College of Business Administration
University of Rhode Island

William A. Orme

WORKING PAPER SERIES

encouraging creative research

On Mean Variance Portfolio Optimiztion:
Improving Performance Through Better Use of Hedgin Relations

Yan Xu, Shingo Goto

2009/2010 No. 14

Office of the Dean
College of Business Administration
Ballentine Hall
7 Lippitt Road
Kingston, RI 02881
401-874-2337
www.cba.uri.edu

This working paper series is intended to facilitate discussion and encourage the exchange of ideas. Inclusion here does not preclude publication elsewhere. It is the original work of the author(s) and subject to copyright regulations.
On Mean Variance Portfolio Optimization: Improving Performance Through Better Use of Hedging Relations

Shingo Goto* Yan Xu†
University of South Carolina University of Rhode Island

November 2009
This version: February 2010

Abstract

In portfolio optimization, the inverse covariance matrix prescribes the hedge trades where a portfolio of stocks hedges each one with all the other stocks to minimize portfolio risk. In practice with finite samples, however, multicollinearity makes the hedge trades too unstable to be reliable. By reducing the number of stocks in each hedge trade to curb estimation errors, we motivate a “sparse” estimator of the inverse covariance matrix with multiple zero off-diagonal elements. Comparing favorably with those under no-short-sale constraints and using shrunk covariance matrix, a portfolio formed on this estimator achieves significant risk reduction out-of-sample. By improving hedge trades, the portfolio delivers higher certainty equivalent returns after transaction costs in a range of situations.

*Moore School of Business, University of South Carolina, 1705 College St., Columbia, SC 29208, USA; phone: +1 803 777 6644; fax +1 803 777 6876; email: shingo.goto@moore.sc.edu.

†College of Business Administration, University of Rhode Island; 7 Lippitt Road, Kingston, RI 02881-0802, USA; phone: +1 410 874 4190; fax: +1 410 874 4312; email: yan.xu@mail.uri.edu.
On Mean Variance Portfolio Optimization: Improving Performance Through Better Use of Hedging Relations

Abstract

In portfolio optimization, the inverse covariance matrix prescribes the hedge trades where a portfolio of stocks hedges each one with all the other stocks to minimize portfolio risk. In practice with finite samples, however, multicollinearity makes the hedge trades too unstable to be reliable. By reducing the number of stocks in each hedge trade to curb estimation errors, we motivate a “sparse” estimator of the inverse covariance matrix with multiple zero off-diagonal elements. Comparing favorably with those under no-short-sale constraints and using shrunk covariance matrix, a portfolio formed on this estimator achieves significant risk reduction out-of-sample. By improving hedge trades, the portfolio delivers higher certainty equivalent returns after transaction costs in a range of situations.
Mean variance portfolio optimization relies on covariances to transform expected returns into optimal portfolio weights. Consider a portfolio of \( N \) stocks. A portfolio manager uses an asset pricing model to predict a vector of expected returns, denoted by \( \mu \), and employs a risk model to predict the covariance matrix, denoted by \( \Sigma \).\(^1\) Then, it is well known that her optimal portfolio weights are summarized by

\[
w \propto \Sigma^{-1} \mu.
\]

In this familiar result, the inverse covariance matrix \( \Sigma^{-1} \) plays a pivotal role in transforming the return forecast \( \mu \) into the optimal portfolio weights \( w \). Therefore, we take the liberty of calling the inverse covariance matrix, \( \Psi \equiv \Sigma^{-1} \), the *mean variance optimizer*.\(^2\)

While simple and forceful in theory, the idea of mean variance portfolio optimization turns out to be surprisingly difficult to implement. In many real world situations, the available number of historical return observations per stock (\( T \)) is not much larger than the number of stocks (\( N \)).\(^3\) Then, the inverse covariance matrix (the mean variance optimizer) estimated

---

\(^1\)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.

\(^2\)Inverse covariance matrix is also called “precision matrix” or “concentration matrix” (Dempster (1972)).

\(^3\)For example, consider an active manager who runs a number of portfolios based on her proprietary sources of information (“signals”). Each signal portfolio tends not to have a large number of historical returns.
from the sample becomes very sensitive to small perturbations in the estimated covariances, making it unstable from one period to another. As a consequence, the mean variance optimizer tends to produce poor out-of-sample performance. The mean variance optimization tends to erode, rather than enhance, the gains from naïve diversification policies such as equal-weighting (e.g., Jobson and Korkie (1981a), DeMiguel, Garlappi, and Uppal (2009).) Michaud (1989) even calls the mean variance optimization the “error maximization.”

To regain our confidence in the mean variance optimization framework, we propose an improvement of the mean variance optimizer $\Psi$. Specifically, since stock returns are correlated, we can use each stock to hedge others. Motivated by the simple and sensible idea to improve hedging relations, our proposed estimator of $\Psi$ delivers a significant portfolio risk reduction out-of-sample. In particular, the risk reduction is prominent in practical situations in which the ratio $N/T$ is large, and especially when the sample covariance matrix is not even invertible. Our approach has three features that are different from previous contributions.

First, our motivation to enhance the hedging relations among stocks differentiates us from many existing papers that start from assuming a structure on the return generating process (e.g., linear factor model). To elaborate, Stevens (1998) shows that the inverse covariance matrix of stock returns reveals the optimal hedging trades among stocks. Specifically, the $i$-th row (or column) of $\Psi$ is proportional to the stock’s minimum variance hedge portfolio. The hedge portfolio consists of a long position in the $i$-th stock, and a short position in the $\Psi$.

---

4Please see Brandt (2009) for a comprehensive review of the literature.
“tracking portfolio” of the other $N - 1$ stocks to hedge the $i$-th stock. We aim to enhance the mean variance optimizer $\Psi$ by improving each hedge portfolio.

Second, we focus on estimating the inverse covariance matrix $\Psi$ by directly imposing a structure on itself, rather than on the covariance matrix $\Sigma$ first then inverting as existing approaches do. For example, Chan, Karceski, and Lakonishok (1999) use a low dimensional factor structure on the covariance matrix to improve the optimized portfolio’s out-of-sample performance.\(^5\) Many commercial risk models (e.g. APT, Axioma, MSCI Barra, Northfield, etc.) also appear to impose multivariate factor structures on the covariance matrix, though their model details are proprietary and unknown.\(^6\) Ledoit and Wolf (2003, 2004a,b) 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. In contrast to them, our approach applies an Occam’s razor to the inverse covariance matrix (mean variance optimizer).

Third, our Occam’s razor promotes sparsity of $\Psi$. That is, our estimation methodology drives some of the off-diagonal elements of $\Psi$ to zero, and these zero elements indeed constitute a significant fraction of our estimator in empirical applications. Note that the sparsity

\(^5\)Fan, Fan, and Lv (2008) also report the advantage of employing a factor structure in optimal portfolio construction.

\(^6\)There is also a voluminous work in the area of “robust portfolio optimization” with emphasis on the optimization process, especially from practitioners’ perspectives. Please see Fabozzi, Kolm, Pachamanova, and Focardi (2007) for a review. However, our approach is quite distinct from this venue of research.
of $\Psi$ implies neither the sparsity of the covariance matrix $\Sigma \equiv \Psi^{-1}$ nor the sparsity of the optimal portfolio weights.\footnote{For example, a diagonal $\Psi$ is sparse because all of its off-diagonal elements are zero. In this extreme case, for global minimum variance portfolio, each portfolio weight is the diagonal element itself thus non-zero.} It generally utilizes most stocks in the portfolio.

Our proposed estimator belongs to a wide class of shrinkage estimators, in that we bias (shrink) the estimator in a direction to reduce estimation errors. However, the gains from reducing estimation errors outweigh the costs of deviating from the optimal hedge suggested by the past evidence. For example, each hedge portfolio is supposed to hedge one stock with the other $N - 1$. But do we really need to use all the remaining $N - 1$ stocks? Does the fact that the $i$-th and $j$-th stocks hedge each other strongly in the sample period imply that this particular hedging relationship will remain in the future? With finite samples, these questions are very difficult to answer. In this situation, we do not shy away from choosing a biased answer – less hedging than the data suggests, or even no hedging – because such a solution reduces the magnitude of estimation errors. As we show, our estimator of $\Psi$ is typically sparse in empirical applications, meaning that each hedge portfolio includes only a subset of the stocks in a given portfolio.

To promote shrinkage and sparsity, we estimate $\Psi$ by maximum likelihood with an additional constraint on the sum of the absolute values (i.e., $l_1$ norm) of its off-diagonal elements.\footnote{We follow Yuan and Lin (2007) and Rothman, Bickel, Levina, and Zhu (2008) in setting up the likelihood function with the penalty, and employ the “glasso” algorithm of Friedman, Hastie, and Tibshirani (2008) to solve the estimation problem.}
Intuitively, this estimation methodology works as follows: The constraint imposes a penalty on the absolute values of the hedge trades, thus shrinking their overall trade size. Meanwhile, different stocks compete with each other to enter the hedged portfolios. The methodology turns off the \( j \)-th stock in the \( i \)-th hedged portfolio if the marginal gain from retaining the stock is not worth the cost. In effect, this restricts the \((i, j)\)th and \((j, i)\)th elements of the inverse covariance matrix to be zero. In this way, we encourage shrinkage and sparsity simultaneously in the estimation of \( \Psi \).

Aiming for risk reduction, we test examine the out-of-sample performance of the proposed sparse mean variance optimizer using a few representative datasets in the US and international stock markets. As such, the proposed sparse estimator \( \hat{\Psi} \) achieves a higher predictive likelihood of stock returns than the sample-based maximum likelihood estimator (MLE), implying the superiority of our proposed estimator in explaining the predictive covariances of stock returns. Furthermore, the proposed estimator of the mean variance optimizer accomplishes a substantial reduction in the out-of-sample risk of the global minimum variance portfolio compared to the one based on the sample covariance matrix.

Empirically, the proposed estimator also compares favorably with the portfolio with no-short-sale restriction (Jagannathan and Ma (2003)) and with the portfolio based on the shrinkage covariance matrix estimator of Ledoit and Wolf (2004a). The relative strength of our proposed estimator is more prominent in datasets with a large \( N/T \). The optimizer produces stable portfolio weights even when \( N/T \) exceeds one, that is, when the sample co-
variance matrix is not invertible. The reduction in the portfolio risk is also accompanied by a high and stable level of Sharpe ratios and certainty equivalent returns. For many datasets, the optimizer \( \hat{\Psi} \) delivers positive gains in certainty equivalent returns after accounting for transaction costs. Furthermore, with the improved optimizer \( \hat{\Psi} \) in place, no-short-sale constraint no longer helps improve the out-of-sample portfolio performance. Consistent with Green and Hollifield (1992), by improving hedge trades, \( \hat{\Psi} \) helps achieve further portfolio risk reduction beyond the no-short-sale constraint.

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

I. 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 \( \Sigma^{-1} \) reveals the optimal hedging relations among stocks. Specifically, the \( i \)-th row (or column) of \( \Sigma^{-1} \) is proportional to the \( i \)-th stock’s hedge portfolio. The \( i \)-th hedge portfolio consists of taking (1) a long position in \( i \)-th stock and (2) a short position in a portfolio of the other \( N - 1 \) stocks that tracks the \( i \)-th stock return. Each tracking portfolio can be estimated
from the following regression:

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

(1)

where \( r_{i,t} \) denotes the \( i \)-th stock return in period \( t \). \( \varepsilon_{i,t} \) is the unhedgeable component of \( r_{i,t} \), whose variance is denoted by \( \upsilon_i = \text{Var} (\varepsilon_{i,t}) \). \( \upsilon_i \) is a measure of the \( i \)-th stock’s unhedgeable risk. The objective of each hedge regression is to minimize \( \upsilon_i \), and consequently, we can view (1) as an OLS estimation problem. Stevens (1998) calls this a “regression hedge.” When \( \beta_{i|j} \) is different from zero in population, it implies that the \( j \)-th stock provides a greater hedge for the \( i \)-th stock beyond the effects of other \( N - 2 \) stocks in the portfolio.

Let us denote the \( N \times N \) covariance matrix and its inverse by \( \Sigma \) and define \( \Psi = \Sigma^{-1} = [\psi_{ij}] \), with \( \psi_{ij} \) as the \((i,j)\)th element. Then, Stevens (1998) establishes the following identity between \( \Psi \) and the hedge regression (1):

\[ \psi_{ij} = \begin{cases} 
-\frac{\beta_{i|j}}{\upsilon_i} & \text{if } i \neq j \\
\frac{1}{\upsilon_i} & \text{if } i = j
\end{cases} \]  

(2)

Remember that the hedge coefficient \( \beta_{i|j} \) represents the ability of the \( j \)-th stock to hedge the \( i \)-th stock beyond the effects of other \( N - 2 \) stocks. We can view \( \psi_{ij} \) as a measure of marginal hedgeability between the \( i \)-th and \( j \)-th stocks conditional on the presence of all

---

9For the symmetry \( \psi_{ij} = \psi_{ji} \), see Stevens’ (1998) footnote 3.
other stocks in the portfolio.\textsuperscript{10} If the \(i\)-th and \(j\)-th stocks are uncorrelated with each other after controlling for all other stocks, then \(\beta_{ij} = 0\) and hence \(\psi_{ij} = 0\) must hold.

We can gain more useful insights about the inverse covariance matrix by looking at its \(i\)-th row:

\[
\Psi(i, \cdot) = \frac{1}{\upsilon_i} \begin{bmatrix} -\beta_{i1}, ..., -\beta_{i,i-1}, 1, -\beta_{i,i+1}, ..., -\beta_{i,N} \end{bmatrix}.
\]  

(3)

The identity (2) and the expression for the hedge portfolio holdings (3) form the basis for subsequent analysis.

We can regard (3) as a vector of stock holdings in the \(i\)-th stock’s hedge portfolio. It involves a unit long position in the \(i\)-th stock and a short position of the hedge portfolio constructed from the regression (1). The hedge portfolio is a long/short portfolio whose holdings do not necessarily sum to a prefixed value (e.g. one). However, holdings of each asset are scaled by \(1/\upsilon_i\), meaning that the portfolio takes a larger (smaller) position in a stock when its unhedgeable risk is smaller (larger). We call \(\Psi\) the mean variance optimizer because it provides a detailed prescription about the hedging trades we need to make to minimize the portfolio risk.

\textsuperscript{10} Much earlier in the statistics literature, Dempster (1972) shows that \(\psi_{ij}\) represents the conditional dependence between the \(i\)-th variable and the \(j\)-th variable given all other variables in the system.
II. An Improved Mean Variance Optimizer

A. Objective

The framework laid out in section I. is useful to understand the sources of potential large estimation errors in mean variance portfolio optimization. As shown above, each hedge regression (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 us to the econometric problem: multicollinearity.\textsuperscript{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 the identity (2), this also implies that the off-diagonal elements of $\Psi$ are also susceptible to large estimation errors and instability.

We conquer the multicollinearity problem in two ways.\textsuperscript{12} First, in the hedged portfolio

\textsuperscript{11}Please see Judge, Griffiths, Hill, Lütkepohl, and Lee (1980), Chapter 12, for a detailed discussion of multicollinearity.

\textsuperscript{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. However, in practical applications, decision makers are often constrained by the available number of historical returns.
of a stock, we shrink the portfolio holdings of other stocks. The fact that the \( i \)-th and \( j \)-th stocks hedge each other in the past is not necessarily indicative of a continued mutual hedging relation going forward. In the presence of this uncertainty, we prefer erring on the non-hedging side, because it reduces the estimation errors. From the identity (2), this is equivalent to shrinking the off-diagonal elements of the inverse covariance matrix. The rationale behind the shrinkage is well known: an interior optimum exists in the trade-off between bias and estimation error if our objective is to minimize the forecast error variances. On the one hand, shrinkage introduces biases in the optimal weights of a hedge portfolio; on the other hand, it reduces the estimation errors in the estimated portfolio weights. For proper levels of shrinkage, gains from the reduction of estimation errors dominate the costs of biases.

Secondly, our estimation method curbs the multicollinearity more directly by just turning off stocks that are not worth including in the hedge portfolio. In other words, our method restricts some of the hedge portfolio weights to zero, which is equivalent to imposing zero restrictions on the off-diagonal elements of \( \Psi \). This corresponds to a strong form of shrinkage. When \((i, j)\)-th element of \( \Psi \) is zero, it simply means that the \( i \)-th and \( j \)-th stocks do not help hedge each other in the presence of other \( N - 2 \) stocks in the portfolio. For example, when \( N \) stock returns are driven by a small number of common factors, we may not need to use all stocks to hedge each other.

The zero restriction may again introduce potential misspecification biases (omitted re-
gressors in hedge regressions), but it leads to lower estimation errors. In this way, we promote the sparsity of the mean variance optimizer $Ψ$. Needless to say, it does not imply the sparsity of the covariance matrix $Σ ≡ Ψ^{-1}$ since all stock returns are still correlated with each other. As a matter of fact, $Σ$ is hardly sparse even when $Ψ$ is sparse.

B. Relation to the Recent Literature

The trade-off between the misspecification biases and estimation errors has been emphasized in a few recent papers. For example, Jagannathan and Ma (2003) demonstrate that “wrong constraints” (leading to misspecifications) can help improve the out-of-sample performance of optimized portfolios to the extent that they reduce estimation errors. Ledoit and Wolf’s (2003, 2004a,b) main impetus for shrinking the sample covariance matrix is also to strike an optimal balance between the misspecification biases and the estimation errors. Our approach shares the same objective, but suggests a different route to achieve it.

An oft-cited criticism against mean variance optimization is that optimized portfolios often involve extreme and unstable weights. Green and Hollifield (1992) argue that the extreme weights are due to the presence of a dominant systematic factor (i.e., the market factor) in the covariance structure: large long and short positions are unavoidable to hedge the systematic risk regardless of the estimation errors. In response, Jagannathan and Ma (2003) demonstrate that the measurement errors in the estimated covariance matrix still play a prominent role in deteriorating the portfolio performance, as imposing no-short-sale
constraint – that is “wrong” in population – can improve the portfolio performance by curtailing estimation errors. Our estimator of $\Psi$ combines some of the salient features in both papers. First, it focuses on reducing estimation errors by promoting the parsimony of $\Psi$. Second, it aims to capture the hedging relations better, so that it aids the portfolio to achieve a better hedging of the systematic risk.

In a recent contribution, DeMiguel, Garlappi, Nogales, and Uppal (2009) (“DGNU” for short) propose a general framework to constrain the norms of the final weighting solutions. Their general and flexible framework encompasses many existing portfolio weighting methods, including those of Jagannathan and Ma (2003) and Ledoit and Wolf (2003, 2004a,b). While we believe DGNU sets a new stage in the portfolio optimization literature, we also feel that it is worthwhile to step back to reexamine the key input ($\Psi$) for portfolio optimization. From a practical perspective, constraining the norms of the portfolio weights is interpreted as using a prior on the weighting solution (output) as DGNU note, but most portfolio managers acquire and process information about the inputs rather than the outputs.

Furthermore, hedging relations among stocks are not necessarily stable, and portfolio managers may choose to rely on their priors in choosing inputs under certain market conditions. For example, it is known that hedging relations among various style factors, such as size and value, have undergone substantial shifts periodically.\(^\text{13}\) With such shifts, historical

\(^{13}\)For example, rolling correlations among the Market, SMB, and HML factor returns have alternated signs between strongly negative and strongly positive.
covariance provides less useful guidance, but some portfolio managers are more adept at predicting the optimal hedging relations among the factors. Our approach provides more flexibility to portfolio managers as they can revise the elements of $\Psi$ to better reflect their private information when necessary. This is possible because our approach makes the relationship between hedging relations and optimal portfolio weights transparent.

Needless to say, out-of-sample performance of the optimized portfolios depends not only on the optimizer $\Psi$, but also on the expected return forecast $\mu$. Given the well-known perils of using sample means, expected return $\mu$ should further depend on an asset pricing model and the manager’s private information.\(^\text{14}\) Since there is a one-to-one mapping between $(w, \Psi)$ and $(\mu, \Psi)$, Tu and Zhou (2009a) propose to build economic objectives into the prior on portfolio weights $w$ rather than on $\mu$. This provides a sensible approach when accompanied by an informative input (or prior) for $\Psi$. Meanwhile, the success of this approach hinges on the quality of $\Psi$, as the optimal weighting solution is susceptible to estimation errors in $\Psi$. Our focus is on the significance of improving $\Psi$ independent of the choice of $\mu$, as an improved estimator of $\Psi$ should complement recent innovations in the literature. Consequently, we leave the problem of finding an optimal input of $\mu$ outside the scope of this study. In empirical applications, our primary focus is on the out-of-sample portfolio risk reduction of the global minimum variance portfolio.

\(^{14}\text{To improve the quality of } \mu, \text{ recent literature incorporates various theoretical restrictions in a prior distribution of } \mu \text{ from a Bayesian perspective. Seminal works include Black and Litterman (1992), Pástor (2000), Pástor and Stambaugh (2000), and MacKinlay and Pástor (2000).}\)
Nevertheless, the importance of improving the mean variance optimizer $\Psi$ should not be understated. While there is a general perception that estimation error in $\mu$ is far more costly than the estimation error in $\Psi$, Kan and Zhou (2007) demonstrate that when $N/T$ is not so small, there is a very significant interactive effect between the estimation errors in $\mu$ and $\Psi$ that can make the optimized portfolios very unstable and unreliable.\textsuperscript{15} This is the situation in which a stable optimizer is particularly called for. It is thus of our interest to see if our estimator of $\Psi$ achieves a significant reduction in the out-of-sample portfolio risk when $N/T$ is large.

C. Empirical Implementation

The two solutions we propose – (i) shrinkage and (ii) selection of stocks in each hedge regression – are not incompatible. We now have a statistical technology to address both objectives simultaneously in a parsimonious manner: the least absolute shrinkage and selection operator (lasso) (Tibshirani (1996)). They key innovation of lasso is to constrain the $l_1$ norm of the parameters that need to be estimated. To take advantage of this innovation, we estimate the inverse covariance matrix $\Psi$ directly by maximum likelihood, but with a constraint on

\textsuperscript{15}See Chan, Karceski, and Lakonishok (1999), Jagannathan and Ma (2003), Ledoit and Wolf (2003, 2004a), among others, for similar perspectives. Kan and Zhou (2007) and Tu and Zhou (2009b) theoretically demonstrate that, in the presence of estimation errors, mean variance investors should allocate a significant fraction of wealth to the global minimum variance portfolio and equal weighted portfolio when $N/T$ is large.
the $l_1$ norm of its off-diagonal elements.\footnote{Lasso has been receiving attention in recent econometrics literature. Caner (2009) studies a lasso-type estimator which is formed by the GMM objective function with the addition of a penalty term. He shows that the lasso-type GMM correctly selects the true model much more often than other regular procedures. Bai and Ng (2008) consider lasso to perform selection and shrinkage simultaneously in factor forecasting of time series.}

Let $R_t = (r_{1,t}, \ldots, r_{N,t})'$ be a vector of $N$ excess stock returns at time $t$. We consider a problem of estimating the inverse covariance matrix $\Psi$ from a sample of $T$ historical observations of $R_t$. Put differently, $T$ specifies the length of the “estimation window.” $\tilde{R}_t$ denotes the “centered” vector of $R_t$, whose elements have zero time-series means in the estimation window.\footnote{The MLE estimator of the mean is always the sample mean.} Then, the log likelihood for the inverse covariance matrix $\Psi = \Sigma^{-1}$ in the estimation window is

$$
\frac{T}{2} \ln |\Psi| - \frac{1}{2} \sum_{l=1}^{T} \tilde{R}_{t-l+1}' \Psi \tilde{R}_{t-l+1}.
$$

(4)

We seek an estimator $\hat{\Psi}$ that maximizes the likelihood function (4) subject to the following constraint:

$$
\sum_{i=1}^{N} \sum_{j=1}^{N} |\psi_{ij}| \leq \tau, \quad (5)
$$

where $\tau \geq 0$ is a tuning parameter so that the sum of $|\psi_{ij}|$ (across all $i$ and $j$, $i \neq j$) must be less than or equal to $\tau$. Expression (5) constrains the sum of the absolute values of its off-diagonal elements, $\psi_{ij}$ ($i \neq j$). In other words, we constrain the $l_1$ norm of the off-diagonal
elements of the inverse covariance matrix $\Psi$.¹⁸

Intuitively, the constraint (5) promotes the overall shrinkage by restricting the sum of $|\psi_{ij}|$. Meanwhile, it imposes the restriction of the form $|\psi_{ij}| = 0$ in the following way. Under the constraint (5), different elements of $\psi_{ij}$ compete with each other to remain non-zero. However, if the marginal gain from retaining the $j$-th stock in the $i$-th stock’s hedged portfolio does not justify the cost, then we turn off the $j$-th stock in the $i$-th hedged portfolio. Then, by the identity (2) and symmetry, this effectively restricts the $(i, j)$th and $(j, i)$th elements of the inverse covariance matrix to be zero.

A Lagrangian expression of this constrained maximum likelihood estimation problem is

$$
\max_{\Psi = [\psi_{ij}]} \frac{T}{2} \ln |\Psi| - \frac{T}{2} tr \left( \hat{S} \Psi \right) - \rho \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i \neq j}^{N} |\psi_{ij}|,
$$

(6)

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

$$
\hat{S} = \frac{1}{T} \sum_{t=1}^{T} \tilde{R}_{t-l+1} \tilde{R}_{t-l+1}^\prime.
$$

The regularization parameter $\rho \geq 0$ denotes the penalty on the $l_1$ norm, $\sum_{i=1}^{N} \sum_{j \neq i}^{N} |\psi_{ij}|$. The estimator, $\hat{\Psi}_\rho$, depends on the regularization parameter $\rho$, as signified by the subscript. A larger value of $\rho$ promotes more sparsity of $\hat{\Psi}_\rho$ whereas $\rho = 0$ makes the solution identical.

to the unconstrained MLE solution. We will discuss the choice of $\rho$ when we evaluate out-of-sample performance in the following section. We conduct our estimation using the graphical lasso (glasso) algorithm of Friedman, Hastie, and Tibshirani (2008).\footnote{The “penalized” maximum likelihood estimation 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 $\hat{\Psi}$ solves: $\hat{\Psi}_{\text{AIC}} = \arg \max \frac{T}{2} \ln |\Psi| - \frac{T}{2} \text{tr}(\hat{S}\hat{\Psi}) - \text{Card}(\Psi)$ where $\text{Card}(\Psi)$ is the number of non-zero elements of $\Psi$, while Schwarz’s BIC uses $\ln \frac{T}{2} \text{Card}(\Psi)$ instead of $\text{Card}(\Psi)$. The information criteria penalize the number of parameters $\text{Card}(\Psi)$ to discourage model complexity and overfitting. Unfortunately, when $N$ is not small, the penalties on $\text{Card}(\Psi)$ make our estimation problem infeasible because $\Psi$ can be sparse in $2^{n(n-1)/2}$ different ways. However, by replacing $\text{Card}(\Psi)$ or $\ln \frac{T}{2} \text{Card}(\Psi)$ with the $l_1$ norm $\rho \sum_{i=1}^{N} \sum_{j \neq i} |\psi_{ij}|$, we can combine model selection with parameter estimation within a feasible convex optimization problem. Indeed, graphical lasso (glasso) algorithm accomplishes two goals in just one step. Friedman, Hastie, and Tibshirani (2008) make the glasso algorithm publicly available in R programming language.} In Appendix, we explain how the estimation problem (6) is related to the system of hedge portfolios introduced in Section I.

III. Out-of-Sample Evaluation: Setup

A. Dataset and Methodology

To test the out-of-sample performance of the proposed mean variance optimizer, $\hat{\Psi}_\rho$, we employ the ten datasets listed in Table 1. The datasets cover the representative portfolios,
both US and international, that are of interest to both academic researchers and industry practitioners. They can be classified as: portfolios formed on size and book-to-market ratio (#1,2), industry portfolios (#3,4), country market indexes (#7,10), value and growth portfolios in international markets (#8), and their combinations (#5,6,9). The number of assets ranges between 15 and 148. Because we are interested in cases in which \( N/T \) is not small, our datasets have more assets than those studied by DeMiguel, Garlappi, Nogales, and Uppal (2009) and Tu and Zhou (2009a,b).

Table 1 about here.

We primarily focus on the out-of-sample performance of the minimum variance portfolio of risky assets (stocks). Since the minimum variance locus must be constructed from risky assets only and without risk-free ones, this portfolio also requires full initial investments. We thus exclude any “spread assets” such as the \( SMB \) and \( HML \) factor portfolios and futures contracts, otherwise they will allow investors to obtain risky exposure without making initial investments (except margins). Then spread assets won’t have bona fide returns unless we further assume that they are covered by risk-free assets, a contradiction with above. Therefore, our out-of-sample analysis excludes spread assets and imposes the usual constraint \( 1_N'w = 1 \), where \( 1_N \) denotes the \( N \times 1 \) vector of ones.\(^{21}\)

\(^{20}\)We thank Ken French for making many of the datasets available.

\(^{21}\)Of course, if our primary focus is on the Sharpe ratio rather than the portfolio risk, then it is certainly interesting to release the \( 1_N'w = 1 \) constraint and incorporate various spread assets into analysis. However,
We evaluate the out-of-sample portfolio performance using the standard “rolling-horizon” approach. In each month $t$, we construct the global minimum variance portfolios using the past 120 months (10 years) of stock returns (the “estimation window,” $T = 120$). Next, we hold such portfolios for one 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 dataset 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. Meanwhile, we are also interested in the case of large $N/T$ where a stable optimizer is called for. Therefore, our analysis is conducted on datasets that have at least 100 assets (#2,6,8,9), and also on those with $N > T$ in which the sample covariance matrix is not invertible.\footnote{To conserve space, we only report the results for $T = 120$. We have also conducted an analysis using a shorter estimation window $T = 60$, and results are available upon request. In fact, with $T = 60$ the proposed estimator achieves more significant gains in forecasting future covariances and reducing out-of-sample risk in many datasets.} Column [5] of Table 2 reports the $N/T$ ratios of our datasets.

B. Choice of the Regularization Parameter $\rho$

Our proposed $\hat{\Psi}_\rho$ depends on the regularization parameter $\rho$ [equation (6)]. Since choosing the parameter after observing out-of-sample performance induces a look-ahead bias and our current paper focuses on the out-of-sample performance of $\hat{\Psi}_\rho$ and abstracts away from the prediction of expected returns,
thus is not appropriate, we fix the first 60 months out-of-sample as a “training sample”, in which we search for the value of $\rho$ that maximizes the predictive likelihood using a grid with increment of 0.1. Then we adhere to this choice throughout the remaining out-of-sample “testing period” and burn in the first 60 out-of-sample portfolio returns. Our approach is simple and conservative, but the optimizer can deliver consistent performance when the optimal value of $\rho$ remains stable over time.

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 1978 for the first 6 datasets and in January 1990 for the next three datasets, and so on. The total number of testing periods is 374 months for the first 6 datasets, 216 months for the next three datasets, and 206 months for the last dataset (MSCI).

Table 2 about here.

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

Column [4] of Table 2 reports the values of $\rho$ chosen in the training period, which range from 0.5 to 2.4 across datasets. The next column [5] reports the “sparsity” of $\hat{\Psi}_\rho$, measured by the percentage of zero off-diagonal elements. The time-series average of the sparsity ranges from 29.0 percent to 80.1 percent, meaning that a significant fraction of the inverse covariance matrix is set to zero. In this sense, the proposed estimator $\hat{\Psi}_\rho$ is indeed highly sparse.

Table 3 reports the condition numbers for $\hat{\Psi}_\rho$ and $\tilde{S}^{-1}$ as well as for Ledoit and Wolf’s
(2004a) shrunk covariance matrix (sample covariance matrix shrunk to constant correlation matrix), the inverse of which is denoted by $\hat{\Sigma}^{-1}_{LW}$. When covariance matrices are ill-conditioned, i.e., when condition numbers are large, estimation errors amplify in the inverse operation. It can be seen that $\hat{\Psi}_\rho$ and $\hat{\Sigma}^{-1}_{LW}$ behave much better than $\hat{S}^{-1}$ as the condition numbers are much smaller for $\hat{\Psi}_\rho$ and $\hat{\Sigma}^{-1}_{LW}$ than for $\hat{S}^{-1}$. As the ratio $N/T$ gets larger, the sample covariance matrix becomes more ill-conditioned. When $N/T$ exceeds one, $\hat{S}^{-1}$ does not exist. Nevertheless, $\hat{\Psi}_\rho$ and $\hat{\Sigma}^{-1}_{LW}$ can still yield stable condition numbers.

Clearly $\hat{\Sigma}_{LW}$ and our proposed estimator $\hat{\Psi}_\rho$ share the same objective to achieve a superior prediction of covariances by reducing estimation errors. However, $\hat{\Sigma}^{-1}_{LW}$ inherits the well-conditionedness from shrinking $\hat{\Sigma}_{LW}$ toward a parsimonious structure. In contrast, our proposed estimator $\hat{\Psi}_\rho$ directly shrinks the inverse covariance matrix toward a sparse structure. Table 3 shows that both $\hat{\Psi}_\rho$ and $\hat{\Sigma}^{-1}_{LW}$ help stabilize $\hat{S}^{-1}$ in situations when the sample covariance matrix is ill-conditioned or even non-invertible.

23 In constructing the Ledoit-Wolf shrunk covariance matrix $\hat{\Sigma}_{LW}$, we replace the sample covariance matrix ($\hat{S}$) with a weighted average of the sample covariance matrix and a shrinkage target. For the latter, we use a constant correlation matrix, that is, the correlation between any two stocks is just the average of all the pairwise correlations from the sample covariance matrix. We appeal to the principle of parsimony and choose this shrinkage target instead of one obtained by further assuming a one-factor structure (such as a market model). In fact, 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. We then solve the optimal weight (shrinkage intensity) by minimizing the loss function of Ledoit and Wolf (2004b), which is represented by the Frobenius norm of the distance between the true covariance and shrinkage estimator.
IV. Out-of-Sample Evaluation: Evidence

A. Predicting Covariances Out-of-Sample

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

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

$$= \ln \left| \hat{\Sigma}_t^{-1} \right| - \tilde{R}_t' \hat{\Sigma}_t^{-1} \tilde{R}_t. \quad (7)$$

where $T_f$ is the total number of out-of-sample testing periods. $\tilde{R}_t = R_t - \frac{1}{T_f} \sum_{j=1}^{T_f} R_j$ denotes the demeaned return vector, that is calculated by subtracting the time-series mean from $R_t$ during the out-of-sample testing period.

A concern is that the Gaussian predictive likelihood function (7) is sensitive to extreme return observations such as those in October 1987 and October 2008. An estimator of $\hat{\Sigma}^{-1}$ that happens to describe these periods better may attain a higher predictive likelihood than other estimators even if its predictive likelihood is lower in many other periods. Therefore, in testing the differences in predictive likelihoods $L(\hat{\Psi}_\rho) - L(\hat{S}^{-1})$ and $L(\hat{\Psi}_\rho) - L(\hat{\Sigma}_{LW}^{-1})$, we adopt
a bootstrap resampling methodology. Because the predictive likelihood for each monthly return $l_t(\hat{\Sigma}^{-1})$ exhibits serial dependence, we apply Politis and Romano’s (1994) “stationary bootstrap,” i.e., the block resampling with block lengths having a geometric distribution with mean 25. This choice of the expected block size is consistent with the optimal block size using Politis and White’s (2004) method.\footnote{Politis and White (2004) provide a methodology of automatic selection/estimation of the optimal block length for the stationary bootstrap or circular bootstrap.} Using 1,000 resamples of $[l_t(\hat{\Psi}_\rho), l_t(\hat{S}^{-1}), l_t(\hat{\Sigma}^{-1}_{LW})]$ for $t = 1,...,T_f$, we calculate the difference in the predictive likelihoods via (7) and test the null hypothesis of no difference indirectly by constructing two-sided bootstrap intervals for the differences with nominal level $1 - \alpha$. If this interval does not contain zero, then the null is rejected at the nominal level $\alpha$ (e.g. Ledoit and Wolf (2008)).

Table 4 reports the values of $L(\hat{\Psi}_\rho) - L(\hat{S}^{-1})$ and $L(\hat{\Psi}_\rho) - L(\hat{\Sigma}^{-1}_{LW})$ for each dataset, along with the significance levels. These values are reliably different from zero at the 1% critical level, meaning that the proposed estimator $\hat{\Psi}_\rho$ delivers a significantly higher predictive likelihood than the in-sample maximum likelihood estimator $\hat{S}^{-1}$ and the shrunk estimator $\hat{\Sigma}^{-1}_{LW}$ in all datasets (except #7). This result suggests that the sparse inverse covariance matrix estimator predicts the covariances of stock returns well and often better than the sample-based (maximum likelihood) covariance matrix and the shrunk covariance matrix do out-of-sample.

Table 4 about here.
B. 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. Our focus is on the global minimum variance 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 global minimum variance portfolio is

\[
  w_{\text{GMV}} = \frac{1}{1' \hat{\Sigma}^{-1} 1} \hat{\Sigma}^{-1} 1.
\]

By replacing \( \hat{\Sigma}^{-1} \) with the proposed sparse estimator \( \hat{\Psi}_\rho \), we propose the global minimum variance portfolio:

\[
  w_{\text{GMV-} \hat{\Psi}_\rho} = (1' \hat{\Psi}_\rho 1)^{-1} \hat{\Psi}_\rho 1. \quad \text{We denote this portfolio by GMV-} \hat{\Psi}_\rho.
\]

We compare the out-of-sample portfolio risk of GMV-\( \hat{\Psi}_\rho \) with those of the following alternative portfolios.

- The sample-based global minimum variance portfolio, denoted by GMV-\( \hat{S}^{-1} \). Its portfolio weights are summarized by

  \[
  w_{\text{GMV-} \hat{S}^{-1}} = (1' \hat{S}^{-1} 1)^{-1} \hat{S}^{-1} 1.
  \]

- The equal-weighted (1/\( N \)) portfolio, \( w_{\text{EW}} = \frac{1}{N} 1 = (1' 1)^{-1} 1 \). We obtain this portfolio (denoted by EW) when we replace \( \hat{S}^{-1} \) with the identity matrix. 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 global minimum variance portfolio with no-short-sale constraint such that every single portfolio weight has to be non-negative, 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 \( i \)-th element of \( w \). \( \hat{S} \) is the sample covariance matrix.\(^{25}\)

• The global minimum variance portfolio constructed from the Ledoit and Wolf’s (2004a) shrinkage estimator, \( \hat{\Sigma}_{LW}^{-1} \). We call this portfolio GMV-LW. Its portfolio weights are summarized by \( w_{LW} = (1' \hat{\Sigma}_{LW}^{-1} N)^{-1} \hat{\Sigma}_{LW}^{-1} 1_N \).

In the out-of-sample test, we first obtain the time-series of out-of-sample returns for the five portfolios: GMV-\( \hat{\Psi}_\rho \), GMV-\( \hat{S}^{-1} \), EW (1/N), GMV-JM, and GMV-LW. Then we calculate the out-of-sample portfolio variances and standard deviations, the latter denoted by \( \sigma_{\hat{\Psi}_\rho} \), \( \sigma_{\hat{S}^{-1}} \), \( \sigma_{EW} \), \( \sigma_{JM} \), and \( \sigma_{LW} \), respectively.

We then inspect if the proposed portfolio GMV-\( \hat{S}^{-1} \) achieves a reduction in the out-of-sample portfolio risk compared to 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 \( \sigma_{\hat{S}^{-1}} - \sigma_{\hat{\Psi}_\rho} \), \( \sigma_{EW} - \sigma_{\hat{\Psi}_\rho} \), \( \sigma_{JM} - \sigma_{\hat{\Psi}_\rho} \), and \( \sigma_{LW} - \sigma_{\hat{\Psi}_\rho} \). We again apply Politis and Romano’s (1994) stationary bootstrap with expected block size of 5, following the standard

\(^{25}\)Jagannathan and Ma (2003) show that the constrained solution corresponds to a shrunk sample covariance matrix, in which covariance between a certain asset and others is reduced when the constraint on this asset is binding.
practice in the recent literature (e.g., Ledoit and Wolf (2008), DeMiguel, Garlappi, and Uppal (2009), DeMiguel, Garlappi, Nogales, and Uppal (2009)).

Columns under heading [1] of Table 5 report the out-of-sample variances of the five portfolios. The portfolio variance of GMV-$\hat{S}^{-1}$ gets very large in datasets with large $N/T$. 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 stable portfolio variance across datasets.

Furthermore, our proposed portfolio GMV-$\hat{\Psi}_\rho$ always generates lower portfolio variance than alternative portfolios in most datasets except for IntMkt (#7), where GMV-LW maintains a slightly even lower variance. Columns under heading [2] tabulate differences in out-of-sample portfolio risks (standard deviations) between GMV-$\hat{\Psi}_\rho$ and GMV-LW. Improvements in out-of-sample risk reduction from GMV-$\hat{S}^{-1}$ or EW ($1/N$) to GMV-$\hat{\Psi}_\rho$ are evident and statistically significant. GMV-$\hat{\Psi}_\rho$ also enhances significant out-of-sample risk reduction from GMV-JM, and GMV-LW for the US datasets (#1-6), but the difference is smaller in magnitude and insignificant for international datasets (#7-10). Still, GMV-$\hat{\Psi}_\rho$ is never significantly dominated in portfolio risk minimization. Therefore, the proposed portfolio compares favorably with the one based on no-short-sale restrictions (Jagannathan and Ma (2003)) and the one with Ledoit and Wolf’s (2004a) shrinkage covariance matrix estimator.\footnote{We would rather add that we do not intend to claim a general superiority of $\hat{\Psi}_\rho$-based portfolios to the ones proposed by Jagannathan and Ma (2003) and Ledoit and Wolf (2003, 2004a,b). All of these portfolios involve some flexibility in implementation. For example, Jagannathan and Ma’s (2003) portfolio strategy}
C. 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, Sharpe ratio is among the most widely used performance measures. It is also known that the global minimum variance portfolio, that ignores the estimated mean returns altogether but exploits covariances among stocks, often achieves a higher Sharpe ratio than 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. Columns under heading [1] of Table 6 tabulate monthly Sharpe ratios. GMV-\(\hat{\Psi}_\rho\) generates the highest or second-highest out-of-sample Sharpe ratios among the five portfolios, except for the datasets consisting of country indexes (#7,10). For the US portfolios (#1-6), GMV-\(\hat{\Psi}_\rho\) yields Sharpe ratios between 0.148 and 0.276 during the testing period between July 1978 and August 2009. These Sharpe ratios are higher than those of the EW (1/N) portfolio, which range can accommodate various upper and lower bounds for individual portfolio weights. Ledoit and Wolf’s (2003, 2004a,b) shrinkage estimator depends on the shrinkage target and the shrinkage intensity parameter. Our choice of the constant correlation matrix follows Ledoit and Wolf’s recommendation, but other choices are certainly possible. The proposed sparse estimator \(\hat{\Psi}_\rho\) also involves a choice of the regularization parameter \(\rho\). The relative performance of the portfolios may depend on datasets, estimation windows, and performance measures as well as testing methodologies.
between 0.130 and 0.142. The value-weighted market portfolio has a Sharpe ratio of 0.113 during the same period. Therefore, GMV-$\hat{\Psi}_\rho$ indeed achieves higher Sharpe ratios than the equally-weighted and value-weighted diversification policies.

Detecting reliable differences in the Sharpe ratios are 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’s (1981b) test is not appropriate. We therefore employ Ledoit and Wolf’s (2008) studentized circular block bootstrap (with block size 5) to test the null hypothesis of no difference in Sharpe ratios. Results are shown in Table 6 under heading [2]. GMV-$\hat{\Psi}_\rho$ achieves significantly higher Sharpe ratios than GMV-$\hat{S}^{-1}$, EW $(1/N)$, and GMV-JM in four to six datasets. However, in one of the other datasets (#1), GMV-$\hat{S}^{-1}$ attains the highest Sharpe ratio that is significantly higher than GMV-$\hat{\Psi}_\rho$ and other alternative portfolios. In addition, for industry portfolios (#3,4) and country indexes (#7,10) we do not find any statistically discernible differences in Sharpe ratios between GMV-$\hat{\Psi}_\rho$ and the four alternative portfolios.

In summary, the proposed portfolio GMV-$\hat{\Psi}_\rho$ achieves high and stable Sharpe ratios that compare favorably with alternative portfolios. However, given large estimation errors in mean returns, we refrain from drawing strong conclusions for the differences in Sharpe ratios in this exercise.

---

27 DeMiguel, Garlappi, Nogales, and Uppal (2009) also adopt this approach.
D. Economic Gains from Improved Portfolio Optimization

In principle, mean variance portfolio optimization accomplishes portfolio risk reduction beyond the EW (1/N) diversification rule by exploiting the hedging relations among stocks. Assessing the economic significance of portfolio risk reduction, however, is difficult without additional assumptions. Furthermore, meaningful assessment of economic gains must account for the effect of transaction costs, the amount of which grows rapidly with the hedge trades induced turnovers.

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

$$\text{TCost-adjusted } CER_q = \hat{\mu}_q - \gamma \hat{\sigma}_q^2 - TCost_q$$

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 $q$. Following the standard practice in the literature (e.g., Brandt (2009)), we set the risk aversion coefficient $\gamma$ to be 5. With this choice, EW (1/N) portfolio of US domestic stocks yield TCost-adjusted $CER$s higher than 3 percent, which we deem to be higher than most investors would require. However, results for the case of $\gamma = 3$ (available upon request) are qualitatively similar to the results reported in this paper. We can interpret TCost-adjusted $CER_q$ as

---

28 Tu and Zhou (2009a,b) use $\gamma = 3$. With this choice, EW (1/N) portfolio of US domestic stocks yield TCost-adjusted $CER$s higher than 3 percent, which we deem to be higher than most investors would require. However, results for the case of $\gamma = 3$ (available upon request) are qualitatively similar to the results reported in this paper.

29 According to Domowitz, Glen, and Madhavan (2001), average proportional trading costs for US equities...
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 TCost-adjusted \( CER \) of one percent. In this case, the investor is indifferent between the portfolio \( q \) and an asset that guarantees riskless return of one percent plus the risk-free rate. A higher value of TCost-adjusted \( CER \) indicates that the portfolio has a more desirable risk-return characteristic.

In order to get a sense of the size of trades, Table 7 reports monthly turnover in columns under heading [1]. 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 does not involve any hedge trades. The no-short-sale constraint also keeps the turnover of the GMV-JM low. On the contrary, GMV-\( \Psi_\rho \) and GMV-LW portfolio strategies actively employ hedge trades to reduce portfolio risk, and hence entail higher turnover. For GMV-\( \Psi_\rho \) and GMV-LW, monthly turnovers range between 0.104 and 0.733 and between 0.127 and 2.391 respectively. However, these portfolios attain a much lower turnover than GMV-\( \hat{S}^{-1} \) by reducing the effects of estimation errors.

The columns of Table 7 under heading [2] report TCost-adjusted \( CERs \) of the five portfolios. GMV-\( \Psi_\rho \) clearly realizes positive economic gains over GMV-\( \hat{S}^{-1} \) in all datasets. While GMV-\( \Psi_\rho \) incurs much larger transaction costs than the EW (1/N) and the no-short-sale con-

\[ (\text{across NYSE, AMEX, NASDAQ for 1990-1998}) \text{ is } 38.1 \text{ basis points per trade. This almost coincides with the earlier estimate of 38 basis points by Stoll and Whaley (1983) for NYSE stocks. Average transaction cost for the 18 countries in the MSCI dataset is 44.6 basis points per trade.} \]
strained GMV-JM portfolios, its gains from portfolio risk reduction still exceed the increased transaction costs for most datasets. For example, GMV-$\hat{\Psi}_\rho$ achieves positive economic gains over EW (1/N) in all datasets but #7. It also achieves positive economic gains over GMV-JM in all datasets but #4. Therefore, for an investor with risk aversion coefficient of $\gamma = 5$, economic gains from increased hedge trades to reduce portfolio risk are worth their transaction costs in most datasets. Economic gains from our proposed estimator $\hat{\Psi}_\rho$ also compare favorably with those from the Ledoit and Wolf’s shrinkage estimator $\Sigma_{LW}^{-1}$, as GMV-$\hat{\Psi}_\rho$ achieves higher TCost-adjusted CERs than GMV-LW in seven datasets. Overall, GMV-$\hat{\Psi}_\rho$ yields the highest or second highest TCost-adjusted CERs among the five portfolios in all datasets, indicating the consistency of its economic gains across different datasets.

Table 7 about here.

Still, we caution a strong conclusion from this analysis. First of all, TCost-adjusted CERs include a time-series mean of out-of-sample portfolio returns that are subject to large estimation errors. Second, this analysis tends to underestimate the gains from the minimum variance optimized portfolios in favor of EW (1/N) strategy, given the assumption that the optimized portfolios are rebalanced in every period. In practice, portfolio managers incorporate transaction costs explicitly in making their optimal portfolio decisions. Consequently, they rebalance portfolios only when predicted utility gains (e.g., certainty equivalent returns) exceed the transaction costs, thereby lowering their turnover. Seeking an optimal trade-off
between portfolio risk reduction and transaction costs, however, requires dynamic optimization with a specific penalty function on the transaction costs. We abstract from this practical implementation issue. Our focus is on improving the estimator of the inverse covariance matrix, because such an improved estimator can help improve our portfolio decisions in any setup.

E. Effects of No-Short-Sale Restriction

In some practical situations, portfolio managers impose no-short-sale constraint (non-negativity constraint) on their portfolios to avoid extreme positions. However, Green and Hollifield (1992) show that imposing no-short-sale constraint inhibits their ability to hedge the dominant systematic risk (i.e., the market factor) in minimizing portfolio risk. In response, Jagannathan and Ma (2003) demonstrate that imposing no-short-sale constraint (non-negativity constraint) does not necessarily hurt the portfolio performance because the estimated covariance matrix contains large measurement errors. In summary, the effects of no-short-sale constraint on portfolio performance depend on the magnitude of measurement errors in the estimated covariance matrix. If large estimation errors are still present in \( \hat{\Psi}_\rho \), then adding the no-short-sale restriction can still help improve the performance of the GMV-\( \hat{\Psi}_\rho \) portfolio.

Consequently, we examine how the non-negativity constraint affects the performance of the GMV-\( \hat{\Psi}_\rho \) portfolio. We use GMV-JM-\( \hat{\Psi}_\rho \) to denote the GMV-\( \hat{\Psi}_\rho \) portfolio with no-short-sale restriction, i.e., all portfolio weights of GMV-JM-\( \hat{\Psi}_\rho \) are non-negative. In other words,
GMV-JM-Ψρ replaces $S^{-1}$ with $\hat{\Psi}_\rho$ in GMV-JM.

Table 8 compares the out-of-sample portfolio performance of GMV-$\hat{\Psi}_\rho$, GMV-JM, and GMV-JM-$\hat{\Psi}_\rho$. Obviously, the non-negativity 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, though GMV-JM-$\hat{\Psi}_\rho$ yields lower turnover than GMV-JM in all datasets. This result suggests that, with the improved estimator of the inverse covariance matrix $\hat{\Psi}_\rho$ in place, the non-negativity constraint no longer helps improve the portfolio performance.$^{30}$ 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, no-short-sale restriction leads to lower portfolio performance once the estimation errors are contained.

\begin{table}[h!]
\centering
\caption{Table 8 about here.}
\end{table}

V. Conclusion

Appealing to the idea 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. This is consistent with Jagannathan and Ma’s (2003) observation that non-negativity constraint leads to 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.
covariance matrix by shrinking the estimated hedge portfolio weights. The proposed estimator of the inverse covariance matrix is sparse, meaning that a significant fraction of its off-diagonal elements are zero.

We show that the proposed estimator produces superior forecasts of covariances, and accomplishes a significantly better task in out-of-sample portfolio risk minimization compared to the sample-based (maximum likelihood) estimator of the covariance matrix, especially in datasets with large $N/T$ ratio. Furthermore, out-of-sample performance of the global minimum variance portfolio with the sparse inverse covariance matrix compares favorably to those of the equal-weighted portfolio, the no-short-sale constrained portfolio (Jagannathan and Ma (2003)), and the portfolio with shrunk covariance matrix (e.g. Ledoit and Wolf (2004a)). These results support the initial motivation of our analysis: By mitigating estimation errors in the hedge portfolios, we can enhance the ability of mean-variance optimizer in reducing the out-of-sample portfolio risk. We further show that, with the proposed estimator, economic gains from improved hedge trades exceed conventional level of transaction costs. Moreover, additional no-short-sale restriction does not help enhance the out-of-sample performance because it inhibits better use of hedge trades.
References


Appendix

In this appendix we show the relationship between the estimation of $\Psi$ and constructing the system of hedge portfolios. Again the constrained maximum likelihood estimation problem is

$$\max_{\Psi = [\psi_{ij}]} \frac{T}{2} \ln |\Psi| - \frac{T}{2} tr (\hat{S}\Psi) - \rho \sum_{i=1}^{N} \sum_{j=1}^{N} |\psi_{ij}|,$$  \hfill (8)

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

By constraining maximum likelihood, graphical lasso algorithm also imposes constraints on hedge regressions. Banerjee, El Ghaoui and d’Asprement (2008) show that the problem (8) 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 $\hat{S}$,

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

Using convex duality, Banerjee, El Ghaoui and d’Asprement show the dual problem of the above turns out to be

$$\min_{\beta} \left\{ \frac{1}{2} \left\| W_{11}^{1/2} \beta - b \right\|^2 + \rho \|\beta\|_1 \right\},$$  \hfill (9)

where $b = W_{11}^{-1/2}\hat{s}_{12}$ and $\beta = W_{11}^{-1}w_{12}$. 38
Now this dual exactly resembles a lasso least-squares problem (Tibshirani (1996)) applied on hedge regression (1) (Stevens (1998)). Without the constraint on its $l_1$ norm, $\beta$ is exactly the vector of hedge coefficients [see expression (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 $\beta$ to zero. Again the lasso estimator contains bias but reduces estimation variation.

This duality further motivates Friedman, Hastie, and Tibshirani’s (2008) glasso algorithm, which recursively solves each row and update the lasso problem until convergence, ensuring the symmetry of the $\hat{\Psi}$. For the equivalence between the above two problems, please see Banerjee, El Ghaoui and d’Asprement (2008) or Friedman, Hastie, and Tibshirani (2008).
Table 1: Data description

<table>
<thead>
<tr>
<th>#</th>
<th>1 Descriptor</th>
<th>2 Description</th>
<th>3 Mkts</th>
<th>4 N</th>
<th>5 N/T</th>
<th>6 Time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SZBM25</td>
<td>25 ((5 \times 5)) portfolios formed on size and book-to-market.</td>
<td>US</td>
<td>25</td>
<td>0.208</td>
<td>07/1963-08/2009</td>
</tr>
<tr>
<td>2</td>
<td>SZBM100</td>
<td>100 ((10 \times 10)) portfolios formed on size and book-to-market.</td>
<td>US</td>
<td>100</td>
<td>0.833</td>
<td>07/1963-08/2009</td>
</tr>
<tr>
<td>3</td>
<td>IND30</td>
<td>30 industry portfolios.</td>
<td>US</td>
<td>30</td>
<td>0.250</td>
<td>07/1963-08/2009</td>
</tr>
<tr>
<td>4</td>
<td>IND48</td>
<td>48 industry portfolios.</td>
<td>US</td>
<td>48</td>
<td>0.400</td>
<td>07/1963-08/2009</td>
</tr>
<tr>
<td>5</td>
<td>SZBM25 + IND30</td>
<td>Combination of SZBM25 and IND30.</td>
<td>US</td>
<td>55</td>
<td>0.458</td>
<td>07/1963-08/2009</td>
</tr>
<tr>
<td>6</td>
<td>SZBM100 + IND48</td>
<td>Combination of SZBM100 and IND48.</td>
<td>US</td>
<td>148</td>
<td>1.233</td>
<td>07/1963-08/2009</td>
</tr>
<tr>
<td>7</td>
<td>IntMkt</td>
<td>Country market portfolios of 15 developed countries.</td>
<td>Int’l</td>
<td>15</td>
<td>0.125</td>
<td>01/1975-12/2007</td>
</tr>
<tr>
<td>8</td>
<td>IntValGro</td>
<td>Value and growth portfolios in 15 developed countries.</td>
<td>Int’l</td>
<td>118</td>
<td>0.983</td>
<td>01/1975-12/2007</td>
</tr>
<tr>
<td>9</td>
<td>IntMkt + IntValGro</td>
<td>Combination of IntMkt and IntValGro.</td>
<td>Int’l</td>
<td>133</td>
<td>1.108</td>
<td>01/1975-12/2007-</td>
</tr>
<tr>
<td>10</td>
<td>MSCI</td>
<td>MSCI country indexes for 18 developed markets, including US.</td>
<td>Int’l</td>
<td>18</td>
<td>0.150</td>
<td>01/1977-02/2009</td>
</tr>
</tbody>
</table>

Notes: This table lists the various testing portfolios we consider. Column [1] gives the abbreviation used to refer to the testing portfolios in column [2], from either US or International capital markets (Column [3]). Column [4] reports the number of assets in each dataset, and column [5] reports the \(N/T\) ratio when \(T = 120\). The sample period of each dataset is shown in Column [6]. For IND48, we augment the data with an old industry classification [from Jul 1963 to Dec 2004; used by DeMiguel, Garlappi, and Uppal (2009)] with those from a new one from Jan 2005 to Dec 2009. IntMkts are US dollar returns of the value-weighted country market portfolios of the following 15 countries: Australia, Belgium, Canada, France, Germany, Hong Kong, Italy, Japan, Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, and UK. IntValGro contains value and growth portfolios in each of the 15 countries using four valuation ratios: book-to-market \((B/M)\); earnings-price \((E/P)\); cash earnings to price \((CE/P)\); and dividend yield \((D/P)\). 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 CE/P based value and growth portfolios for Norway, due to a large number of missing values. MSCI contains US dollar returns on the MSCI country indexes for: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Italy, Japan, Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, UK, and US.
Table 2: Out-of-Sample Periods, Regularization Parameter, and Sparsity for Estimated $\hat{\Psi}_\rho$

<table>
<thead>
<tr>
<th>#</th>
<th>Dataset</th>
<th>Out-of-Sample Analysis Period</th>
<th>Proposed Optimizer: $\Psi_\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SZBM25</td>
<td>1973:07-1978:06</td>
<td>1978:07-2009:08</td>
</tr>
<tr>
<td>2</td>
<td>SZBM100</td>
<td>1973:07-1978:06</td>
<td>1978:07-2009:08</td>
</tr>
<tr>
<td>3</td>
<td>IND30</td>
<td>1973:07-1978:06</td>
<td>1978:07-2009:08</td>
</tr>
<tr>
<td>4</td>
<td>IND48</td>
<td>1973:07-1978:06</td>
<td>1978:07-2009:08</td>
</tr>
<tr>
<td>5</td>
<td>SZBM25 + IND30</td>
<td>1973:07-1978:06</td>
<td>1978:07-2009:08</td>
</tr>
<tr>
<td>6</td>
<td>SZBM100 + IND48</td>
<td>1973:07-1978:06</td>
<td>1978:07-2009:08</td>
</tr>
<tr>
<td>7</td>
<td>IntMkt</td>
<td>1980:01-1984:12</td>
<td>1990:01-2007:12</td>
</tr>
<tr>
<td>8</td>
<td>IntValGro</td>
<td>1985:01-1989:12</td>
<td>1990:01-2007:12</td>
</tr>
<tr>
<td>9</td>
<td>IntMkt + IntValGro</td>
<td>1985:01-1989:12</td>
<td>1990:01-2007:12</td>
</tr>
<tr>
<td>10</td>
<td>MSCI</td>
<td>1987:01-1991:12</td>
<td>1992:01-2009:02</td>
</tr>
</tbody>
</table>

Notes: For each dataset, this table reports: [1] the training period; [2] testing period; and [3] the number of testing periods (months) for an out-of-sample analysis; [4] the regularization parameter $\rho$ chosen from the training period; and [5] the degree of sparsity, which is the average percentage of zero off-diagonal elements in the estimated inverse covariance matrix $\hat{\Psi}_\rho$. 
### Table 3: Condition Numbers for Estimated Inverse Covariance Matrices

<table>
<thead>
<tr>
<th>#</th>
<th>Dataset</th>
<th>N/T</th>
<th>1 $\hat{\Psi}_\rho$</th>
<th>2 $\hat{S}^{-1}$</th>
<th>3 $\hat{\Sigma}_{LW}^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean     std   max  min</td>
<td>mean    std   max  min</td>
<td>mean   std   max  min</td>
</tr>
<tr>
<td>1</td>
<td>SZBM25</td>
<td>0.21</td>
<td>414      106  642  226</td>
<td>1871    446  3285  1035</td>
<td>300    56    451  216</td>
</tr>
<tr>
<td>2</td>
<td>SZBM100</td>
<td>0.83</td>
<td>571      158  943  287</td>
<td>89897   40655 262970 29483</td>
<td>890    233   1466 566</td>
</tr>
<tr>
<td>3</td>
<td>IND30</td>
<td>0.25</td>
<td>139      44   209  67</td>
<td>601     268  1086  208</td>
<td>173    53    267  75</td>
</tr>
<tr>
<td>4</td>
<td>IND48</td>
<td>0.40</td>
<td>191      58   282  91</td>
<td>2110    943  3826  592</td>
<td>343    127   551  135</td>
</tr>
<tr>
<td>5</td>
<td>SZBM25 + IND30</td>
<td>0.46</td>
<td>567      153  879  307</td>
<td>12754   4789 30593 4333</td>
<td>610    176   1121 382</td>
</tr>
<tr>
<td>6</td>
<td>SZBM100 + IND48</td>
<td>1.23</td>
<td>675      193  1131 359</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>7</td>
<td>IntMkt</td>
<td>0.13</td>
<td>18       2    23   13</td>
<td>109     36   215  50</td>
<td>50     22    96   20</td>
</tr>
<tr>
<td>8</td>
<td>IntValGro</td>
<td>0.98</td>
<td>534      51   621  403</td>
<td>VLN     VLN  VLN  VLN</td>
<td>631    209   1038 339</td>
</tr>
<tr>
<td>9</td>
<td>IntMkt + IntValGro</td>
<td>1.11</td>
<td>354      36   416  257</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>10</td>
<td>MSCI</td>
<td>0.15</td>
<td>27       5    48   20</td>
<td>163     70   432  95</td>
<td>63     23    108  35</td>
</tr>
</tbody>
</table>

Notes: For each dataset, this table reports condition numbers of the three estimators of the inverse covariance matrix: 

- [1] $\hat{\Psi}_\rho$ is the proposed sparse estimator with the regularization parameter $\rho$ given in column [4] of Table 2;
- [2] $\hat{S}^{-1}$ is the inverse of the sample covariance matrix;
- [3] $\hat{\Sigma}_{LW}^{-1}$ is the inverse of the Ledoit-Wolf shrunk covariance matrix (sample covariance matrix shrunk to constant correlation matrix). Condition number for the (inverse) covariance matrix is the ratio of the maximal eigenvalue to the minimal eigenvalue. Higher value of the condition number indicates greater numerical instability of the inverse operation. Condition number of $\hat{S}^{-1}$ is shown as “infinite ($\infty$)” when $N > T$ ($T = 120$), as the minimal eigenvalue of $\hat{S}^{-1}$ is zero. $VLN$ means “very large number.”
Table 4: Difference in Log Predictive Likelihoods

<table>
<thead>
<tr>
<th>#</th>
<th>Dataset</th>
<th>$L(\hat{\Psi}_\rho) - L(\hat{S}^{-1})$</th>
<th>$L(\hat{\Psi}<em>\rho) - L(\hat{\Sigma}</em>{LW}^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$SZBM25$</td>
<td>2.35***</td>
<td>4.05***</td>
</tr>
<tr>
<td>2</td>
<td>$SZBM100$</td>
<td>602.63***</td>
<td>9.01***</td>
</tr>
<tr>
<td>3</td>
<td>$IND30$</td>
<td>4.09***</td>
<td>0.57***</td>
</tr>
<tr>
<td>4</td>
<td>$IND48$</td>
<td>14.69***</td>
<td>1.27***</td>
</tr>
<tr>
<td>5</td>
<td>$SZBM25 + IND30$</td>
<td>25.09***</td>
<td>3.33***</td>
</tr>
<tr>
<td>6</td>
<td>$SZBM100 + IND48$</td>
<td>$T &lt; N$</td>
<td>12.48***</td>
</tr>
<tr>
<td>7</td>
<td>$IntMkt$</td>
<td>-0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>$IntValGro$</td>
<td>468372.4***</td>
<td>19.39***</td>
</tr>
<tr>
<td>9</td>
<td>$IntMkt + IntValGro$</td>
<td>$T &lt; N$</td>
<td>26.37***</td>
</tr>
<tr>
<td>10</td>
<td>$MSCI$</td>
<td>1.60***</td>
<td>1.54***</td>
</tr>
</tbody>
</table>

Notes: This table reports differences in the log predictive likelihoods, $L(\hat{\Psi}_\rho) - L(\hat{S}^{-1})$ and $L(\hat{\Psi}_\rho) - L(\hat{\Sigma}_{LW}^{-1})$, where $L(\hat{\Psi}_\rho)$, $L(\hat{S}^{-1})$, and $L(\hat{\Sigma}_{LW}^{-1})$ are the log predictive likelihood values per observation. $T < N$ 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 expected block length of 25. ***, **, and * indicate significance at the 1%, 5% and 10% level.
<table>
<thead>
<tr>
<th>#</th>
<th>Dataset</th>
<th>( \sigma_\hat{\Psi}^2 )</th>
<th>( \sigma_{s^{-1}}^2 )</th>
<th>( \sigma_{EW}^2 )</th>
<th>( \sigma_{JM}^2 )</th>
<th>( \sigma_{LW}^2 )</th>
<th>( \sigma_{s^{-1}} - \sigma_\hat{\Psi}^2 )</th>
<th>( \sigma_{EW} - \sigma_\hat{\Psi}^2 )</th>
<th>( \sigma_{JM} - \sigma_\hat{\Psi}^2 )</th>
<th>( \sigma_{LW} - \sigma_\hat{\Psi}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SZBM25</td>
<td>13.05</td>
<td>14.28</td>
<td>25.20</td>
<td>18.61</td>
<td>20.38</td>
<td>0.16**</td>
<td>1.41***</td>
<td>0.70***</td>
<td>0.90***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SZBM100</td>
<td>12.25</td>
<td>650.93</td>
<td>25.74</td>
<td>18.03</td>
<td>17.80</td>
<td>19.85***</td>
<td>1.59***</td>
<td>0.75***</td>
<td>0.71***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IND30</td>
<td>11.97</td>
<td>13.97</td>
<td>22.78</td>
<td>13.10</td>
<td>12.63</td>
<td>0.28***</td>
<td>1.31***</td>
<td>0.16*</td>
<td>0.09*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>IND48</td>
<td>11.39</td>
<td>15.18</td>
<td>23.35</td>
<td>13.02</td>
<td>11.97</td>
<td>0.52</td>
<td>1.44***</td>
<td>0.23***</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>SZBM25 + IND30</td>
<td>9.72</td>
<td>14.90</td>
<td>23.28</td>
<td>12.61</td>
<td>10.78</td>
<td>0.74***</td>
<td>1.70***</td>
<td>0.43***</td>
<td>0.17***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SZBM100 + IND48</td>
<td>9.61</td>
<td>T &lt; N</td>
<td>24.55</td>
<td>24.39</td>
<td>10.48</td>
<td>T &lt; N</td>
<td>1.85***</td>
<td>0.43***</td>
<td>0.14**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>IntMkt</td>
<td>15.84</td>
<td>16.57</td>
<td>18.69</td>
<td>16.23</td>
<td>15.24</td>
<td>0.09*</td>
<td>0.34**</td>
<td>0.05</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>IntValGro</td>
<td>13.05</td>
<td>1193.88</td>
<td>18.85</td>
<td>14.60</td>
<td>13.63</td>
<td>30.67***</td>
<td>0.72***</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>IntMkt + IntValGro</td>
<td>12.99</td>
<td>T &lt; N</td>
<td>18.81</td>
<td>14.60</td>
<td>13.79</td>
<td>T &lt; N</td>
<td>0.72***</td>
<td>0.21</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>MSCI</td>
<td>13.89</td>
<td>14.75</td>
<td>23.11</td>
<td>16.16</td>
<td>13.98</td>
<td>0.11*</td>
<td>1.06***</td>
<td>0.29*</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: For each dataset, this table reports the monthly variances of out-of-sample returns for the five portfolios: GMV-\(\hat{\Psi}\), GMV-\(S^{-1}\), EW (1/N), GMV-JM, and GMV-LW. \(\sigma_\hat{\Psi}^2\), \(\sigma_{s^{-1}}^2\), \(\sigma_{EW}^2\), \(\sigma_{JM}^2\), and \(\sigma_{LW}^2\) 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 non-invertibility of the sample covariance matrix. Columns under [1] tabulate the point estimates of the out-of-sample monthly return variances in \(\%^2\). Numbers below the variances in italics are the ranking of portfolio variance among the five portfolios: 1 indicates the smallest variance and 5 indicates the largest. Columns under [2] tabulate the mean differences in the out-of-sample portfolio standard deviations, based on Politis and Romano’s (1994) stationary bootstrap with expected block size of 5. 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% level.
Table 6: Out-of-Sample Sharpe Ratio (Monthly)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GMV-Ψρ</td>
<td>GMV-Š⁻¹</td>
<td>EW (1/N)</td>
<td>GMV-JM</td>
<td>GMV-LW</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SZBM25</td>
<td>0.276</td>
<td>0.308</td>
<td>0.140</td>
<td>0.141</td>
<td>0.203</td>
<td>−0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SZBM100</td>
<td>0.253</td>
<td>0.000</td>
<td>0.142</td>
<td>0.160</td>
<td>0.203</td>
<td>0.253**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IND30</td>
<td>0.156</td>
<td>0.132</td>
<td>0.130</td>
<td>0.134</td>
<td>0.150</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>IND48</td>
<td>0.148</td>
<td>0.112</td>
<td>0.130</td>
<td>0.148</td>
<td>0.139</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>SZBM25 + IND30</td>
<td>0.263</td>
<td>0.300</td>
<td>0.136</td>
<td>0.160</td>
<td>0.252</td>
<td>−0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SZBM100 + IND48</td>
<td>0.271</td>
<td>T &lt; N</td>
<td>0.139</td>
<td>0.179</td>
<td>0.272</td>
<td>T &lt; N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>IntMkt</td>
<td>0.141</td>
<td>0.126</td>
<td>0.164</td>
<td>0.129</td>
<td>0.150</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>IntValGro</td>
<td>0.275</td>
<td>0.075</td>
<td>0.181</td>
<td>0.184</td>
<td>0.252</td>
<td>0.200**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>IntMkt + IntValGro</td>
<td>0.282</td>
<td>T &lt; N</td>
<td>0.179</td>
<td>0.184</td>
<td>0.251</td>
<td>T &lt; N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>MSCI</td>
<td>0.076</td>
<td>0.076</td>
<td>0.072</td>
<td>0.070</td>
<td>0.098</td>
<td>−0.000</td>
</tr>
</tbody>
</table>

Notes: This table reports the out-of-sample monthly Sharpe ratios for the following five portfolios: GMV-Ψρ, GMV-Š⁻¹, EW (1/N), GMV-JM, and GMV-LW. T < N indicates that the portfolio cannot be constructed due to the non-invertibility of the sample covariance matrix. Columns under [1] report the level of monthly Sharpe ratios. Numbers below the Sharpe ratios in italics indicate the ranking of Sharpe ratios among the five portfolios: 1 indicates the highest Sharpe ratio and 5 indicates the lowest. Columns under [2] report differences in Sharpe ratios between GMV-Ψρ and other four 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 a block size equal to 5. ***, **, and * indicate significance at the 1%, 5% and 10% level.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dataset</td>
<td>GMV-Ψρ</td>
<td>GMV-S⁻¹</td>
</tr>
<tr>
<td>1</td>
<td>SZBM25</td>
<td>0.374</td>
<td>0.801</td>
</tr>
<tr>
<td>2</td>
<td>SZBM100</td>
<td>0.544</td>
<td>7.955</td>
</tr>
<tr>
<td>3</td>
<td>IND30</td>
<td>0.192</td>
<td>0.475</td>
</tr>
<tr>
<td>4</td>
<td>IND48</td>
<td>0.252</td>
<td>0.812</td>
</tr>
<tr>
<td>5</td>
<td>SZBM25 + IND30</td>
<td>0.438</td>
<td>1.700</td>
</tr>
<tr>
<td>6</td>
<td>SZBM100 + IND48</td>
<td>0.522</td>
<td>T &lt; N</td>
</tr>
<tr>
<td>7</td>
<td>IntMkts</td>
<td>0.104</td>
<td>0.177</td>
</tr>
<tr>
<td>8</td>
<td>IntValGro</td>
<td>0.586</td>
<td>147.8</td>
</tr>
<tr>
<td>9</td>
<td>IntValGroMkt</td>
<td>0.733</td>
<td>T &lt; N</td>
</tr>
<tr>
<td>10</td>
<td>MSCI</td>
<td>0.141</td>
<td>0.228</td>
</tr>
</tbody>
</table>

Notes: Columns under [1] report turnover of the five portfolios: GMV-Ψρ, GMV-S⁻¹, EW (1/N), GMV-JM, and GMV-LW. Please see the text for their definition. Calculation of the portfolio turnover follows DeMiguel, Garlappi, and Uppal (2009). T < N indicates that the portfolio cannot be constructed due to the non-invertibility of the sample covariance matrix. Columns under [2] report the values of the transaction cost adjusted certainty equivalent excess returns (TCost-adjusted CERs), shown in annual percentage points. The transaction cost of each is calculated as 50 basis points times monthly turnover times 12 (to annualize).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SZBM25</td>
<td>13.05</td>
<td>0.276</td>
<td>0.374</td>
</tr>
<tr>
<td>2</td>
<td>SZBM100</td>
<td>12.25</td>
<td>0.253</td>
<td>0.544</td>
</tr>
<tr>
<td>3</td>
<td>IND30</td>
<td>11.97</td>
<td>0.156</td>
<td>0.192</td>
</tr>
<tr>
<td>4</td>
<td>IND48</td>
<td>11.39</td>
<td>0.148</td>
<td>0.252</td>
</tr>
<tr>
<td>5</td>
<td>SZBM25 + IND30</td>
<td>9.72</td>
<td>0.263</td>
<td>0.438</td>
</tr>
<tr>
<td>6</td>
<td>SZBM100 + IND48</td>
<td>9.61</td>
<td>0.271</td>
<td>0.522</td>
</tr>
<tr>
<td>7</td>
<td>IntMkts</td>
<td>15.84</td>
<td>0.141</td>
<td>0.104</td>
</tr>
<tr>
<td>8</td>
<td>IntValGro</td>
<td>13.05</td>
<td>0.275</td>
<td>0.586</td>
</tr>
<tr>
<td>9</td>
<td>IntValGroMkt</td>
<td>12.99</td>
<td>0.282</td>
<td>0.733</td>
</tr>
<tr>
<td>10</td>
<td>MSCI</td>
<td>13.89</td>
<td>0.076</td>
<td>0.141</td>
</tr>
</tbody>
</table>

Notes: This table compares GMV-$\hat{\Psi}_\rho$, GMV-JM, and GMV-JM-$\hat{\Psi}_\rho$ in terms of [1] out-of-sample monthly return variance; [2] monthly Sharpe Ratio, and [3] monthly turnover. 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 no-short-sale restriction (e.g., Jagannathan and Ma (2003)).
Founded in 1892, the University of Rhode Island is one of eight land, urban, and sea grant universities in the United States. The 1,200-acre rural campus is less than ten miles from Narragansett Bay and highlights its traditions of natural resource, marine and urban related research. There are over 14,000 undergraduate and graduate students enrolled in seven degree-granting colleges representing 48 states and the District of Columbia. More than 500 international students represent 59 different countries. Eighteen percent of the freshman class graduated in the top ten percent of their high school classes. The teaching and research faculty numbers over 600 and the University offers 101 undergraduate programs and 86 advanced degree programs. URI students have received Rhodes, Fulbright, Truman, Goldwater, and Udall scholarships. There are over 80,000 active alumnae.

The University of Rhode Island started to offer undergraduate business administration courses in 1923. In 1962, the MBA program was introduced and the PhD program began in the mid 1980s. The College of Business Administration is accredited by The AACSB International - The Association to Advance Collegiate Schools of Business in 1969. The College of Business enrolls over 1400 undergraduate students and more than 300 graduate students.

**Mission**

Our responsibility is to provide strong academic programs that instill excellence, confidence and strong leadership skills in our graduates. Our aim is to (1) promote critical and independent thinking, (2) foster personal responsibility and (3) develop students whose performance and commitment mark them as leaders contributing to the business community and society. The College will serve as a center for business scholarship, creative research and outreach activities to the citizens and institutions of the State of Rhode Island as well as the regional, national and international communities.

The creation of this working paper series has been funded by an endowment established by William A. Orme, URI College of Business Administration, Class of 1949 and former head of the General Electric Foundation. This working paper series is intended to permit faculty members to obtain feedback on research activities before the research is submitted to academic and professional journals and professional associations for presentations.

An award is presented annually for the most outstanding paper submitted.