfolio's out-of-sample performance. Fan, Fan, and Lv (2008) also report the advantage of employing a factor structure in optimal portfolio construction. Many commercial risk models (e.g., Morgan Stanley Capital International (MSCI) Barra, Northfield, etc.) also impose multivariate factor structures on the covariance matrix, although their model details are proprietary and unknown to us. Ledoit and Wolf (2003), (2004a), (2004b) shrink the sample covariance matrix toward a more parsimonious target matrix, such as a constant correlation matrix or a covariance matrix with a one-factor structure. Their main impetus for shrinking the sample covariance matrix is also to strike an optimal balance between the misspecification biases and the estimation errors.[19]

Our approach shares the same objective as that of these papers, but suggests a different route to achieve it. Specifically, our approach differs from existing ones in three respects. First, our approach imposes a structure directly on the inverse covariance matrix $\Psi \equiv \Sigma^{-1}$, rather than on the covariance matrix Σ first and then inverting it, as existing approaches do. Second, our aim to shrink the hedging relations among stocks differentiates our approach from existing ones. Third, our approach promotes subset selection and hence sparsity of the inverse covariance matrix.

The literature suggests that economic intuition is very useful in improving covariance estimates. In particular, industry classifications often dominate statistical factor models in predicting stock return comovements (e.g., Connor (1995), Chan et al. (1999), Bhojraj, Lee, and Oler (2003), and Chan, Lakonishok, and Swaminathan (2007)). Why, then, would we expect our glasso estimator (a purely statistical approach) to compare well with a structural factor model based on economic intuition (e.g., industry classification)?

[19]Recent contributions to the mean variance portfolio choice literature include DeMiguel, Garlappi, Nogales, and Uppal (2009), Tu and Zhou (2010), (2011), and Kirby and Ostdiek (2012a), (2012b), among others. Our sparse mean variance optimizer $\hat{\Psi}_\rho$ can be blended with the methods proposed by Tu and Zhou (2010), (2011) and Kirby and Ostdiek (2012a), (2012b).

Let us consider a spectral decomposition (or a principal component analysis) of Σ, $\Sigma = U\Lambda U'$, where Λ is a $N \times N$ diagonal matrix whose diagonal entries are eigenvalues (λ's) of Σ; that is, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N > 0$. The ith column of U, u_i ($i = 1, \ldots, N$), is the eigenvector corresponding to the ith eigenvalue ($u_i'u_i = 1,\ u_i'u_j = 0$ if $i \neq j$). We can express this decomposition as

$$\Sigma = \lambda_1 u_1 u_1' + \lambda_2 u_2 u_2' + \cdots + \lambda_N u_N u_N'.$$

This decomposition is helpful to understand the superior explanatory power of a structural model based on economic intuition (e.g., an industry factor model). We can improve our estimate of Σ most efficiently by focusing on a small number of dominant eigenvectors (principal components), say, u_1, u_2, and u_3. It is relatively easy to associate these dominant principal components (with some rotations) with industry-level return comovements. Consequently, it is possible to improve our estimate of Σ by imposing an industry factor structure.

However, this argument does not necessarily imply that an industry factor model also works well in estimating the inverse, Σ^{-1}. The spectral decomposition of $\Sigma^{-1} = U\Lambda^{-1}U'$ implies

$$\Sigma^{-1} = \frac{1}{\lambda_N} u_N u_N' + \frac{1}{\lambda_{N-1}} u_{N-1} u_{N-1}' + \cdots + \frac{1}{\lambda_1} u_1 u_1'.$$

The first principal component of Σ now has relatively small effects on Σ^{-1} because it is now associated with the smallest eigenvalue of Σ^{-1}. Therefore, the economic intuition about a small number of dominant eigenvectors of Σ does not necessarily help improve our estimate of Σ^{-1}. This simple decomposition illustrates why economic intuition is generally hard to come by when we work directly with the inverse covariance matrix itself.

Nevertheless, based on the insight of Stevens (1998), we are able to offer an economic intuition for our glasso estimator as a way to improve hedging relations among assets in the portfolio.

IV. Out-of-Sample Evaluation: Setup

A. Data Sets and Methodology

To test the out-of-sample performance of the proposed mean variance optimizer, $\hat{\Psi}_\rho$, we employ the following data sets listed in Table 1. Compared to recent studies (e.g., DeMiguel, Garlappi, Nogales, and Uppal (2009), Tu and Zhou (2010), (2011), and Kirby and Ostdiek (2012a), (2012b)), our data sets have more test assets (larger N/T) and hence have greater concerns for estimation errors.

We consider two broad groups of data sets. The first group consists of four data sets of portfolio returns that are easily available through the courtesy of Ken French on his Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). They include the following: 100 portfolios formed on size and book-to-market ratio (data set 1), Fama and French's (1997) 48 industry

TABLE 1

Data Description

Table 1 lists the various testing portfolios we consider. Column 1 gives the abbreviation used to refer to the testing portfolios in column 2, from either U.S. or international capital markets (column 3). Column 4 reports the number of assets in each data set, and column 5 reports the N/T ratio when $T = 120$. The sample period of each data set is shown in column 6. For IND48, we augment the data with an old industry classification (from July 1963 to Dec. 2004; used by DeMiguel, Garlappi, and Uppal (2009)) with those from a new one from Jan. 2005 to Dec. 2010. NON_US combines two data sets, INT_MKT and INT_VAL_GRO. INT_MKT is U.S. dollar returns of the value-weighted country market portfolios of the following 15 countries: Australia, Belgium, Canada, France, Germany, Hong Kong, Italy, Japan, the Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, and the United Kingdom. INT_VAL_GRO contains value and growth portfolios in each of the 15 countries using four valuation ratios: BM, EP, CEP, and DP. The value portfolios contain firms in the top 30% of a ratio, and the growth portfolios contain firms in the bottom 30%. We exclude the CEP-based value and growth portfolios for Norway, due to a large number of missing values. Individuals contains 100 data sets of 100 stocks randomly chosen from NYSE and AMEX.

	Descriptor	Description	Market	N	N/T	Time Period
Data Set	1	2	3	4	5	6
1	SZBM100	100 (10 × 10) portfolios formed on size and BM	U.S.	100	0.833	July 1963–Dec. 2010
2	IND48	48 industry portfolios	U.S.	48	0.400	July 1963–Dec. 2010
3	SZBM100 + IND48	Combination of SZBM100 and IND48	U.S.	148	1.233	July 1963–Dec. 2010
4	NON_US	Combination of INT_MKT and INT_VAL_GRO	Int'l	133	1.108	Jan. 1975–Dec. 2010
5	Individuals	100 stocks from NYSE and AMEX	U.S.	100	0.833	Jan. 1973–Dec. 2010

portfolios (data set 2),[20] the combination of the first two totaling 148 portfolios (data set 3), and 133 portfolios from 15 non-U.S. developed countries (NON_US) (data set 4).[21]

These data sets consist of large and well-diversified portfolios. Although widely used in current literature, the gains from shrinkage and subset selection (sparsity restriction) should be more evident for portfolios of individual stocks among which the hedging relations are more difficult to estimate. Therefore, we consider the second group of our test assets, which consists of 100 data sets, each of which contains 100 individual stocks randomly chosen from the NYSE and AMEX universe.[22] Specifically, at the beginning of Jan. 1983, we draw individual stocks from NYSE and AMEX with stock prices no less than $5, and with price records available for the previous 120 months. We follow Jagannathan and Ma's (2003) method closely and replace missing returns with the value-weighted market returns in the testing period. We randomly choose 100 individual stocks at a time for 100 times to form 100 test assets (data set 5).

[20]For the 48 industry portfolios (data set 2), we augment the data with an old industry classification (from July 1963 to Dec. 2004, used by DeMiguel, Garlappi, Nogales, and Uppal (2009)) with those from the new industry classification from Jan. 2005 to Dec. 2010.

[21]NON_US (data set 4) includes portfolios from the following 15 developed countries: Australia, Belgium, Canada, France, Germany, Hong Kong, Italy, Japan, the Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, and the United Kingdom. For each country, we use the value-weighted market portfolio as well as the value and growth portfolios formed according to the following four valuation ratios: book/market (BM), earnings/price (EP), cash earnings/price (CEP), and dividend/price (DP). For each valuation ratio, the value portfolio is formed by the stocks in the top 30% of the ratio, and the growth portfolio is formed by the stocks in the bottom 30% of the ratio. For Norway, we exclude the value and growth portfolios based on CEP, due to a large number of missing values. In total, we have 133 portfolios in NON_US.

[22]We consider 100 portfolios of 100 individual stocks simultaneously over the out-of-sample testing period, as we are interested in evaluating the portfolio turnover and turnover-cost-adjusted certainty equivalent returns. In tables we report average quantities of the 100 portfolios.

We primarily focus on the out-of-sample performance of the minimum variance portfolio of risky assets (stocks). Because the minimum variance locus must be constructed from risky assets only and without risk-free ones, this portfolio also requires full initial investments. Therefore, our out-of-sample analysis excludes spread assets and imposes the usual constraint $\mathbf{1}_N'w = 1$, where $\mathbf{1}_N$ denotes the $N \times 1$ vector of 1s.[23]

We evaluate the out-of-sample portfolio performance using the standard "rolling-horizon" approach. In each month t, we construct the GMV portfolios using the past 120 months (10 years) of stock returns (the "estimation window," $T = 120$). Next, we hold such portfolios for 1 month and calculate the portfolio returns in month $t + 1$ out-of-sample. We continue this process by adding the return for the next period in the data set and dropping the earliest return from the estimation window. The choice of the rolling estimation window size, $T = 120$, follows the standard practice in the literature.[24] Meanwhile, we are also interested in the case of large N/T, in which a robust optimizer is called for. Therefore, our analysis involves data sets that have at least 100 assets (data sets 1, 3, 4, 5), and also those with $N > T$ in which the sample covariance matrix is singular (data sets 3, 4). Column 5 of Table 2 reports the N/T ratios of our data sets. For a given regularization parameter ρ, we can obtain the unique mean variance optimizer $\hat{\Psi}_\rho$ even when $N > T$. (See Section III.)

TABLE 2

Out-of-Sample Periods, Regularization Parameter, and Sparsity for Estimated $\hat{\Psi}_\rho$

For each data set, Table 2 reports: training period (column 1); testing period (column 2); number of testing periods (months) for an out-of-sample analysis (column 3); regularization parameter ρ chosen from the training period (column 4); degree of sparsity (column 5), which is the average percentage of 0 off-diagonal elements in the estimated inverse covariance matrix $\hat{\Psi}_\rho$.

		Out-of-Sample Analysis Period			Proposed Optimizer: $\hat{\Psi}_\rho$	
		Training Period	Testing Period	T_f	ρ	Sparsity
Data Set	Descriptor	1	2	3	4	5
1	SZBM100	July 1973–June 1983	July 1983–Dec. 2010	330	1.7	45.0%
2	IND48	July 1973–June 1983	July 1983–Dec. 2010	330	1.3	32.2%
3	SZBM100 + IND48	July 1973–June 1983	July 1983–Dec. 2010	330	1.9	47.1%
4	NON_US	Jan. 1985–Dec. 1994	Jan. 1995–Dec. 2010	192	0.8	44.0%
5	Individuals	Jan. 1983–Dec. 1992	Jan. 1993–Dec. 2010	216	5.9	32.4%

B. Choice of the Regularization Parameter ρ

Our proposed $\hat{\Psi}_\rho$ depends on the regularization parameter ρ (expression (6)). Because choosing the parameter after observing out-of-sample performance

[23]Of course, if our primary focus is on the Sharpe ratio rather than the portfolio risk, then it is certainly interesting to release the $\mathbf{1}_N'w = 1$ constraint and incorporate various spread assets into analysis. However, our current paper focuses on the out-of-sample performance of $\hat{\Psi}_\rho$ and abstracts away from the prediction of expected returns.

[24]To conserve space, we report the results only for $T = 120$. We have also conducted an analysis using a shorter estimation window of $T = 60$, and results are available from the authors. In fact, with $T = 60$, the proposed estimator achieves more significant gains in forecasting future covariances and reducing out-of-sample risk over alternative methods.

induces a look-ahead bias and thus is not appropriate, we fix the first 120-month in-sample period before the out-of-sample testing period. We use the in-sample period to search for the value of ρ that maximizes the predictive likelihood using a grid with increments of 0.1. Then we adhere to this choice throughout the out-of-sample "testing period" after burning in the first 120-months "training period." Our approach is simple and conservative, but the optimizer can deliver consistent performance when the optimal value of ρ remains stable over time.[25]

For each data set, columns 1 and 2 of Table 2 summarize the out-of-sample training and testing periods. For example, the out-of-sample testing period starts in July 1983 for the first 3 data sets and in Jan. 1995 for the non-U.S. developed countries (data set 4) and in Jan. 1993 for the 100 randomized samples of 100 individual stocks (data set 5). The total number of testing periods is 330 months for the first 3 data sets, 192 months for those of the non-U.S. developed countries (data set 4), and 216 months for the portfolios of individual stocks.

C. Proposed Method and Alternatives

Our empirical investigation focuses on the GMV portfolio, because this portfolio depends only on the estimator of the inverse covariance matrix $\hat{\Sigma}^{-1}$ but not on expected returns. For a given estimator of the inverse covariance matrix, $\hat{\Sigma}^{-1}$, the GMV portfolio is

$$w_{\text{GMV}} \quad = \quad \frac{1}{\mathbf{1}_N'\hat{\Sigma}^{-1}\mathbf{1}_N}\hat{\Sigma}^{-1}\mathbf{1}_N.$$

By replacing $\hat{\Sigma}^{-1}$ with the proposed sparse optimizer $\hat{\Psi}_\rho$, we propose the GMV portfolio: $w_{\text{GMV-}\hat{\Psi}_\rho} = (\mathbf{1}_N'\hat{\Psi}_\rho\mathbf{1}_N)^{-1}\hat{\Psi}_\rho\mathbf{1}_N$. We denote this portfolio by GMV-$\hat{\Psi}_\rho$.

For all of our test assets (data sets 1–4), we compare the portfolio performance of GMV-$\hat{\Psi}_\rho$ with those of the following alternative portfolios:

- The sample-based GMV portfolio, denoted by GMV-$\hat{S}^{-1}$. Its portfolio weights are summarized by $w_{\text{GMV-}\hat{S}^{-1}} = (\mathbf{1}_N'\hat{S}^{-1}\mathbf{1}_N)^{-1}\hat{S}^{-1}\mathbf{1}_N$.

- The equal-weighted (1/*N*) portfolio, $w_{\text{EW}} = (1/N)\,\mathbf{1}_N = (\mathbf{1}_N'\mathbf{1}_N)^{-1}\mathbf{1}_N$. EW (1/*N*) does not require any estimation and hence is free of estimation errors. Its strong and stable out-of-sample performance is well known (see DeMiguel, Garlappi, and Uppal (2009) for a recent review).

- The sample-based GMV portfolio with no-short-sale constraint such that every single portfolio weight has to be nonnegative, as proposed by Jagannathan and Ma (2003). We denote this portfolio by GMV-JM. Specifically, GMV-JM minimizes $w'\hat{S}w$ subject to $1_N'w = 1$ and $w_i \geq 0$ for $i = 1, \ldots, N$, where w_i denotes ith element of w. $\hat{S}$ is the sample covariance matrix.

[25]Given our "rolling-horizon" approach, it is natural to ask if we can improve the portfolio performance by reestimating the tuning parameter ρ period by period. We do not adopt this approach because it is computationally intensive, particularly for individual stocks. Still, to examine the potential benefits by updating the ρ, we experiment with updating ρ only once in the middle of the testing period. Overall, we find a fairly obvious benefit of updating ρ even only once, as it leads to lower out-of-sample portfolio risk.

- The GMV portfolio constructed from Ledoit and Wolf's (2004b) shrinkage estimator, $\hat{\Sigma}_{\text{LW}}^{-1}$. We call this portfolio GMV-LW. Its portfolio weights are summarized by $w_{\text{LW}} = (\mathbf{1}_N' \hat{\Sigma}_{\text{LW}}^{-1} \mathbf{1}_N)^{-1} \hat{\Sigma}_{\text{LW}}^{-1} \mathbf{1}_N$.

For the 100 randomized samples of 100 individual stocks (data set 5), we also consider the GMV portfolio based on an industry factor model. Intuitively, individual stocks are more highly correlated in the same industry than in different ones, because they are likely to be affected by common industry-wide news. Among various methods, a priori industry classifications often dominate more sophisticated statistical approaches in grouping stocks according to their homogeneity in stock price movements (e.g., Connor (1995), Chan et al. (1999), Bhojraj et al. (2003), and Chan et al. (2007)). For portfolios of individual stocks, therefore, a predicted covariance matrix from an industry factor model (based on industry affiliation) should serve as a natural benchmark for evaluating the out-of-sample performance of the proposed mean variance optimizer.

For each of the 100 randomized samples of 100 individual stocks, we implement an industry factor model as follows. We assign each stock to one of the 30 industry groups on Ken French's Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). With the industry classification, the covariance for any two stocks is given by the sample covariance of the corresponding industry portfolio (value-weighted) returns calculated from the rolling window of 120 months. The diagonal elements of the covariance matrix are the sample variances of 100 individual stock returns estimated from the 120-month rolling window.[26] We use Σ_{IND} to denote the covariance matrix from the industry factor model. We then construct the GMV portfolio, $w_{\text{IND}} = (\mathbf{1}_N' \hat{\Sigma}_{\text{IND}}^{-1} \mathbf{1}_N)^{-1} \hat{\Sigma}_{\text{IND}}^{-1} \mathbf{1}_N$, which we denote by GMV-IND. We also note that, due to its nature, GMV-IND is not applicable to data sets 1–4.

D. Descriptives for $\hat{\Psi}_\rho$

Column 4 of Table 2 reports the values of ρ chosen in the training period, which range from 0.8 (data set 4) to 5.9 (data set 5) across data sets.[27] Column 5 reports the "sparsity" of $\hat{\Psi}_\rho$, measured by the percentage of 0 off-diagonal elements. The time-series average of the sparsity ranges from 32.2% (data set 2) to 47.1% (data set 3), meaning that a significant fraction of the inverse covariance matrix is set to 0.

The sensitivity of the mean variance portfolio optimization problem $\hat{\Sigma} w = \hat{\mu}$ to small changes in the estimated inputs $(\hat{\Sigma}, \hat{\mu})$ can be quantified by the (2-norm) condition number of $\hat{\Sigma}$, which is the maximum singular value to the minimum singular value of the covariance matrix $\hat{\Sigma}$.[28] Moreover, the condition number of $\hat{\Sigma}$ and the condition number of $\hat{\Sigma}^{-1}$ are identical.

[26] Thus our industry factor model is based on each individual stock's industry affiliation, similar to the approach used by Connor (1995), Chan et al. (1999), and MSCI Barra, among others.

[27] Recall that we have 100 randomized samples of 100 individual stocks from AMEX and NYSE; 5.9 thus is the average value from these 100 samples.

[28] Let $w(\epsilon)$ be the solution to $(\hat{\Sigma} + \tilde{\Sigma})w(\epsilon) = \hat{\mu} + \tilde{\mu}$ with $||\tilde{\Sigma}|| \leq \epsilon||\hat{\Sigma}||$ and $||\tilde{\mu}|| \leq \epsilon||\hat{\mu}||$. If $\epsilon\kappa(\hat{\Sigma}) < 1$, where $\kappa(\hat{\Sigma})$ is the condition number of $\hat{\Sigma}$, then $\hat{\Sigma} + \tilde{\Sigma}$ is nonsingular and $(||w(\epsilon) - w||)/(||w||) \leq (2\epsilon\kappa(\hat{\Sigma}))/(1 - \epsilon\kappa(\hat{\Sigma}))$ for any vector norm and consistent matrix norm. See Golub and Van Loan (1996).

Table 3 reports the condition numbers for $\hat{\Psi}_\rho$ and $\hat{S}^{-1}$ as well as for Ledoit and Wolf's (2004b) shrunk covariance matrix (sample covariance matrix shrunk to constant correlation matrix),[29] the inverse of which is denoted by $\hat{\Sigma}_{\text{LW}}^{-1}$.[30] Clearly, the condition numbers are much smaller for $\hat{\Psi}_\rho$ than for $\hat{S}^{-1}$ in data sets 1–4. For the 100 randomized samples of 100 individual stocks (data set 5), $\hat{S}^{-1}$ tends to be highly ill-conditioned because the average condition number is very large (VLN). The average condition numbers for $\hat{\Sigma}_{\text{LW}}^{-1}$ and $\hat{\Sigma}_{\text{IND}}^{-1}$ are much more moderate at 769 and 568, respectively. Finally, $\hat{\Psi}_\rho$ achieves the best conditionedness, with an average condition number of 321.

TABLE 3

Condition Numbers for Estimated Inverse Covariance Matrices

For each data set, Table 3 reports condition numbers of the four estimators of the inverse covariance matrix: $\hat{\Psi}_\rho$ (Panel A) is the proposed sparse estimator with the regularization parameter ρ given in column 4 of Table 2; $\hat{S}^{-1}$ (Panel B) is the inverse of the sample covariance matrix; $\hat{\Sigma}_{\text{LW}}^{-1}$ (Panel C) is the inverse of the Ledoit–Wolf shrunk covariance matrix (sample covariance matrix shrunk to constant correlation matrix); $\hat{\Sigma}_{\text{IND}}^{-1}$ (Panel D) is the inverse of the covariance matrix from an industry factor model based on 30 industry classifications. The condition number for the (inverse) covariance matrix is the ratio of the maximal eigenvalue to the minimal eigenvalue. The condition number of a covariance matrix and that of its inverse are identical. A higher value of the condition number indicates greater numerical instability of the inverse operation. The condition number of $\hat{S}^{-1}$ is shown as "infinite (∞)" when $T < N$ ($T = 120$), as the minimal eigenvalue of $\hat{S}^{-1}$ is 0. NA means not applicable, and VLN means very large number.

			Panel A. $\hat{\Psi}_\rho$		*Panel B.* $\hat{S}^{-1}$		*Panel C.* $\hat{\Sigma}_{\text{LW}}^{-1}$		*Panel D.* $\hat{\Sigma}_{\text{IND}}^{-1}$	
Data Set	Descriptor	N/T	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	SZBM100	0.83	881	189	84,463	31,138	3,306	3,449	NA	NA
2	IND48	0.40	327	140	1,497	841	303	114	NA	NA
3	SZBM100 + IND48	1.23	1,153	252	∞	∞	4,279	4,740	NA	NA
4	NON_US	1.11	2,844	527	∞	∞	902	238	NA	NA
5	Individuals	0.83	321	86	VLN	VLN	709	432	568	165

Table 3 shows that $\hat{\Psi}_\rho$, $\hat{\Sigma}_{\text{LW}}^{-1}$, and $\hat{\Sigma}_{\text{IND}}^{-1}$ all help stabilize $\hat{S}^{-1}$ in situations when the sample covariance matrix is ill-conditioned or even noninvertible.

V. Out-of-Sample Evaluation: Evidence

A. Out-of-Sample Portfolio Risk Minimization

We are primarily interested in the ability of the proposed mean variance optimizer $\hat{\Psi}_\rho$ in reducing the out-of-sample portfolio risk. In the out-of-sample test, we first obtain the time series of out-of-sample returns for the six portfolios: GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, EW (1/*N*), GMV-JM, and GMV-LW, as well as GMV-IND (data set 5 only). Then we calculate the out-of-sample portfolio variances and

[29]The constant correlation matrix assumes that the correlations between any two stocks are just the average of all the pairwise correlations. The usefulness of the constant correlation matrix is well known in the literature and by industry practitioners. For example, Elton and Gruber (1973) and Elton, Gruber, and Urich (1978) show that the constant correlation model produces better forecasts of the future correlation matrix than those obtained from the market model or the sample correlation matrix.

[30]In constructing the Ledoit–Wolf shrunk covariance matrix $\hat{\Sigma}_{\text{LW}}$, we replace the sample covariance matrix ($\hat{S}$) with a weighted average of the sample covariance matrix and the shrinkage target. We then solve the optimal weight (shrinkage intensity) by minimizing the loss function of Ledoit and Wolf (2004a), which is represented by the Frobenius norm of the distance between the true covariance and shrinkage estimator.

standard deviations, the latter denoted by $\sigma_{\hat{\Psi}_\rho}$, $\sigma_{\hat{S}^{-1}}$, σ_{EW}, σ_{JM}, σ_{LW}, and σ_{IND}, respectively.

We then inspect if the proposed portfolio GMV-$\hat{\Psi}_\rho$ achieves a reduction in the out-of-sample portfolio risk as compared with the four alternatives. We test the null hypothesis of no difference in out-of-sample portfolio risk by computing bootstrap two-sided confidence intervals for $\sigma_{\hat{S}^{-1}} - \sigma_{\hat{\Psi}_\rho}$, $\sigma_{\text{EW}} - \sigma_{\hat{\Psi}_\rho}$, $\sigma_{\text{JM}} - \sigma_{\hat{\Psi}_\rho}$, $\sigma_{\text{LW}} - \sigma_{\hat{\Psi}_\rho}$, and $\sigma_{\text{IND}} - \sigma_{\hat{\Psi}_\rho}$, with the nominal level $1 - \alpha$.[31] Because portfolio standard deviations exhibit serial dependence, we apply Politis and Romano's (1994) "stationary bootstrap" (i.e., the block resampling with block lengths having a geometric distribution). We choose the mean block size for each data set using Politis and White's (2004) optimal block size.

The columns under Panel A of Table 4 report the out-of-sample variances of the six portfolios. The portfolio variance of GMV-$\hat{S}^{-1}$ enlarges substantially when N approaches T in magnitude. The portfolio cannot even be constructed when $N > T$, because $\hat{S}^{-1}$ does not exist then. On the other hand, EW (1/N), GMV-JM, and GMV-LW attain more moderate portfolio variance across data sets. For portfolios of individual stocks (data set 5), GMV-IND also achieves a significant risk reduction. Its out-of-sample portfolio risk is the second lowest, after that of GMV-$\hat{\Psi}_\rho$.

TABLE 4

Out-of-Sample Portfolio Risk (monthly)

For each data set, Table 4 reports the monthly variances of out-of-sample returns for the following six portfolios: GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, EW (1/N), GMV-JM, GMV-LW, and GMV-IND. $\sigma^2_{\hat{\Psi}_\rho}$, $\sigma^2_{\hat{S}^{-1}}$, σ^2_{EW}, σ^2_{JM}, σ^2_{LW}, and σ^2_{IND} denote their respective variances, calculated from the time series of out-of-sample returns in the testing period. $T < N$ indicates that the portfolio cannot be constructed due to the noninvertibility of the sample covariance matrix. The columns under Panel A tabulate the point estimates of the out-of-sample monthly return variances in %^{2}. Numbers below the variances (in parentheses) are the ranking of portfolio variance among the six portfolios: 1 indicates the smallest variance, and 5 (or 6 for data set 5) indicates the largest. The columns under Panel B tabulate the mean differences in the out-of-sample portfolio standard deviations between $\sigma_{\hat{\Psi}_\rho}$ and the other five alternative portfolio standard deviations, based on Politis and Romano's (1994) stationary bootstrap with the optimal expected block size as used by Politis and White (2004). We test the null of no difference indirectly by constructing the two-sided bootstrap intervals for the difference. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. NA means not applicable.

		Panel A. Return Variance (%^{2})						*Panel B.* $\sigma_{\text{ALT}} - \sigma_{\hat{\Psi}_\rho}$ *for* σ_{ALT} (%)				
Data Set	Descriptor	$\sigma^2_{\hat{\Psi}_\rho}$	$\sigma^2_{\hat{S}^{-1}}$	σ^2_{EW}	σ^2_{JM}	σ^2_{LW}	σ^2_{IND}	$\sigma_{\hat{S}^{-1}}$	σ_{EW}	σ_{JM}	σ_{LW}	σ_{IND}
1	SZBM100	13.30	65.30	25.43	58.57	21.98	NA	4.43***	1.40***	4.01*	1.04***	NA
		(1)	(5)	(3)	(4)	(2)						
2	IND48	12.45	17.59	22.88	16.22	13.15	NA	0.66***	1.25***	0.50*	0.10	NA
		(1)	(4)	(5)	(3)	(2)						
3	SZBM100 + IND48	10.70	$T < N$	24.12	16.43	12.74	NA	$T < N$	1.64***	0.78***	0.30	NA
		(1)	(5)	(4)	(3)	(2)						
4	NON_US	15.25	$T < N$	29.56	23.64	15.92	NA	$T < N$	1.53***	0.96*	0.08	NA
		(1)	(5)	(4)	(3)	(2)						
5	Individuals	10.10	19.44	23.79	13.40	11.36	11.01	1.21***	1.72***	0.45***	0.19***	0.14***
		(1)	(5)	(6)	(4)	(3)	(2)					

Furthermore, our proposed portfolio GMV-$\hat{\Psi}_\rho$ generates lower out-of-sample portfolio variance than alternative portfolios in all data sets. The columns under

[31] If this interval does not contain 0, then the null is rejected at the nominal level α (e.g., Ledoit and Wolf (2008)).

Panel A of Table 4 tabulate differences in out-of-sample portfolio risks (standard deviations) between GMV-$\hat{\Psi}_\rho$ and the alternative portfolios. Improvements in out-of-sample risk reduction from GMV-$\hat{S}^{-1}$, EW ($1/N$), or GMV-JM to GMV-$\hat{\Psi}_\rho$ are evident and statistically significant. GMV-$\hat{\Psi}_\rho$ also enhances significant out-of-sample risk reduction to a greater degree than GMV-LW does for data sets 1 (100 portfolios formed on size and BM) and 5 (individual stocks), but the difference is smaller in magnitude and insignificant for data sets 2–4. Still, GMV-$\hat{\Psi}_\rho$ is never significantly dominated in portfolio risk minimization. In the 100 randomized samples of 100 individual stocks (data set 5), GMV-$\hat{\Psi}_\rho$ achieves the lowest out-of-sample average portfolio risk, followed by GMV-IND, GMV-LW, and GMV-JM. Reductions in the out-of-sample portfolio risk from alternative portfolios to GMV-$\hat{\Psi}_\rho$ are all significant at the 1% level. Overall, GMV portfolios with regularized/shrunk covariance matrices (GMV-$\hat{\Psi}_\rho$, GMV-LW, GMV-IND) achieve further risk reduction beyond the naïve diversification rule, EW ($1/N$).

Note that we do not intend to claim the general superiority of $\hat{\Psi}_\rho$-based portfolios to the ones proposed by Jagannathan and Ma (2003) and Ledoit and Wolf (2003), (2004a), (2004b), or to the one based on an industry factor model. All of these portfolios involve some flexibility in implementation. For example, Jagannathan and Ma's (2003) portfolio strategy can accommodate various upper and lower bounds for individual portfolio weights. Ledoit and Wolf's (2003), (2004a), (2004b) shrinkage estimator depends on the shrinkage target and the shrinkage intensity parameter. Industry factor models depend on industry classification. The proposed sparse optimizer $\hat{\Psi}_\rho$ also involves a choice of the regularization parameter ρ. More generally, the relative performance of the portfolios may depend on data sets, estimation windows, and performance measures as well as testing methodologies. It suffices for us to show that the proposed sparse optimizer $\hat{\Psi}_\rho$ has a formidable ability to reduce out-of-sample portfolio risk, and its ability compares favorably with those of existing methods in many reasonable situations.

B. Out-of-Sample Sharpe Ratio

In principle, reduction in the portfolio risk can improve the Sharpe ratio if the mean returns remain the same. Although mean returns are susceptible to estimation errors, the Sharpe ratio is among the most widely used performance measures. It is also known that the GMV portfolio often achieves a higher Sharpe ratio than do other portfolios (e.g., Jorion (1985), (1986), DeMiguel, Garlappi, and Uppal (2009)). Therefore, we calculate out-of-sample Sharpe ratios for GMV-$\hat{\Psi}_\rho$ and the four alternative portfolios.

The columns under Panel A of Table 5 tabulate monthly Sharpe ratios. GMV-$\hat{\Psi}_\rho$ generates the highest or second-highest out-of-sample Sharpe ratios, except for the 48-industry portfolio (data set 2). For the U.S. portfolios (data sets 1–3), GMV-$\hat{\Psi}_\rho$ yields Sharpe ratios between 0.126 and 0.267 during the testing period between July 1983 and Dec. 2010. For comparison, the value-weighted U.S. market portfolio has a Sharpe ratio of 0.113 during the same period (not tabulated). For the non-U.S. developed countries (data set 4), GMV-$\hat{\Psi}_\rho$ attains the highest Sharpe ratio, followed by GMV-LW and GMV-JM.

TABLE 5

Out-of-Sample Sharpe Ratio (monthly)

Table 5 reports the out-of-sample monthly Sharpe ratios (SRs) for the following six portfolios: GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, EW (1/N), GMV-JM, GMV-LW, and GMV-IND. These portfolios are denoted by $\hat{\Psi}_\rho$, $\hat{S}^{-1}$, 1/N, JM, LW, and IND, respectively. $T < N$ indicates that the portfolio cannot be constructed due to the noninvertibility of the sample covariance matrix. The columns under Panel A report the level of monthly Sharpe ratios. Numbers below the Sharpe ratios (in parentheses) indicate the ranking of Sharpe ratios among the five portfolios: 1 indicates the highest Sharpe ratio, and 5 (or 6 for data set 5) indicates the lowest. The columns under Panel B report differences in Sharpe ratios between GMV-$\hat{\Psi}_\rho$ and the other five portfolios. We test the null of no difference in Sharpe ratios by constructing two-sided bootstrap intervals for the difference, using the studentized circular block bootstrap of Ledoit and Wolf (2008) with an optimal block size. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

		Panel A. Monthly SRs						*Panel B. Differences in SRs between GMV-$\hat{\Psi}_\rho$ and the Other Five Portfolios*				
Data Set	Descriptor	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	1/N	JM	LW	IND	$\hat{S}^{-1}$	1/N	JM	LW	IND
1	SZBM100	0.260 (1)	0.112 (4)	0.113 (3)	0.065 (5)	0.215 (2)	NA	0.148**	0.147***	0.195***	0.045	NA
2	IND48	0.126 (3)	0.070 (5)	0.129 (2)	0.151 (1)	0.119	NA	0.056**	−0.003	−0.025	0.007	NA
3	SZBM100 + IND48	0.267 (2)	$T < N$ (5)	0.119 (4)	0.123 (3)	0.315	NA	$T < N$	0.148***	0.144***	−0.048	NA
4	NON_US	0.273 (1)	$T < N$ (5)	0.138 (4)	0.142 (3)	0.227	NA	$T < N$	0.136**	0.132**	0.046**	NA
5	Individuals	0.147 (2)	0.100 (6)	0.181 (1)	0.101 (5)	0.105 (4)	0.138 (3)	0.047***	−0.034***	0.046***	0.042***	0.010**

For the 100 randomized samples of 100 individual stocks, GMV-$\hat{\Psi}_\rho$ is dominated by the EW (1/N) strategy (0.181); still, its Sharpe ratio of 0.147 is considerably above those of the other remaining strategies, whose Sharpe ratios range between 0.100 and 0.138. The high Sharpe ratio of the equal-weighted portfolio is likely attributable to the size effect and the short-term reversal effect (see Plyakha, Uppal, and Vilkov (2012)).

Detecting reliable differences in the Sharpe ratios is difficult due to large estimation errors in mean returns. Furthermore, in the presence of fat tails, serial correlation, and volatility clustering, the conventional Jobson and Korkie (1981a) test is not appropriate. We therefore employ Ledoit and Wolf's (2008) studentized circular block bootstrap (with Politis and White's (2004) optimal block size) to test the null hypothesis of no difference in Sharpe ratios.[32] Results are shown in Table 5 under Panel B. GMV-$\hat{\Psi}_\rho$ achieves significantly higher Sharpe ratios than GMV-$\hat{S}^{-1}$, EW (1/N), and GMV-JM in data sets 1, 3, and 4. Consistent with Plyakha et al. (2012), the equal-weighted portfolio EW (1/N) achieves significantly higher Sharpe ratios than any other method for individual stocks (data set 5). Still, GMV-$\hat{\Psi}_\rho$ achieves significantly higher Sharpe ratios than GMV-$\hat{S}^{-1}$, GMV-JM, GMV-LW, and GMV-IND in the data sets of individual stocks (data set 5). However, given large estimation errors in mean returns, we refrain from drawing strong conclusions for the differences in Sharpe ratios in this exercise.

C. Predicting Out-of-Sample Return Covariances

Let Σ and Σ^{-1} be the population covariance matrix and its inverse, and $\hat{\Sigma}_t^{-1}$ the time t estimate of the inverse covariance matrix, meaning a generic term for

[32]DeMiguel, Garlappi, Nogales, and Uppal (2009) also adopt this approach.

either $\hat{\Psi}_\rho$, $\hat{S}^{-1}$, or $\hat{\Sigma}_{\text{LW}}^{-1}$. An empirical measure of the ability of $\hat{\Sigma}_t^{-1}$ to predict out-of-sample covariances of stock returns is the log predictive Gaussian likelihood function (per observation):

$$L(\hat{\Sigma}^{-1}) = \frac{1}{T_f}\sum_{t=1}^{T_f} l_t(\hat{\Sigma}^{-1}) \tag{7}$$

$$\text{with} \quad l_t(\hat{\Sigma}^{-1}) = \ln\left(\det\left(\hat{\Sigma}_{t-1}^{-1}\right)\right) - \tilde{R}_t'\hat{\Sigma}_{t-1}^{-1}\tilde{R}_t,$$

where T_f is the total number of out-of-sample testing periods. $\tilde{R}_t = R_t - (1/T_f)\sum_{j=1}^{T_f} R_j$ denotes the demeaned return vector, which is calculated by subtracting the time-series mean from R_t during the out-of-sample testing period. We note that $L(\hat{\Sigma}^{-1})$ is twice the log likelihood function, as in the likelihood ratio test.

We test the differences in predictive likelihoods with the stationary bootstrap. Table 6 reports the values of $L(\hat{\Psi}_\rho)-L(\hat{S}^{-1})$ and $L(\hat{\Psi}_\rho)-L(\hat{\Sigma}_{\text{LW}}^{-1})$ for each data set, plus the value of $L(\hat{\Psi}_\rho)-L(\hat{\Sigma}_{\text{IND}}^{-1})$ for the data set of individual stocks (data set 5), along with the significance levels. The proposed glasso estimator (the sparse mean variance optimizer) $\hat{\Psi}_\rho$ delivers a significantly higher predictive likelihood than does the in-sample maximum likelihood estimator (MLE) $\hat{S}^{-1}$ in all data sets. The proposed estimator $\hat{\Psi}_\rho$ loses significantly to $\hat{\Sigma}_{\text{LW}}^{-1}$ for the non-U.S. developed countries (data set 4), but dominates $\hat{\Sigma}_{\text{LW}}^{-1}$ in data sets 1, 3, and 5. Our proposed estimator $\hat{\Psi}_\rho$ achieves particularly strong prediction of covariances for the data sets of individual stocks. Its predictive likelihood is significantly higher than those of the shrunk estimator $\hat{\Sigma}_{\text{LW}}^{-1}$ and the industry factor model $\hat{\Sigma}_{\text{IND}}^{-1}$.

TABLE 6

Out-of-Sample Risk Prediction: Differences in Log Predictive Likelihoods

Table 6 reports differences in the log predictive likelihoods, $L(\hat{\Psi}_\rho) - L(\hat{S}^{-1})$, $L(\hat{\Psi}_\rho) - L(\hat{\Sigma}_{\text{LW}}^{-1})$, and $L(\hat{\Psi}_\rho) - L(\Sigma_{\text{IND}}^{-1})$, where $L(\hat{\Psi}_\rho)$, $L(\hat{S}^{-1})$, $L(\hat{\Sigma}_{\text{LW}}^{-1})$, and $L(\hat{\Sigma}_{\text{IND}}^{-1})$ are the log predictive likelihood values per observation, corresponding to the inverse covariance matrices $\hat{\Psi}_\rho$, $\hat{S}^{-1}$, $\hat{\Sigma}_{\text{LW}}^{-1}$, and $\hat{\Sigma}_{\text{IND}}^{-1}$. $T < N$ means that $\hat{S}^{-1}$ does not exist. We test the null hypothesis of no difference indirectly by constructing two-sided bootstrap intervals for the difference using Politis and Romano's (1994) stationary bootstrap with the optimal expected block size as used by Politis and White (2004). ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively. NA means not applicable.

Data Set	Descriptor	$L(\hat{\Psi}_\rho) - L(\hat{S}^{-1})$	$L(\hat{\Psi}_\rho) - L(\hat{\Sigma}_{\text{LW}}^{-1})$	$L(\hat{\Psi}_\rho) - L(\Sigma_{\text{IND}}^{-1})$
1	SZBM100	102.08***	7.91***	NA
2	IND48	17.05***	1.68	NA
3	SZBM100 + IND48	$T < N$	31.94***	NA
4	NON_US	$T < N$	−29.90***	NA
5	Individuals	VLN	40.95***	25.07***

D. The Behavior of Optimized Portfolio Weights

The gist of the problem in portfolio risk minimization is that estimation errors in the sample covariance matrix yield unstable portfolio weights that can take extreme values. To address this problem, Ledoit and Wolf (2003), (2004a), (2004b) propose to shrink the sample covariance matrix. Jagannathan and Ma's (2003) procedure explicitly constrains the portfolio weights themselves, which is also equivalent to penalizing extreme values in the sample covariance matrix.

Our procedure tackles the problem by penalizing extreme coefficients in hedge regressions, thereby shrinking the inverse covariance matrix directly, rather than shrinking the covariance matrix and then inverting it. It is thus interesting to compare the behavior of the optimized portfolio weights of GMV-$\hat{\Psi}_\rho$ to those of GMV-$\hat{S}^{-1}$, GMV-JM, and GMV-LW.

We tabulate the minimum, 1st, 5th, 95th, and 99th percentiles, as well as the maximum of the portfolio weights for each method and for each data set, in Table 7. The table excludes the trivial case of EW (1/*N*). The last column of the table reports the Herfindahl index of optimized portfolio weights, $(1/T)\sum_{t=1}^{T}(\sum_{i=1}^{N}\hat{w}_{i,t}^2)$, where $\hat{w}_{i,t}$ here denotes the optimized weight on asset i at period t. The Herfindahl index takes the lowest value when all assets receive equal weights. The index takes larger values as the variability of weights across assets increases.

TABLE 7

Distribution of Portfolio Weights

Table 7 reports the distributions of monthly weights of the following four portfolios: GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, GMV-JM, and GMV-LW (we suppress the trivial case for the EW (1/*N*) portfolio). These portfolios are denoted by $\hat{\Psi}_\rho$, $\hat{S}^{-1}$, JM, and LW, respectively. For each asset, we pool all the monthly weights and report the minimum value; 1%, 5%, 95%, and 99% distributions; and maximum value. For individuals, we further pool all 100 samples. In the last column, we report the Herfindahl index of the weights.

Data Set	Descriptor	Portfolio	Minimum	1%	5%	95%	99%	Maximum	Herfindahl (%^{2})
1	SZBM100	GMV-$\hat{\Psi}_\rho$	−0.200	−0.129	−0.098	0.131	0.185	0.258	4,782
		GMV-$\hat{S}^{-1}$	−1.693	−0.820	−0.533	0.550	0.860	1.591	VLN
		GMV-JM	0.000	0.000	0.000	0.000	0.467	1.000	8,515
		GMV-LW	−0.457	−0.267	−0.155	0.192	0.341	0.676	12,354
2	IND48	GMV-$\hat{\Psi}_\rho$	−0.235	−0.151	−0.105	0.161	0.375	0.510	4,105
		GMV-$\hat{S}^{-1}$	−0.804	−0.383	−0.202	0.246	0.531	0.872	VLN
		GMV-JM	0.000	0.000	0.000	0.000	0.948	1.000	8,029
		GMV-LW	−0.200	−0.144	−0.103	0.175	0.488	0.655	5,182
3	SZBM100 + IND48	GMV-$\hat{\Psi}_\rho$	−0.158	−0.096	−0.064	0.085	0.136	0.252	3,169
		GMV-$\hat{S}^{-1}$				$T < N$			
		GMV-JM	0.000	0.000	0.000	0.000	0.167	1.000	7,891
		GMV-LW	−0.364	−0.172	−0.101	0.139	0.243	0.493	8,626
4	NON_US	GMV-$\hat{\Psi}_\rho$	−0.245	−0.150	−0.103	0.139	0.183	0.301	6,732
		GMV-$\hat{S}^{-1}$				$T < N$			
		GMV-JM	0.000	0.000	0.000	0.000	0.225	1.000	7,780
		GMV-LW	−0.135	−0.082	−0.055	0.102	0.151	0.223	2,979
5	Individuals	GMV-$\hat{\Psi}_\rho$	−0.075	−0.039	−0.024	0.060	0.144	0.315	1,250
		GMV-$\hat{S}^{-1}$	−28.364	−0.448	−0.141	0.184	0.517	27.720	VLN
		GMV-JM	0.000	0.000	0.000	0.001	0.396	0.965	5,922
		GMV-LW	−0.166	−0.077	−0.047	0.080	0.203	0.469	2,561

In data sets 1, 2, and 5, GMV-$\hat{S}^{-1}$ takes very large portfolio weights. For example, in the data set of individual stocks (data set 5), the maximum and minimum weights of GMV-$\hat{S}^{-1}$ can be as extreme as 2,772% and −2,836%. The very large numbers (VLNs) of the Herfindahl index are also consistent with the presence of extreme portfolio weights. Due to singularity, we cannot calculate these statistics for GMV-$\hat{S}^{-1}$ in data sets 3 and 4. Table 7 also reveals that the Jagannathan and Ma (2003) no-short-sale constrained portfolio (GMV-JM) tends to produce highly concentrated portfolios. The 95% value of the portfolio weight is essentially 0 in most data sets, meaning that only 5% or less of the assets receive positive portfolio weights. The maximum portfolio weight is 1 in data sets 1–4, meaning that the GMV-JM chooses only one asset in certain periods.

GMV-$\hat{\Psi}_\rho$ and GMV-LW, on the other hand, use both long and short positions of many assets. They produce similar distribution in optimized portfolio weights, but GMV-$\hat{\Psi}_\rho$ tends to be a little more aggressive in reducing extreme weights than does GMV-LW in all data sets except data set 4.

To see how the optimized weights from each procedure vary over time as the rolling horizon advances from one month to the next, Table 8 reports monthly turnover. The reported turnover can be interpreted as the average fraction of wealth traded in each rebalancing period. Not surprisingly, the EW (1/N) portfolio has the lowest turnover because the diversification rule involves only small contrarian rebalancing but does not involve any hedge trades. The no-short-sale constraint also keeps the turnover of GMV-JM low. On the contrary, the GMV-$\hat{\Psi}_\rho$ and GMV-LW portfolio strategies actively employ hedge trades to reduce portfolio risk, and hence entail higher turnover. However, these portfolios attain a much lower turnover than GMV-$\hat{S}^{-1}$. For GMV-$\hat{\Psi}_\rho$ and GMV-LW, monthly turnovers range between 0.160 and 0.886 and between 0.319 and 1.220, respectively. For the 100 portfolios of individual stocks (data set 5), the average monthly turnover of GMV-$\hat{\Psi}_\rho$ (0.160) is also lower than that of GMV-IND (0.281).

TABLE 8

Portfolio Turnover (monthly)

Table 8 reports turnover of the following six portfolios: GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, EW (1/N), GMV-JM, GMV-LW, and GMV-IND. These portfolios are denoted by $\hat{\Psi}_\rho$, $\hat{S}^{-1}$, 1/N, JM, LW, and IND, respectively. Calculation of the portfolio turnover follows DeMiguel, Garlappi, Nogales, and Uppal (2009). $T < N$ indicates that the portfolio cannot be constructed due to the noninvertibility of the sample covariance matrix.

Data Set	Descriptor	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	1/N	JM	LW	IND
1	SZBM100	0.534	7.970	0.025	0.282	1.220	NA
2	IND48	0.298	0.778	0.033	0.062	0.327	NA
3	SZBM100 + IND48	0.527	$T < N$	0.028	0.066	1.140	NA
4	NON_US	0.886	$T < N$	0.033	0.102	0.534	NA
5	Individuals	0.160	5.220	0.033	0.109	0.319	0.281

E. Economic Gains from Improved Portfolio Optimization

Mean variance portfolio optimization accomplishes portfolio risk reduction beyond the EW (1/N) diversification rule by exploiting the hedging relations among stocks. To assess the economic significance of portfolio risk reduction, however, we have to account for the effects of transaction costs, the amount of which grows rapidly with the hedge-trade-induced turnovers.

To this end, we calculate the annualized certainty equivalent excess return (CER) of each portfolio after subtracting its transaction cost (T_COST). Specifically, the T_COST-ADJUSTED_CER is

$$\text{T\_COST-ADJUSTED\_CER}_q = \hat{\mu}_q - \frac{\gamma}{2}\hat{\sigma}_q^2 - \text{T\_COST}_q,$$

where $\hat{\mu}_q$ and $\hat{\sigma}_q^2$ are the time-series (annualized) mean and variance of out-of-sample excess returns for portfolio q, respectively. Following Brandt (2009), we set the risk aversion coefficient γ to be 5.[33] T_COST$_q$ is the annualized transaction costs associated with portfolio q, measured by the annualized turnover multiplied by proportional transaction costs of 50 basis points (bps) per trade, following the standard practice in the recent literature. We can interpret the T_COST-ADJUSTED_CER$_q$ as the increase in the risk-free rate that an investor is willing to trade for a risky portfolio q after accounting for transaction costs. For example, suppose that a portfolio q has a T_COST-ADJUSTED_CER of 1%. In this case, the investor is indifferent between the portfolio q and an asset that guarantees a riskless return of 1% plus the risk-free rate. A higher value of the T_COST-ADJUSTED_CER indicates that the portfolio has a more desirable risk-return characteristic.

The columns of Table 9 report the T_COST-ADJUSTED_CERs of the five portfolios. GMV-$\hat{\Psi}_\rho$ clearly realizes positive economic gains over GMV-$\hat{S}^{-1}$ in all data sets. Although GMV-$\hat{\Psi}_\rho$ incurs much larger transaction costs than the EW ($1/N$) and the no-short-sale constrained GMV-JM portfolios, its gains from portfolio risk reduction still exceed the increased transaction costs for many data sets. Therefore, for an investor with a risk-aversion coefficient of $\gamma=5$, economic gains from the improved hedge trades to reduce portfolio risk are worth their transaction costs in many situations. Economic gains from our proposed estimator $\hat{\Psi}_\rho$ also compare favorably with those from the Ledoit and Wolf's (2004b) shrinkage estimator Σ_{LW}^{-1}, as GMV-$\hat{\Psi}_\rho$ achieves higher T_COST-ADJUSTED_CERs than GMV-LW in all data sets we consider here. The average T_COST-ADJUSTED_CER of GMV-$\hat{\Psi}_\rho$ is also higher than that of GMV-IND in the randomized samples of individual stock returns (data set 5).

TABLE 9

T_COST-ADJUSTED_CERs (annual %)

Table 9 reports the values of the T_COST-ADJUSTED_CERs of the following six portfolios: GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, EW ($1/N$), GMV-JM, GMV-LW, and GMV-IND, shown in annual percentage points. These portfolios are denoted by $\hat{\Psi}_\rho$, $\hat{S}^{-1}$, $1/N$, JM, LW, and IND, respectively. The transaction cost of each is calculated as 50 bps times monthly turnover times 12 (to annualize). For each data set, the ranking is also reported (in parentheses) below the value of the CERs. NA means not applicable.

Data Set	Descriptor	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	$1/N$	JM	LW	IND
1	SZBM100	4.17	−56.55	−0.94	−13.32	−1.83	NA
		(1)	(5)	(3)	(2)	(4)	
2	IND48	−0.19	−6.42	0.34	2.08	−0.73	NA
		(3)	(5)	(2)	(1)	(4)	
3	SZBM100 + IND48	4.11	$T < N$	−0.38	0.64	2.82	NA
		(1)	(5)	(4)	(3)	(2)	
4	NON_US	2.92	$T < N$	−0.09	0.55	2.89	NA
		(1)	(5)	(4)	(3)	(2)	
5	Individuals	1.54	−31.99	3.27	−0.24	−0.75	0.43
		(2)	(6)	(1)	(4)	(5)	(3)

F. Effects of No-Short-Sale Restriction

Next we examine how the nonnegativity constraint affects the performance of the GMV-$\hat{\Psi}_\rho$ portfolio. Green and Hollifield (1992) show that imposing the

[33]Tu and Zhou (2010), (2011) use $\gamma = 3$. Results for the case of $\gamma = 3$ are qualitatively similar to the results reported in this paper.

no-short-sale constraint inhibits the portfolio's ability to hedge the dominant systematic risk in minimizing portfolio risk. In response, Jagannathan and Ma (2003) demonstrate that imposing the no-short-sale constraint (nonnegativity constraint) does not necessarily hurt the portfolio performance because the estimated covariance matrix contains large measurement errors. In fact, if large estimation errors are still present in $\hat{\Psi}_\rho$, adding the no-short-sale restriction can help improve the performance of the GMV-$\hat{\Psi}_\rho$ portfolio. In the following, we use GMV-JM-$\hat{\Psi}_\rho$ to denote the GMV-$\hat{\Psi}_\rho$ portfolio with the no-short-sale restriction; that is, all portfolio weights of GMV-JM-$\hat{\Psi}_\rho$ are nonnegative. In other words, GMV-JM-$\hat{\Psi}_\rho$ replaces $\hat{S}^{-1}$ with $\hat{\Psi}_\rho$ in GMV-JM.

Table 10 compares the out-of-sample portfolio performance of GMV-$\hat{\Psi}_\rho$, GMV-JM, and GMV-JM-$\hat{\Psi}_\rho$. Obviously, the nonnegativity restriction limits the optimizer's ability to diversify portfolio risk. With the same restriction but using different inverse covariances, GMV-JM and GMV-JM-$\hat{\Psi}_\rho$ have very similar portfolio risk and Sharpe ratio out-of-sample, although GMV-JM-$\hat{\Psi}_\rho$ yields lower turnover than GMV-JM in all data sets. This result suggests that with the improved estimator of the inverse covariance matrix $\hat{\Psi}_\rho$ in place, the nonnegativity constraint no longer helps improve the portfolio performance.[34] It also indicates that the improved performance of GMV-$\hat{\Psi}_\rho$ is achieved through improved hedge trades that entail short positions in some constituents of the portfolio. By constraining the hedge trades, the no-short-sale restriction leads to lower portfolio performance once the estimation errors are contained. Put differently, the additional no-short-sale restriction does not help enhance the out-of-sample performance because it inhibits better use of hedge trades.

TABLE 10

Effects of No-Short-Sale Restriction

Table 10 compares GMV-$\hat{\Psi}_\rho$, GMV-JM, and GMV-JM-$\hat{\Psi}_\rho$ (denoted by $\hat{\Psi}_\rho$, JM, and JM-$\hat{\Psi}_\rho$, respectively) in terms of out-of-sample monthly return variance (Panel A), monthly Sharpe ratio (SR) (Panel B), and monthly turnover (Panel C). GMV-JM-$\hat{\Psi}_\rho$ replaces the inverse sample covariance matrix used in GMV-JM with the proposed sparse inverse covariance matrix $\hat{\Psi}_\rho$. In other words, GMV-JM-$\hat{\Psi}_\rho$ is GMV-$\hat{\Psi}_\rho$ with the no-short-sale restriction (e.g., Jagannathan and Ma (2003)).

		Panel A. Monthly Variance (%^{2})			*Panel B. Monthly SR*			*Panel C. Monthly Turnover*		
Data Set	Descriptor	GMV-$\hat{\Psi}_\rho$	GMV-JM	GMV-JM-$\hat{\Psi}_\rho$	GMV-$\hat{\Psi}_\rho$	GMV-JM	GMV-JM-$\hat{\Psi}_\rho$	GMV-$\hat{\Psi}_\rho$	GMV-JM	GMV-JM-$\hat{\Psi}_\rho$
1	SZBM100	13.30	58.57	57.90	0.260	0.065	0.047	0.534	0.282	0.282
2	IND48	12.45	16.22	16.45	0.126	0.151	0.151	0.298	0.062	0.058
3	SZBM100 + IND48	10.70	16.43	16.44	0.267	0.123	0.126	0.527	0.066	0.082
4	NON_US	15.25	23.64	23.51	0.273	0.142	0.145	0.886	0.102	0.096
5	Individuals	10.10	13.40	13.91	0.147	0.101	0.105	0.160	0.109	0.105

G. Detailed Look at the Randomized Samples of Individual Stock Returns

Data set 5 involves 100 runs with random draws of 100 individual stocks from the NYSE/AMEX universe. In preceding sections, we have reported average

[34]This is consistent with Jagannathan and Ma's (2003) observation that the nonnegativity constraint leads to a reduction in portfolio performance when we apply a factor structure or a shrinkage to the covariance matrix estimation, or when we use daily returns (low N/T) to estimate the sample covariance matrix.

estimates of the 100 randomized samples for out-of-sample portfolio risk and other estimates, for GMV-$\hat{\Psi}_\rho$ and alternative portfolio procedures. Although these average estimates are useful to assess the ability of each portfolio procedure in mitigating the effects of estimation risk, we can give a fuller picture by presenting the whole distribution of estimation results across the 100 samples. This exercise also allows us to examine the relation between the level of sparsity restriction on the estimated inverse covariance matrix and the magnitude of out-of-sample portfolio risk reduction ($\sigma_{\hat{S}^{-1}} - \sigma_{\hat{\Psi}_\rho}$) across the 100 samples. A high level of optimal sparsity (chosen by a grid search) can indicate that the sample (inverse) covariance matrix is noisier. If so, a lack of restrictions on hedge regressions would be more effective in clearing up estimation errors to achieve larger portfolio risk reduction.[35]

In Table 11, we first present the monthly mean, standard deviation, and minimum and maximum values of estimation results for i) the out-of-sample portfolio variance, ii) the Herfindahl index of optimized portfolio weights, and iii) the out-of-sample Sharpe ratio, for each of the six portfolio procedures: GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, EW ($1/N$), GMV-JM, GMV-LW, and GMV-IND. In particular, we report the fraction of runs for which the proposed procedure GMV-$\hat{\Psi}_\rho$ yields higher and lower portfolio variance, Herfindahl index of optimized weights, and Sharpe ratio as compared with alternative procedures.

In Panel A of Table 11, we can see that GMV-$\hat{\Psi}_\rho$ achieves lower out-of-sample portfolio risk than GMV-$\hat{S}^{-1}$ and EW ($1/N$) in all 100 runs. GMV-$\hat{\Psi}_\rho$ also achieves lower out-of-sample portfolio risk than GMV-JM, GMV-LW, and GMV-IND in 95 runs out of 100. This is a compelling evidence for the ability of GMV-$\hat{\Psi}_\rho$ to achieve significant risk reduction in portfolios of individual stocks, even beyond the industry factor model and the portfolios formed on the procedures of Jagannathan and Ma (2003) and Ledoit and Wolf (2004b). GMV-$\hat{\Psi}_\rho$ is also superior to GMV-JM, GMV-LW, and GMV-IND in curtailing extreme weights. Its Herfindahl index of optimized portfolio weights is lower than that of any of the alternative procedures in all 100 runs (Panel B).

In Panel C of Table 11, we turn to the Sharpe ratio and find that although the equal-weighted portfolio EW ($1/N$) achieves higher Sharpe ratio in 88 runs, GMV-$\hat{\Psi}_\rho$ still achieves higher Sharpe ratios in the majority of runs as compared with any of the other alternative procedures.

Finally, in Panels D and E of Table 11, we report the relation between the level of sparsity of $\hat{\Psi}_\rho$ and the magnitude of out-of-sample portfolio risk reduction ($\sigma_{\hat{S}^{-1}} - \sigma_{\hat{\Psi}_\rho}$) across the 100 randomized samples (each consisting of 100 individual stocks). In Panel D, we show that when we regress $\sigma_{\hat{S}^{-1}} - \sigma_{\hat{\Psi}_\rho}$ on the level of the sparsity, the slope coefficient is 0.112 (t-statistic = 3.30). That is, when the sparsity of $\hat{\Psi}_\rho$ is 1% higher in a sample than in other samples, GMV-$\hat{\Psi}_\rho$ tends to achieve 0.112% more portfolio risk reduction (per month) over GMV-$\hat{S}^{-1}$ in that sample than in other samples. In Panel E, we sort these 100 randomized samples by the sparsity level into three groups: bottom 30, middle 40, and top 30 subsamples. We observe that groups with higher sparsity levels tend to achieve greater out-of-sample portfolio risk reduction. Specifically, the average

[35]We thank an anonymous referee for this insight.

TABLE 11

Distribution of Realized Volatility, Sharpe Ratio and Weights, and Risk Reduction versus Sparsity for Individuals

Panels A–C of Table 11 compare six portfolios, GMV-$\hat{\Psi}_\rho$, GMV-$\hat{S}^{-1}$, EW(1/*N*), GMV-JM, GMV-LW, and GMV-IND, in terms of distributions of out-of-sample monthly return variance (%^{2}) (Panel A), Herfindahl index of optimized weights (%^{2}) (Panel B), and out-of-sample monthly Sharpe ratio (Panel C) over 100 randomized samples, each consisting of 100 individual stocks. The portfolios are denoted by $\hat{\Psi}_\rho$, $\hat{S}^{-1}$, 1/*N*, JM, LW, and IND, respectively. For each panel, we take the 100-sample average of the mean, standard deviation, minimum and maximum of realized monthly variance, monthly Sharpe ratio, and Herfindahl index of monthly weights for each randomized sample. "No. of GMV-$\hat{\Psi}_\rho > $ ALT" counts the number of times the portfolio GMV-$\hat{\Psi}_\rho$ has achieved either a higher realized volatility, Sharpe ratio, or Herfindahl index than each alternative portfolio does in the 100 randomized samples. Similarly, "No. of GMV-$\hat{\Psi}_\rho <$ ALT" counts the number of times the portfolio GMV-$\hat{\Psi}_\rho$ has achieved either a lower realized volatility, Sharpe ratio, or Herfindahl index than an alternative portfolio. Panels D and E report the relationship between the level of sparsity of $\hat{\Psi}_\rho$ and the magnitude of out-of-sample portfolio risk reduction ($\sigma_{\hat{S}-1} - \sigma_{\hat{\Psi}_\rho}$) across the 100 randomized samples. Panel D reports the results regressing risk reduction on sparsity, whereas Panel E groups the 100 randomized samples by the sparsity level into bottom 30, middle 40, and top 30 subsamples. We then report the average risk reduction within each subsample.

	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	1/*N*	JM	LW	IND
Panel A. Out-of-Sample Return Variance (monthly; %^{2})						
Mean	10.10	19.44	23.79	13.40	11.37	11.01
Standard deviation	2.23	6.09	1.19	4.69	2.67	2.56
Minimum	6.05	9.53	21.21	6.52	6.48	6.52
Maximum	14.92	58.09	26.91	31.51	17.17	15.87
No. of GMV-$\hat{\Psi}_\rho >$ ALT		0	0	5	5	5
No. of GMV-$\hat{\Psi}_\rho <$ ALT		100	100	95	95	95
Panel B. Herfindahl Index of Optimized Portfolio Weights (%^{2})						
Mean	1,250	VLN	100	5,922	2,561	1,909
Standard deviation	329	VLN	0	1,773	614	395
Minimum	807	VLN	100	3,247	1,625	1,356
Maximum	2,022	VLN	100	8,854	3,875	2,866
No. of GMV-$\hat{\Psi}_\rho >$ ALT		0	100	0	0	0
No. of GMV-$\hat{\Psi}_\rho <$ ALT		100	0	100	100	100
Panel C. Out-of-Sample Sharpe Ratio (monthly)						
Mean	0.147	0.100	0.181	0.101	0.115	0.138
Standard deviation	0.031	0.057	0.006	0.045	0.043	0.035
Minimum	0.065	−0.046	0.167	−0.012	0.019	0.036
Maximum	0.237	0.232	0.198	0.197	0.229	0.229
No. of GMV-$\hat{\Psi}_\rho >$ ALT		81	12	85	91	67
No. of GMV-$\hat{\Psi}_\rho <$ ALT		19	88	15	9	33

Panel D. Risk Reduction (monthly % $\sigma_{\hat{S}-1} - \sigma_{\hat{\Psi}_\rho}$) Regressed on Sparsity

Intercept	*t*-Statistic	Sparsity	*t*-Statistic	R^2
−2.198	(−2.13)	0.112	(3.30)	10%

Panel E. Risk Reduction (monthly % $\sigma_{\hat{S}-1} - \sigma_{\hat{\Psi}_\rho}$) across Sparsity Groups

Sparsity Group	Bottom 30%	Middle 40%	Top 30%
Risk reduction	1.078	1.105	1.468

risk reductions in the bottom 30, middle 40, and top 30 groups are 1.08%,1.11%, and 1.47% per month, respectively. These results suggest that our sparse mean variance optimizer $\hat{\Psi}_\rho$ tends to achieve greater portfolio risk reduction in samples that call for higher levels of sparsity restrictions.

VI. Conclusion

Appealing to the insight that the inverse covariance matrix prescribes the optimal hedging relations among stocks, we propose a method to reduce the estimation errors in the inverse covariance matrix by shrinking the estimated hedge portfolio weights. The proposed mean variance optimizer, or the glasso estimator of the inverse covariance matrix, is sparse, meaning that a significant fraction of its off-diagonal elements are 0.

We show that the proposed glasso estimator of the inverse covariance matrix accomplishes a significant and robust out-of-sample risk reduction, especially when the sample-based (maximum likelihood) estimator of the covariance matrix is ill-conditioned or singular. Furthermore, the out-of-sample performance of the GMV portfolio with the glasso estimator compares favorably to those of the equal-weighted portfolio, the no-short-sale constrained GMV portfolio (Jagannathan and Ma (2003)), and the GMV portfolio with shrunk covariance matrix (e.g., Ledoit and Wolf (2004b)), as well as the GMV portfolio using an industry factor model. The gains in the portfolio risk reduction from the glasso estimator are particularly strong in portfolios of individual stocks, among which the hedging relations tend to be more difficult to estimate than among portfolios.

These results support the initial motivation of our analysis: By mitigating estimation errors in the hedge portfolios, we can enhance the ability of the mean variance optimizer in reducing the out-of-sample portfolio risk.

Appendix. Glasso Algorithm and Hedge Regressions

This Appendix shows the relationship between the estimation of Ψ and constructing the system of hedge portfolios. Again the constrained quasi-maximum likelihood estimation problem is

$$\text{(A-1)} \qquad \max_{\Psi=[\psi_{ij}]} \frac{T}{2}\ln(\det(\Psi)) - \frac{T}{2}\text{trace}(\hat{S}\Psi) - \rho \sum_{i=1, i\neq j}^{N} \sum_{j=1, j\neq i}^{N} |\psi_{ij}|,$$

where $\hat{S}$ is the sample covariance matrix.

By constraining the quasi-maximum likelihood, the graphical lasso algorithm also imposes constraints on hedge regressions. Banerjee et al. (2008) show that problem (A-1) is convex and consider estimation as follows. Letting W be a perturbation of the sample estimator $\hat{S}$, they show that one can solve the problem by optimizing over each row and corresponding column of W. Suppose we rearrange the stocks so that the last row and column correspond to the stock one wants to hedge by other stocks. Then, partitioning W and $\hat{S}$,

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w'_{12} & w_{22} \end{bmatrix}, \qquad S = \begin{bmatrix} \hat{S}_{11} & \hat{s}_{12} \\ \hat{s}'_{12} & \hat{s}_{22} \end{bmatrix}.$$

Using convex duality, Banerjee et al. (2008) show that the dual problem of the previous equation turns out to be

$$\text{(A-2)} \qquad \min_{\beta} \left\{ \frac{1}{2} \left\| W_{11}^{1/2}\beta - b \right\|^2 + \rho \left\|\beta\right\|_1 \right\},$$

where $b = W_{11}^{-1/2}\hat{s}_{12}$ and $\beta = W_{11}^{-1}w_{12}$.

Now this dual problem exactly resembles a lasso least-squares problem (Tibshirani (1996)) applied on hedge regression (1) (Stevens (1998)). Without the constraint on its l_1 norm, β is exactly the vector of hedge coefficients (see equation (3)), the result of regressing the last stock return on all the previous $N - 1$ stocks (thus the expression $W_{11}^{-1}w_{12}$).

However, when the solution to the minimization problem is subject to the l_1 norm constraint, the lasso will shrink some element of the row vector $\boldsymbol{\beta}$ to 0. Again the lasso estimator contains bias but reduces estimation variation.

This duality further motivates Friedman et al. (2008) glasso algorithm, which recursively solves each row and updates the lasso problem with the pathwise coordinate descent algorithm until convergence, ensuring the symmetry of $\hat{\Psi}$. For the equivalence between the two problems, please see Banerjee et al. (2008) or Friedman et al. (2008).

References

Akaike, H. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control*, 19 (1974), 716–723.

Banerjee, O.; L. El Ghaoui; and A. d'Aspremont. "Model Selection through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data." *Journal of Machine Learning Research*, 9 (2008), 485–516.

Best, M., and R. Grauer. "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results." *Review of Financial Studies*, 4 (1991), 315–342.

Bhojraj, S.; C. Lee; and D. Oler. "What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research." *Journal of Accounting Research*, 41 (2003), 745–774.

Black, F., and R. Litterman. "Global Portfolio Optimization." *Financial Analysts Journal*, 48 (1992), 28–43.

Brandt, M. "Portfolio Choice Problems." In *Handbook of Financial Econometrics*, Y. Ait-Sahalia and L. Hansen, eds. Amsterdam: North-Holland (2009).

Chan, L.; J. Karceski; and J. Lakonishok. "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model." *Review of Financial Studies*, 12 (1999), 937–974.

Chan, L.; J. Lakonishok; and B. Swaminathan. "Industry Classifications and Return Comovement." *Financial Analysts Journal*, 63 (2007), 56–70.

Connor, G. "The Three Types of Factor Models: A Comparison of Their Explanatory Power." *Financial Analysts Journal*, 51 (1995), 42–46.

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

DeMiguel, V.; L. Garlappi; and R. Uppal. "Optimal versus Naïve Diversification: How Inefficient Is the 1/*N* Portfolio Strategy?" *Review of Financial Studies*, 22 (2009), 1915–1953.

Dempster, A. "Covariance Selection." *Biometrics*, 28 (1972), 157–175.

Elton, E., and M. Gruber. "Estimating the Dependence Structure of Share Prices—Implications for Portfolio Selection." *Journal of Finance*, 28 (1973), 1203–1232.

Elton, E.; M. Gruber; and T. Urich. "Are Betas Best?" *Journal of Finance*, 33 (1978), 1375–1384.

Fama, E., and K. French. "Industry Costs of Equity." *Journal of Financial Economics*, 43 (1997), 153–193.

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

Friedman, J.; T. Hastie; H. Höfling; and R. Tibshirani. "Pathwise Coordinate Optimization." *Annals of Applied Statistics*, 2 (2007), 302–332.

Friedman, J.; T. Hastie; and R. Tibshirani. "Sparse Inverse Covariance Estimation with the Graphical Lasso." *Biostatistics*, 9 (2008), 432–441.

Golub, G., and C. van Loan. *Matrix Computations.* Baltimore, MD: Johns Hopkins University Press (1996).

Green, R., and B. Hollifield. "When Will Mean-Variance Efficient Portfolios Be Well Diversified?" *Journal of Finance*, 47 (1992), 1785–1809.

Huang, J., and Z. Shi. "Determinants of Bond Risk Premia." Working Paper, Penn State University (2011).

Jagannathan, R., and T. Ma. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance*, 58 (2003), 1651–1683.

Jobson, J. D., and B. Korkie. "Performance Hypothesis Testing with the Sharpe and Treynor Measures." *Journal of Finance*, 36 (1981a), 889–908.

Jobson, J. D., and B. Korkie. "Putting Markowitz Theory to Work." *Journal of Portfolio Management*, 7 (1981b), 70–74.

Jorion, P. "International Portfolio Diversification with Estimation Risk." *Journal of Business*, 58 (1985), 259–278.

Jorion, P. "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis*, 21 (1986), 279–292.
Judge, G.; W. Griffiths; C. Hill; H. Lütkepohl; and T. Lee. *The Theory and Practice of Econometrics*. Hoboken, NJ: John Wiley & Sons (1980).
Kan, R., and G. Zhou. "Optimal Portfolio Choice with Parameter Uncertainty." *Journal of Financial and Quantitative Analysis*, 42 (2007), 621–656.
Kirby, C., and B. Ostdiek. "It's All in the Timing: Simple Active Portfolio Strategies That Outperform Naïve Diversification." *Journal of Financial and Quantitative Analysis*, 47 (2012a), 437–467.
Kirby, C., and B. Ostdiek. "Optimizing the Performance of Sample Mean-Variance Efficient Portfolios." Working Paper, Rice University (2012b).
Ledoit, O., and M. Wolf. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection." *Journal of Empirical Finance*, 10 (2003), 603–621.
Ledoit, O., and M. Wolf. "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." *Journal of Multivariate Analysis*, 88 (2004a), 365–411.
Ledoit, O., and M. Wolf. "Honey, I Shrunk the Sample Covariance Matrix." *Journal of Portfolio Management*, 30 (2004b), 110–119.
Ledoit, O., and M. Wolf. "Robust Performance Hypothesis Testing with the Sharpe Ratio." *Journal of Empirical Finance*, 15 (2008), 850–859.
Markowitz, H. "Portfolio Selection." *Journal of Finance*, 7 (1952), 77–91.
Meinshausen, N., and P. Bühlmann. "High Dimensional Graphs and Variable Selection with the Lasso." *Annals of Statistics*, 34 (2006), 1436–1462.
Michaud, R. "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal*, 45 (1989), 31–42.
Plyakha, Y.; R. Uppal.; and G. Vilkov. "Why Does an Equal-Weighted Portfolio Outperform Value- and Price-Weighted Portfolios?" Working Paper, EDHEC Business School (2012).
Politis, D., and J. Romano. "The Stationary Bootstrap." *Journal of the American Statistical Association*, 89 (1994), 1303–1313.
Politis, D., and H. White. "Automatic Block-Length Selection for the Dependent Bootstrap." *Econometric Reviews*, 23 (2004), 53–70.
Rothman, A.; P. Bickel; E. Levina; and J. Zhu. "Sparse Permutation Invariant Covariance Estimation." *Electronic Journal of Statistics*, 2 (2008), 494–515.
Schwarz, G. "Estimating the Dimension of a Model." *Annals of Statistics*, 6 (1978), 461–464.
Stevens, G. "On the Inverse of the Covariance Matrix in Portfolio Analysis." *Journal of Finance*, 53 (1998), 1821–1827.
Tibshirani, R. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society, Series B (Methodological)*, 58 (1996), 267–288.
Tibshirani, R. "The Lasso Problem and Uniqueness." *Electronic Journal of Statistics*, 7 (2013), 1456–1490.
Tu, J., and G. Zhou. "Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty." *Journal of Financial and Quantitative Analysis*, 45 (2010), 959–986.
Tu, J., and G. Zhou. "Markowitz Meets Talmud: A Combination of Sophisticated and Naïve Diversification Strategies." *Journal of Financial Economics*, 99 (2011), 204–215.
Witten, D., and R. Tibshirani. "Covariance-Regularized Regression and Classification for High Dimensional Problems." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 71 (2009), 615–636.
Yuan, M., and Y. Lin. "Model Selection and Estimation in the Gaussian Graphical Model." *Biometrika*, 94 (2007), 19–35.