JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 55, No. 8, Dec. 2020, pp. 2700–2731
COPYRIGHT 2019, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S0022109019000899

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

Improving Minimum-Variance Portfolios by Alleviating Overdispersion of Eigenvalues

Fangquan Shi, Lianjie Shu, Aijun Yang, and Fangyi He*

Abstract

In portfolio risk minimization, the inverse covariance matrix of returns is often unknown and has to be estimated in practice. Yet the eigenvalues of the sample covariance matrix are often overdispersed, leading to severe estimation errors in the inverse covariance matrix. To deal with this problem, we propose a general framework by shrinking the sample eigenvalues based on the Schatten norm. The proposed framework has the advantage of being computationally efficient as well as structure-free. The comparative studies show that our approach behaves reasonably well in terms of reducing out-of-sample portfolio risk and turnover.

I. Introduction

In mean-variance portfolio optimization, originally proposed by Markowitz (1952), the optimal portfolio weights are determined to minimize the portfolio variance for a particular return objective. Although the mean-variance framework is important in theoretical research, it is difficult to use it in practical portfolio management. This is because the optimal portfolio weights depend on the mean and inverse covariance matrix of asset returns. In practice, the mean and covariance matrix of asset returns are often unknown and need to be estimated. The sample mean and covariance matrix are usually used to replace the true mean and covariance matrix to implement the mean-variance optimal portfolio. However, the estimation uncertainties in parameters could lead to suboptimal portfolios for which the asset weights are noisy and unstable over time (e.g., Jobson and Korkie (1980), (1981), Best and Grauer (1991), and Kan and Zhou (2007)). The unstable weights lead to significant rebalancing and transaction costs.

*Shi, shi.fangquan@connect.um.edu.mo, University of Macau Faculty of Business Administration; Shu (corresponding author), ljshu@um.edu.mo, University of Macau Faculty of Business Administration; Yang, ajyang81@163.com, Nanjing Forest University College of Economics and Management; and He, fangyi_he@swufe.edu.cn, Southwestern University of Finance and Economics School of Finance. We thank the editor and an anonymous referee for their valuable comments. Shu's work was supported in part by the Macau Science and Technology Development Fund (FDCT/0064/2018/A2) and the University Research Grant (MYRG2018-00087-FBA). He's work was supported by the National Natural Science Foundation of China (No. 71772153).

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

The estimation uncertainties come from two sources: estimation of the mean of returns and estimation of the covariance matrix of returns. It is well known that estimation errors in the mean tend to have more influence than in the covariance matrix (Best and Grauer (1991)). Using the sample mean in portfolio optimization could lead to extremely unstable portfolios. For this reason, recent research has focused on minimum-variance portfolios (MVPs) by ignoring the mean estimate and relying on the covariance estimate only. Some empirical evidence from Jagannathan and Ma (2003) indeed shows that the MVP usually has better out-of-sample performance than any mean-variance portfolio.

Nevertheless, the MVP still suffers from the source of uncertainty in estimating the covariance matrix of returns. A sizable literature documents that the textbook estimator, the sample covariance matrix, is inappropriate (see Scherer and Martin (2005) for an overview). In order to reduce sampling errors in estimating the covariance matrix, numerous approaches have been proposed. These approaches could be grouped into three categories: i) to shrink the covariance matrix, ii) to restrict portfolio weights, and iii) to shrink the inverse covariance matrix directly. A typical way to shrink the covariance matrix is based on Bayes–Stein shrinkage estimators (Jorion (1986)), which combine the sample covariance matrix with a prior matrix. The basic idea behind this approach is to introduce a balance between variance and bias by shrinking the sample covariance matrix toward a prior matrix. See, for example, Ledoit and Wolf (2003), (2004), Candelon, Hurlin, and Tokpavi (2012), and Gillen (2014).

Methods in the second category aim at reducing sampling errors in the covariance matrix by constraining portfolio weights. This idea is motivated by the fact that the optimal portfolio weights based on the sample covariance matrix tend to take large positive and negative positions in the underlying assets (Jagannathan and Ma (2003)). One of the main drawbacks of these very leveraged portfolio weights is that they are unstable and thus generate considerable turnover. To cope with this problem, Jagannathan and Ma (2003) add no-short-sale constraints (nonnegative weights) to the MVP model. DeMiguel, Garlappi, Nogales, and Uppal (2009) further provide a general framework of constraining portfolio weights based on norm constraints. This general framework nests the shrinking approaches of Jagannathan and Ma (2003) and Ledoit and Wolf (2003), (2004) as special cases. Some recent variations in this category include Behr, Guettler, and Miebs (2013), Levy and Levy (2014), and Xing, Hu, and Yang (2014).

In addition to the shrinkage of the sample covariance matrix, the inverse covariance matrix can also be shrunk. Some up-to-date approaches focus on this category. One reason is perhaps due to the fact that the inverse covariance matrix has a more immediate impact on portfolio weights than the covariance matrix. After obtaining the inverse covariance matrix estimate, one can directly determine portfolio weights without the need for computing the inverse. Estimation of a high-dimensional covariance matrix or its inverse has become a large and active research field in recent times. A comprehensive review of the entire literature is clearly beyond the scope of this article. Nevertheless, the graphical lasso (Glasso) estimator suggested by Goto and Xu (2015) and the condition-number-regularized estimator proposed by Won, Lim, Kim, and Rajaratnam (2013) represent two important avenues recently employed in portfolio optimization for

estimating a high-dimensional inverse covariance matrix: structure-based estimation and structure-free estimation. Structure-based estimation makes the problem more amenable by assuming additional structure in the inverse covariance matrix. One of the most common structures in this framework is sparsity. Structure-free estimation requires no prespecification of a structured form of the covariance matrix and its inverse. The Glasso estimator is a structure-based approach. It uses the ℓ_1 penalty to impose sparsity, aimed at shrinking the off-diagonal elements of the inverse covariance matrix toward 0. In contrast, the condition-number-regularized approach is a structure-free approach that shrinks the sample eigenvalues with a condition-number constraint. The inverse covariance matrix is estimated based on the original sample eigenvectors and the shrunk eigenvalues without resort to any hypothesized matrix structure.

When there are no a priori beliefs about the matrix structure, structure-free estimation is the most general and neutral approach. In practice, numerous portfolios are composed of many assets whose returns are more or less related to economic fundamentals in a complex manner. The covariance structure of asset returns or its inverse is hard to specify explicitly. This is why a structure-free estimate is favorable in portfolio optimization.

The goal of this article is to propose a structure-free but computationally efficient estimator of the inverse covariance matrix for improving portfolio risk optimization, motivated by alleviating overdispersion of the sample eigenvalues. Notice that the inverse covariance matrix can be uniquely determined by its eigenvalues and eigenvectors according to spectral decomposition. It is well known that when the dimension is not small relative to the sample size, the eigenvalues of the sample covariance matrix will be more dispersed than the true ones (Marčenko and Pastur (1967), Stein (1975)). Moreover, as shown in Section II, sampling errors in the inverse covariance matrix are mainly dominated by errors in the eigenvalues. The effect of sampling errors in eigenvectors is only marginal compared with that of errors in eigenvalues. This reveals that one can greatly reduce estimator errors in the inverse covariance matrix by alleviating overdispersion of the sample eigenvalues. This inspires us to regularize the sample eigenvalues when estimating the inverse covariance matrix.

In particular, under our approach, the inverse covariance matrix is estimated based on maximum likelihood with a constraint on the Schatten q-norm of the inverse covariance matrix. The Schatten q-norm is very general and includes many classical norms ($q = 1, 2, +\infty$) as special cases. An important advantage of this framework is that the matrix-optimization problem can be transformed into an equivalent unconstrained univariate minimization problem, which enables us to derive a closed-form solution to the optimal shrinkage density of the sample eigenvalues for a given regularization parameter. This is a significant improvement in the solution method compared with the majority of methods for estimating a high-dimensional matrix based on numerical solutions. This property substantially reduces the computation time for portfolio optimization and thus greatly facilitates the implementation of the proposed method.

Unlike the Glasso approach, both the condition-number-regularized approach and our approach do not impose sparsity conditions, and both are structure-free estimators. Notice that the condition-number-regularized approach bounds

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

the condition number with the goal of obtaining a well-conditioned estimator. It is different from our approach in a fundamental sense in that our approach bounds the Schatten norm with the direct goal of reducing overdispersion associated with the sample eigenvalues. The condition-number-regularized approach winsorizes the sample eigenvalues at the lower end and the upper end of the spectrum but keeps the middle sample eigenvalues. Therefore, it cannot adjust for the distortion of the middle eigenvalues. In contrast, the proposed approach can suppress fluctuations of all the sample eigenvalues in the regularization.

II. Sampling Errors in the Covariance Matrix of Asset Returns

Let r_t denote the vector of asset returns on p stocks with covariance matrix Σ at time t. In the traditional MVP, the optimal portfolio weights are obtained by minimizing the overall risk of the portfolio under the full investment constraint:

$$\begin{aligned} w^* &= \arg\min_w w'\Sigma w \\ \text{s.t. } w'e &= 1, \end{aligned} \tag{1}$$

where w is the column vector of length p of the portfolio weights, and e is the vector of length p with all entries equal to 1. The explicit solution to the MVP is given by

$$w^* = \frac{\Sigma^{-1}e}{e'\Sigma^{-1}e}. \tag{2}$$

Theoretically, equation (2) illustrates that the optimal weights of the MVP depend on the inverse covariance matrix Σ^{-1} or the covariance matrix, which is unobservable. In practice, different methods could be used to estimate the covariance matrix or the inverse covariance matrix. Under the assumption that r_t is independent and identically distributed multivariate normal with mean μ and covariance matrix Σ, a natural estimate (maximum likelihood) of Σ is the sample covariance matrix

$$S = \frac{1}{n}\sum_{t=1}^{n}(r_t - \hat{\mu})(r_t - \hat{\mu})', \tag{3}$$

where n is the sample size, and $\hat{\mu}$ is the sample mean of asset returns. However, substituting the true covariance matrix with its estimate could lead to the MVP being highly unstable because the sample covariance matrix is a noisy estimator, especially when n is small relative to p.

In order to better understand the sources of sampling errors in estimating Σ and Σ^{-1}, we consider the spectral decomposition of the covariance matrix $\Sigma = U\Lambda U'$, where Λ is a $p \times p$ diagonal matrix whose diagonal entries are eigenvalues of Σ. That is, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$, where $0 < \lambda_1 \leq \cdots \leq \lambda_p$. The ith column of U, u_i, is the eigenvector corresponding to the ith eigenvalue. The spectral decomposition of Σ implies that

$$\Sigma = \lambda_1 u_1 u_1' + \lambda_2 u_2 u_2' + \cdots + \lambda_p u_p u_p'. \tag{4}$$

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

From the spectral decomposition, one can split variations in estimating Σ into two sources: variation from estimating eigenvalues and variation from estimating eigenvectors. Similarly, the spectral decomposition of the inverse covariance matrix implies that

$$\Sigma^{-1} = UMU' = \frac{1}{\lambda_p}u_p u_p' + \frac{1}{\lambda_{p-1}}u_{p-1}u_{p-1}' + \cdots + \frac{1}{\lambda_1}u_1 u_1', \tag{5}$$

where $M = \text{diag}(m_1, \ldots, m_p)$, and $m_i = 1/\lambda_i$.

It is well known that when the dimension p is not small relative to the sample size n (i.e., $p/n \to c > 0$), the sample eigenvalues are overdispersed (i.e., the larger eigenvalues of the sample covariance matrix are overestimated, whereas the smaller eigenvalues are underestimated) (Marčenko and Pastur (1967), Stein (1975)). Therefore, unless $p/n \to 0$, sampling errors in both eigenvalues and eigenvectors will affect the estimation accuracy of the inverse covariance matrix.

The following simulation demonstrates the effects of the sample eigenvalues and eigenvectors on estimating Σ and Σ^{-1}. Following Tu and Zhou (2011), the stock excess returns are generated using the Fama–French 3-factor model

$$r_t = \beta f_t + \varepsilon_t, \tag{6}$$

where f_t is the 3×1 vector of excess returns on the 3 factors, β is the $p \times 3$ matrix of factor loadings, and ε_t is the $p \times 1$ vector of residuals. The factors are generated from a multivariate normal distribution whose mean and covariance matrix are computed using the monthly data about the Fama–French 3 factors during the period from July 1963 to Aug. 2007. The factor loadings are randomly drawn from the uniform distribution over the interval $[0.9, 1.2]$ for the market factor, an interval $[-0.3, 1.4]$ for the size factor, and an interval $[-0.5, 0.9]$ for the book-to-market factor. The residuals are randomly drawn from the multivariate normal distribution with the diagonal covariance matrix whose diagonal elements are drawn from the uniform distribution over the interval $[0.1, 0.3]$.

To quantify sampling errors in estimating Σ, the Frobenius norm of the difference between the true covariance matrix and its estimate, $\|\hat{\Sigma} - \Sigma\|_F$, is calculated, where the Frobenius norm of a matrix A is defined as $\|A\|_F = \{\text{trace}(AA')\}^{1/2}$. In order to compare the respective contribution of eigenvectors and eigenvalues to sampling errors in estimating Σ, we consider three different ways to estimate the covariance matrix based on the spectral decomposition in equation (4): i) by using both sample eigenvalues and eigenvectors (denoted as SS), ii) by using true eigenvalues but sample eigenvectors (denoted as TS), and iii) by using sample eigenvalues but true eigenvectors (denoted as ST). Analogously, sampling errors in estimating the inverse covariance matrix are measured.

Table 1 displays the mean estimation error in estimating both Σ and Σ^{-1} across 1,000 simulations under different sample sizes ($n = 60, 120, 180$). A pronounced observation from Table 1 is that when estimating Σ^{-1}, the TS method can substantially reduce the mean sampling error compared with the SS method, whereas the reduction in the sampling error by using the ST method is only marginal relative to the SS method. This clearly indicates that the contribution of the sample eigenvalues to sampling errors in estimating Σ^{-1} is much more substantial than that of the sample eigenvectors. In order to reduce sampling errors

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

TABLE 1
Sampling Errors in the Eigenstructure

Table 1 reports the average Frobenius norm of the difference between the true and estimated (inverse) covariance matrices of asset returns based on 1,000 simulated replicates. The returns on 50 assets with different sample sizes (n) are generated from the Fama–French 3-factor model. "SS" represents the estimator that uses sample eigenvalues and sample eigenvectors, "TS" represents the estimator that uses true eigenvalues but sample eigenvectors, and "ST" represents the estimator that uses sample eigenvalues but true eigenvectors.

	Covariance Matrix			Inverse Covariance Matrix		
	SS	TS	ST	SS	TS	ST
$n=60$	0.1323	0.0763	0.0883	9,001	172	8,925
$n=120$	0.0927	0.0595	0.0549	758	157	670
$n=180$	0.0758	0.0518	0.0414	434	147	347

in the inverse covariance matrix, it is important to obtain accurate and reliable estimates of eigenvalues, especially the first few dominating eigenvalues for the inverse covariance matrix. Moreover, it can be seen from Table 1 that the errors in the covariance matrix are much smaller than those in the inverse covariance matrix. This is because the eigenvalues of Σ^{-1} are much more dispersed than those of the covariance matrix of asset returns.

III. An Estimator of the Inverse Covariance Matrix Constrained by Schatten Norms

A. Regularization via Schatten Norms

The preceding section is useful to understand the main source of potentially large errors in estimating Σ^{-1} arising from eigenvalues. In this section, we propose an improved estimator of the inverse covariance matrix by shrinking its eigenvalues. For simplicity, denote the inverse covariance matrix as $\Omega=\Sigma^{-1}$. The Schatten q-norm is the q-norm of the vector of singular values of a matrix. The Schatten q-norm of Ω is defined as

$$\|\Omega\|_q = \left(\sum_{i=1}^{p} m_i^q\right)^{1/q}, \tag{7}$$

where m_i is the ith largest eigenvalue of matrix Ω (i.e., $m_1 \geq m_2 \cdots \geq m_p$), and $m_i = 1/\lambda_i$. In order to provide a reliable estimate of Ω, we minimize the negative log-likelihood with the Schatten norm constraint as follows:

$$\begin{aligned} &\min_{\Omega} \quad \operatorname{tr}(\Omega S) - \log|\Omega| \\ &\text{s.t.} \quad \left(\sum_{i=1}^{p} m_i^q\right)^{1/q} \leq \delta. \end{aligned} \tag{8}$$

The Schatten q-norm represents a family of norms. When $q=1$, the Schatten norm restricts the sum of the eigenvalues of the inverse covariance matrix to be less than a threshold, which is equivalent to the trace norm. When $q=2$, it imposes a constraint on the sum of the squared eigenvalues of the inverse covariance matrix, which is equivalent to the Frobenius norm of the inverse covariance matrix. When $q=+\infty$, it imposes a constraint on the largest eigenvalue of the inverse covariance matrix, equivalent to the spectral norm.

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

Because the constraint $(\sum_{i=1}^{p} m_i^q)^{1/q} \leq \delta$ can be written in the form $\sum_{i=1}^{p} m_i^q \leq \delta^q$, the previous optimization problem can be transformed into the equivalent Lagrangian form as

$$\min_{\Omega} \text{tr}(\Omega S) - \log|\Omega| + \rho \sum_{i=1}^{p} m_i^q, \tag{9}$$

where $\rho \geq 0$ is a regularization parameter. Notice that $\sum_{i=1}^{p} m_i^q = \text{tr}(\Omega^q)$, and the objective function in equation (9) can be rewritten as

$$\text{tr}(\Omega S) - \log|\Omega| + \rho \text{tr}(\Omega^q). \tag{10}$$

Note that the negative log-likelihood function is a convex function of Ω. Moreover, $\text{tr}(\Omega^q)$ is also a convex function. Therefore, the objective function in equation (10) is strictly convex. With the convex property, there exists a unique 0 subgradient at the global minimizer $\Omega \succ 0$ (i.e., Ω is positive definite) (Witten and Tibshirani (2009)). Differentiating equation (10) with respect to Ω leads to

$$S - \Omega^{-1} + \rho q \Omega^{q-1} \quad = \quad \mathbf{0}. \tag{11}$$

This equation implies that Ω and S share the same eigenvectors. Denote the spectral decomposition of the sample covariance matrix as $S = QLQ'$, where L is a $p \times p$ diagonal matrix whose diagonal entries are eigenvalues of S. That is, $L = \text{diag}(l_1, \ldots, l_p)$, where $0 < l_1 \leq \cdots \leq l_p$. The ith column of Q, q_i, is the eigenvector corresponding to the ith eigenvalue of the sample covariance matrix. According to equation (11), it follows that

$$l_i - \frac{1}{m_i} + \rho q m_i^{q-1} \quad = \quad 0. \tag{12}$$

When $q = 1$, it is easy to obtain the regularized eigenvalues of Ω from equation (12) as

$$\hat{m}_i \quad = \quad \frac{1}{l_i + \rho}, \quad i \quad = \quad 1, \ldots, p.$$

When $q = 2$, one can obtain

$$\hat{m}_i \quad = \quad \frac{2}{\sqrt{l_i^2 + 8\rho} + l_i}.$$

When $q = \infty$, the Schatten q-norm reduces to $m_{\max}(\Omega) \leq \delta$. The eigenvalue-regularized estimate of Ω can be obtained by solving the following optimizing problem:

$$\begin{array}{ll} \min_{\Omega} & \text{tr}(\Omega S) - \log|\Omega| \\ \text{s.t.} & m_{\max}(\Omega) \quad \leq \quad \delta. \end{array} \tag{13}$$

As shown in Appendix A, the optimal estimate of m_i is given by

$$\hat{m}_i \quad = \quad \min\left\{\frac{1}{l_i}, \delta\right\}. \tag{14}$$

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

In summary, the relationship between the regularized eigenvalues of the inverse covariance matrix $\hat{m}_i$ and the eigenvalues of the sample covariance matrix, l_i, is as follows:

$$\hat{m}_i = \begin{cases} 1/(l_i+\rho), & q = 1, \\ 2/\left(\sqrt{l_i^2+8\rho}+l_i\right), & q = 2, \\ \min\left\{\frac{1}{l_i}, \delta\right\}, & q = \infty. \end{cases} \tag{15}$$

Based on the regularized eigenvalues of the inverse covariance matrix, one can estimate Ω based on

$$\hat{\Omega}_{\text{SN}} = Q\hat{M}Q', \tag{16}$$

where $\hat{M} = \text{diag}(\hat{m}_1, \ldots, \hat{m}_p)$ is a $p \times p$ diagonal matrix, and $\hat{m}_1 \geq \cdots \geq \hat{m}_p$. Notice that when estimating Ω, the proposed estimator shrinks the sample eigenvalues but keeps the sample eigenvectors. There are errors in the sample eigenvectors because the sample eigenvectors are not consistent when the dimension p is not small relative to the sample size n (Johnstone and Lu (2009)). The fact that we keep the sample eigenvectors does not mean that we assume they are close to the true eigenvectors. Instead, it only means that we don't know how to improve upon them, as pointed out by Stein (1975). The proposed estimator belongs to the class of rotation-equivariant estimators introduced by Stein (1975): Rotating the original variables results in the same rotation being applied to the covariance matrix estimator. Rotation equivariance is appropriate in the general case when there is no a priori information about the orientation of the true eigenvectors of the covariance matrix.

B. Relation to Existing Approaches

As mentioned earlier, the proposed approach shares certain similarities with the most up-to-date approaches, such as the Glasso approach of Goto and Xu (2015) and the condition-number-regularized approach of Won et al. (2013), in the sense that all of them belong to the category of shrinking the inverse covariance matrix. All of them are likelihood-based approaches to estimating the inverse covariance matrix. However, they are developed for different purposes. The Glasso estimator is sparse, aimed at shrinking the off-diagonal elements of the inverse covariance matrix toward 0. Unlike the Glasso approach, both the condition-number-regularized approach and our approach are structure-free, and both shrink the sample eigenvalues. The condition-number-regularized approach imposes a constraint on the condition number for improving the numerical stability of a matrix estimator, whereas our approach imposes a constraint on the Schatten norm for reducing overdispersion of the sample eigenvalues.

The proposed approach is also closely related to some traditional portfolio approaches. The following propositions show that the Schatten 1-norm-constrained portfolio weights, w_{SN_1}, coincide with those using the covariance matrix estimator proposed by Ledoit and Wolf (2004) and the 2-norm-constrained portfolio weights proposed by DeMiguel et al. (2009). Moreover, it can be shown in Proposition 3 that the improved MVP weights converge to $1/N$ when the

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

regularization parameter $\rho \to \infty$ or when δ is smaller than or equal to the reciprocal of the maximum eigenvalue of the sample covariance matrix (i.e., $\delta \leq 1/l_{\max}$).

Proposition 1. The MVP using the covariance matrix estimator $\hat{\Sigma}_{\text{SN}_1} = S + \rho I$ coincides with that using the estimator of Ledoit and Wolf (2004), $\hat{\Sigma}_{\text{LW}} = (1-\alpha)S + \alpha\bar{\sigma}^2 I$, where $\bar{\sigma}^2$ is the average of the diagonal elements of the sample covariance matrix.

The Schatten 1-norm-constrained portfolio weights, w_{SN_1}, are the solution to the traditional portfolio problem in equation (2) with the inverse covariance matrix estimated by the maximum likelihood subject to the additional constraint that the sum of eigenvalues (or trace) of Ω is smaller than or equal to a certain threshold δ; that is, $\left(\sum_{i=1}^{p} m_i\right) \leq \delta$. It is straightforward that based on the Schatten 1-norm constraint, the covariance matrix estimator can be written as

$$\hat{\Sigma}_{\text{SN}_1} = Q\text{diag}(l_i + \rho, \ldots, l_p + \rho)Q' = S + \rho I. \tag{17}$$

The MVP weights solved using $\hat{\Sigma}_{\text{SN}_1} = S + \rho I$ can be obtained as

$$w_{\text{SN}_1} = \frac{\hat{\Sigma}^{-1}e}{e'\hat{\Sigma}^{-1}e} = \frac{(S+\rho I)^{-1}e}{e'(S+\rho I)^{-1}e}. \tag{18}$$

The MVP weights solved using the covariance matrix estimator of Ledoit and Wolf (2004), $\hat{\Sigma}_{\text{LW}} = (1-\alpha)S + \alpha\bar{\sigma}^2 I$, can be obtained as

$$w_{\text{LW}} = \frac{\hat{\Sigma}_{\text{LW}}^{-1}e}{e'\hat{\Sigma}_{\text{LW}}^{-1}e} = \frac{((1-\alpha)S + \alpha\bar{\sigma}^2 I)^{-1}e}{e'((1-\alpha)S + \alpha\bar{\sigma}^2 I)^{-1}e} = \frac{(S + \frac{\alpha\bar{\sigma}^2}{1-\alpha} I)^{-1}e}{e'(S + \frac{\alpha\bar{\sigma}^2}{1-\alpha} I)^{-1}e}. \tag{19}$$

Clearly, w_{SN_1} is the same as w_{LW} when $\rho = \alpha\bar{\sigma}^2/(1-\alpha)$.

Proposition 2. Provided that the sample covariance matrix S is nonsingular, for each $\rho \geq 0$, there exists a δ such that the MVP constructed using the covariance matrix estimator $\hat{\Sigma}_{\text{SN}_1} = S + \rho I$ coincides with the MVP with the additional constraint

$$w'w \leq \delta. \tag{20}$$

The 2-norm-constrained MVP proposed by DeMiguel et al. (2009) is obtained by solving the following optimization problem:

$$\begin{aligned} \min_w \quad & w'Sw \\ \text{s.t.} \quad & w'e = 1, \\ & w'w \leq \delta. \end{aligned} \tag{21}$$

If S is nonsingular, then there exists a unique global minimizer to the problem in equation (21). That is, at the first-order condition, there exist Lagrange multipliers ρ and ϕ such that

$$2Sw + 2\rho w - \phi e = 0.$$

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

The MVP constructed using $\hat{\Sigma}_{\mathrm{SN}_1}$ can be obtained by solving the following problem:

$$\begin{array}{lll} \min_w & w'(S+\rho I)w \\ \text{s.t.} & w'e = 1. \end{array} \tag{22}$$

Similarly, if S is nonsingular, there exists a unique global minimizer to the problem in equation (22). At the first-order condition, it follows that

$$2Sw + 2\rho w - \phi e = 0. \tag{23}$$

The same first-order condition for the problems in equations (21) and (22) proves that the MVPs under these two settings are equivalent.

Proposition 3. The improved MVP converges to the equally weighted portfolio when the penalty parameter ρ approaches positive infinity or when δ is smaller than the reciprocal of the maximum eigenvalue of the sample covariance matrix. That is,

$$\lim_{\rho\to\infty} w(\rho) = \frac{e}{p} \quad \text{or} \quad \lim_{\delta\to 1/l_{\max}} w(\delta) = \frac{e}{p}. \tag{24}$$

C. Selection of the Regularization Parameter

The proposed estimator, $\hat{\Omega}_{\mathrm{SN}}$, depends on the regularization parameter ρ (for $q=1,2$) or δ (for $q=\infty$). If ρ is 0, the resulting portfolio is the sample MVP. If $\rho\to\infty$ or $\delta\le 1/l_{\max}$, the resulting portfolio is the $1/N$ portfolio. Usually, neither of the two portfolios can achieve the best performance because the sample MVP contains sampling errors, whereas the $1/N$ portfolio ignores information in the sample covariance matrix completely. It is crucial to select the optimal regularization parameter in order to balance information and noise.

The general way to select the regularization parameter is to maximize likelihood by using the K-fold cross-validation. We randomly split the data X into K approximately equal groups, with n_k observations in the kth group. The kth group $X^{(k)}$ is the validation data used to estimate the out-of-sample covariance $S^{(k)}$. The remaining $K-1$ groups, $X^{(-k)}$, are the training data used to estimate the inverse covariance matrix $\hat{\Omega}^{(-k)}$ with a particular regularization parameter. The optimal regularization parameter can be obtained by minimizing the cross-validated negative log-likelihood, as follows:

$$\min \sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)}) = \min \sum_{k=1}^{K} \mathrm{tr}(\hat{\Omega}^{(-k)} S^{(k)}) - \log(\det(\hat{\Omega}^{(-k)})). \tag{25}$$

The sample covariance matrix $S^{(-k)}$ of the training data has the spectral decomposition $S^{(-k)} = Q^{(-k)} L^{(-k)} Q^{(-k)}$. The ith sample eigenvalue, $L_{i,i}^{(-k)}$, can be interpreted as the sample variance of the training data along an axis defined by the ith eigenvector of $S^{(-k)}$. Define $\Gamma^{(k)} = (Q^{(-k)})' S^{(k)} Q^{(-k)}$. Then the ith diagonal element of $\Gamma^{(k)}$, $\Gamma_{i,i}^{(k)}$, can be interpreted as an out-of-sample estimate of this same quantity. Similar to the study of Warton (2008), some insights into the regularization parameter can be obtained from the following results. The proof of Theorems 1 and 2 is similar to that of Warton (2008), which is available from the authors but not presented here for brevity. The proof of Theorem 3 is given in Appendix B.

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

Theorem 1. When $q=1$, the stationary point of the observed negative likelihood estimated by the K-fold cross-validation, $\sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)})$, has the following form:

$$\rho^* = \left[\sum_{k=1}^{K}\sum_{i=1}^{p} v_{i,k}\right]^{-1} \sum_{k=1}^{K}\sum_{i=1}^{p} v_{i,k}(\Gamma_{i,i}^{(k)} - L_{i,i}^{(-k)}), \tag{26}$$

where $v_{i,k} = 1/(L_{i,i}^{(-k)} + \rho^*)^2$.

Theorem 2. When $q=2$, the stationary point of the observed negative likelihood estimated by the K-fold cross-validation, $\sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)})$, satisfies

$$\sum_{k=1}^{K}\sum_{i=1}^{p} \frac{\sqrt{(L_{i,i}^{(-k)})^2 + 8\rho^*}}{G_{i,k}} = \sum_{k=1}^{K}\sum_{i=1}^{p} \frac{2\Gamma_{i,i}^{(k)} - L_{i,i}^{(-k)}}{G_{i,k}}, \tag{27}$$

where $G_{i,k} = \left(\sqrt{(L_{i,i}^{(-k)})^2 + 8\rho^*} + L_{i,i}^{(-k)}\right)^2 \sqrt{(L_{i,i}^{(-k)})^2 + 8\rho^*} \Big/ 4$.

Theorem 3. Consider increasing sample sizes such that as $n_k \to \infty \ \forall k$, if $S \xrightarrow{p} \Sigma$, then $\rho^* \xrightarrow{p} 0$ or $\delta^* \geq 1/\lambda_{\min}$.

Theorem 1 suggests that the expected amount of shrinkage is closely related to the amount of bias in the sample eigenvalues. If the eigenvalues estimated from the training and validation data are similar (i.e., if cross-validation suggests little bias in the sample eigenvalues), then ρ^* is near to 0, implying that there will be little shrinkage. On the contrary, when there are substantial discrepancies in the sample eigenvalues between the training and validation data, we would expect ρ^* to be large and positive. Analogously, we can expect that this property of ρ^* holds in the case of $q=2$. In the case of $q=\infty$, the parameter δ is used to restrict the maximum eigenvalue of the inverse covariance matrix. A smaller δ value implies a tighter constraint. As the discrepancies in the sample eigenvalues between the training and validation data increase, larger shrinkage is expected, and thus a tighter constraint (i.e., a smaller δ) is needed.

Moreover, Theorem 3 suggests that for fixed p, the optimal regularization parameter ρ^* decreases toward 0 or δ^* increases to $1/\lambda_{\min}$ as the sample size increases. This is very straightforward. As the sample size increases, sampling errors in estimating the covariance matrix decrease. Therefore, we expect no shrinkage as the sample size goes to infinity (i.e., ρ decreases toward 0, and δ increases to $1/\lambda_{\min}$.

IV. Evaluation of Portfolio Performance: Setup

A. Data Sets and Performance Metrics

To evaluate the out-of-sample performance of the proposed approach, we use some public data sets that are widely used in the current literature, as listed in Table 2. These data sets consist of well-diversified portfolios. The 100SP data set contains 100 stocks randomly selected from the Standard & Poor's (S&P) 500

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

TABLE 2
Data Description

Table 2 lists the various testing portfolios considered (column 1). Column 2 reports the number of assets in each data set, and column 3 reports the sample period of each data set. Column 4 gives the abbreviation of each portfolio. The 100SP data set contains 100 stocks randomly selected from the Standard & Poor's (S&P) 500 index, which can be downloaded from the Center for Research in Security Prices (CRSP). Columns 5 and 6 show the maximum correlation and average of absolute correlation of the sample covariance matrix for each data set, respectively. The remaining data sets are extracted from the website of Ken French (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

Data	p	Time Period	Descriptor	Maximum Correlation	Mean Absolute Correlation
1	2	3	4	5	6
17 industry portfolios representing the U.S. stock market	17	07/1973–12/2015	17Ind	0.86	0.63
30 industry portfolios representing the U.S. stock market	30	07/1973–12/2015	30Ind	0.87	0.60
49 industry portfolios representing the U.S. stock market	49	07/1973–12/2015	49Ind	0.85	0.56
100 Fama and French portfolios formed on size and book to market (BM)	100	07/1973–12/2015	100FF	0.95	0.73
100 random stocks from S&P 500 universe	100	07/1973–12/2015	100SP	0.72	0.29
Combination of 100 portfolios and 32 portfolios	132	07/1973–12/2015	132S	0.96	0.75
Combination of two 100 Fama and French portfolios	200	07/1973–12/2015	200S	0.98	0.76

index, which can be downloaded from the Center for Research in Security Prices (CRSP). The remaining data sets are extracted from the website of Ken French (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), including 17-, 30-, and 49-industry portfolios (17Ind, 30Ind, and 49Ind, respectively); 100 portfolios formed on size and book-to-market ratio (100FF); the combination of the 100FF and the 32 portfolios formed on size, book-to-market, and operating profitability (132S); and the combination of the 100 portfolios formed on size and operating profitability and the 100 portfolios formed on size and investment (200S).

The out-of-sample performance of portfolios is often assessed based on the following criteria: i) out-of-sample variance, ii) out-of-sample Sharpe ratio (SR), and iii) out-of-sample portfolio turnover (TO). The out-of-sample variance measures the risk of the portfolio. The Sharpe ratio is a common way to examine the performance of an investment by adjusting for its risk, which measures how well the return of an asset compensates the investor for the risk taken. The turnover measures the stability of the portfolio. Lower turnover means lower transaction cost.

The "rolling-window" procedure is carried out to compare portfolio performance. First, a rolling window is selected to perform the estimation. Denote the rolling-window length by n ($n < T$), where T is the total number of periods for a data set. Following DeMiguel et al. (2009), the estimation window of $n = 120$ data points is used in the comparisons, which corresponds to the monthly data over 10 years. Second, the various portfolio weights are computed based on the return data over the estimation window n. Third, this rolling-window procedure is repeated for the next month by including the data for the next month and dropping the data for the earliest month. This procedure is repeated by moving, each time, the rolling window until the end of the data. At the end of this procedure, there

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

are $T-n$ portfolio-weight vectors generated for each portfolio strategy, denoted as w_t, for $t=n,n+1,\ldots,T-1$.

Holding the portfolio w_t for 1 month gives the out-of-sample return at time $t+1$, defined as

$$R_t = w_t' r_{t+1}, \tag{28}$$

where r_{t+1} is a vector of length p of asset returns at time $t+1$. The out-of-sample mean and variance of R_t are given by

$$\hat{\mu}_w = \frac{1}{T-n}\sum_{t=n}^{T-1} R_t, \tag{29}$$

and

$$\hat{\sigma}_w^2 = \frac{1}{T-n-1}\sum_{t=n}^{T-1}(R_t-\hat{\mu}_w)^2, \tag{30}$$

respectively. The out-of-sample Sharpe ratio is defined as the ratio of the mean of R_t to its standard deviation, given by

$$\widehat{\mathrm{SR}} = \frac{\hat{\mu}_w}{\hat{\sigma}_w}. \tag{31}$$

Based on the sequence of $\{w_t\}$, the out-of-sample turnover is computed by

$$\mathrm{TO}_w = \frac{1}{T-n-1}\sum_{t=n}^{T-1}\sum_{j=1}^{p}(|w_{j,t+1}-w_{j,t^+}|), \tag{32}$$

where $w_{j,t}$ is the portfolio weight of asset j at time t, w_{j,t^+} is the portfolio weight before rebalancing, and $w_{j,t+1}$ is the desired portfolio weight of asset j at time $t+1$ (after rebalancing). The turnover measure can be interpreted as the average percentage of wealth traded across the p available assets and over the $T-n-1$ trading periods.

B. Proposed Method and Alternatives

For a given estimator of the inverse covariance matrix, $\hat{\Sigma}^{-1}$, the MVP is

$$w_{\mathrm{MVP}} = \frac{\hat{\Sigma}^{-1}e}{e'\hat{\Sigma}^{-1}e}.$$

Replacing $\hat{\Sigma}^{-1}$ with different estimators leads to different portfolio strategies. By replacing $\hat{\Sigma}^{-1}$ with the proposed Schatten q-norm–constrained estimator $\hat{\Omega}_{\mathrm{SN}}$, we propose the MVP $w_{\text{MVP-SN}}=(e'\hat{\Omega}_{\mathrm{SN}}e)^{-1}\hat{\Omega}_{\mathrm{SN}}e$, which will be denoted as the SN_q portfolio for simplicity. The SN_1, SN_2, and SN_∞ portfolios represent the MVPs with Schatten 1-norm-constrained, 2-norm-constrained, and max-norm-constrained estimates of Ω, respectively.

In addition to maximizing the likelihood, some other calibration criteria could also be considered for the selection of the regularization parameter. For

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

example, DeMiguel, Martin-Utrera, and Nogales (2013) study the calibration criteria based on minimizing the out-of-sample variance and maximizing the out-of-sample Sharpe ratio by using the leave-one-out cross-validation. To illustrate the effect of different ways of selecting the regularization parameter, three different parameter-selection criteria based on maximum likelihood, minimum out-of-sample variance, and maximum out-of-sample Sharpe ratio are considered for the SN_2 portfolio, our primary focus, which are denoted by SN_2-L, SN_2-Var, and SN_2-SR, respectively. For the sake of simplicity, we limit discussion about the other portfolios to the case with the regularization parameter selected based on the maximum-likelihood criterion, unless otherwise stated.

The proposed portfolio will be compared with the following alternative portfolios: i) the MVP based on the sample covariance matrix S, denoted as MVP-S; ii) the equally weighted $(1/N)$ portfolio, $w_{\mathrm{EW}}=(1/p)e$, denoted as EW; iii) the MVP constructed from the shrinkage estimator proposed by Ledoit and Wolf (2004), $\hat{\Sigma}_{\mathrm{LW}}=(1-\alpha)S+\alpha\bar{\sigma}^2 I$, denoted LW-I; iv) the sample-based MVP with the no-short-sale constraint proposed by Jagannathan and Ma (2003), denoted by JM, which minimizes $w'Sw$ subject to $w'e=1$ and $w_i\geq 0$ for $i=1,\ldots,p$; v) the MVP based on the covariance matrix estimated by the first few principal components, denoted by PC (in this article, the number of principal components is selected based on the rule of thumb that they can explain at least 99% of the total variance of asset returns); vi) the MVP based on condition-number-regularized covariance estimation investigated by Won et al. (2013), denoted by CondReg, which minimizes the negative log-likelihood, $\min_{\Omega}\mathrm{tr}(\Omega S)-\log|\Omega|$ subject to $\mathrm{cond}(\Sigma)\leq k_{\max}$, where the condition number of Σ is defined as the ratio of the maximum eigenvalue to the minimum eigenvalue of Σ, (i.e., $\mathrm{cond}(\Sigma)=\lambda_{\max}/\lambda_{\min}$); and vii) the MVP based on the Glasso approach proposed by Goto and Xu (2015). All these MVPs are listed in Table 3.

TABLE 3
Portfolio Strategies

Table 3 lists all the portfolio strategies to be evaluated. SN_2-L,SN_2-Var, and SN_2-SR represent the portfolio strategies based on Schatten 2-norm-regularized inverse covariance matrices with the regularization parameter calibrated by maximizing the likelihood, minimizing the out-of-sample variance, and maximizing the out-of-sample Sharpe ratio, respectively.

Strategies	Abbreviation
Minimum-variance portfolio (MVP) with the sample covariance matrix	MVP-S
MVP based on Schatten 1-norm-regularized inverse covariance matrix calibrated by maximizing the likelihood function	SN_1
MVP based on Schatten ∞-norm-regularized inverse covariance matrix calibrated by maximizing the likelihood function	SN_∞
MVP based on Schatten 2-norm-regularized inverse covariance matrix calibrated by maximizing the likelihood function	SN_2-L
MVP based on Schatten 2-norm-regularized inverse covariance matrix calibrated by minimizing the out-of-sample variance	SN_2-Var
MVP based on Schatten 2-norm-regularized inverse covariance matrix calibrated by maximizing the out-of-sample Sharp ratio	SN_2-SR
MVP with the graphical lasso inverse covariance matrix estimator	Glasso
MVP with Ledoit and Wolf (2004) shrinkage covariance matrix	LW-I
MVP with no-short-sale constraints (Jagannathan and Ma (2003))	JM
Equally weighted $(1/N)$ portfolio	EW
MVP with the covariance matrix computed using a few principal components	PC
MVP with the condition-number-regularized inverse covariance matrix	CondReg

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

C. Descriptives for the Regularization Parameter

Given the rolling-horizon approach for evaluating the out-of-sample portfolio performance, we can search for the optimal regularization parameter of the proposed estimator $\hat{\Omega}_{SN}$ over each rolling window. A grid-searching algorithm can be used to approximate the optimal regularization parameter based on different calibration criteria. Note that the tuning parameter $\rho > 0$ is unbounded, whereas δ is bounded by $0 < \delta \leq 1/l_1$. When searching the optimal ρ, we need to choose a finite interval of ρ initially. In order to ensure that the initially selected interval can cover the optimal ρ, we can choose a relatively wide range for ρ at first and plot the diagnostic check of the objective function against ρ. If the objective function is a monotonic function of ρ over the initial interval, we need to further widen it until it can cover the optimal ρ. If the objective function has a stationary point over the interval selected, the point will be considered as the nearly optimal ρ.

Experience with the cross-validation approach has always yielded a single stationary point. Nevertheless, the possibility of multiple extreme points has not been disapproved. In practice, in order to safeguard this possibility, one can use multiple starting points for the optimizing algorithm or a diagnostic plot of the maximum likelihood or minimum out-of-sample variance.

The proposed approach provides a closed-form of the shrinkage density for the sample eigenvalues for a given regularization parameter. It is very computationally efficient, and we adjust the regularization parameter window by window. For illustration, Figure 1 plots the time-varying regularization parameters for some Schatten-norm-constrained estimators under the 100FF data set. Plots of the regularization parameters for other data sets are also available. As can be seen from Figure 1, when the cross-validated negative likelihood (or variance) decreases, little shrinkage is used (i.e., ρ tends to decrease, and δ tends to increase as expected).

Compared with the proposed estimator, the Glasso estimator is computationally intensive. It would be computationally burdensome to update the regularization parameter window by window. For this reason, Goto and Xu (2015) only optimize the regularization parameter of the Glasso estimator for the first in-sample period and keep it stable over time. In this article, following Goto and Xu (2015), the regularization parameter for the Glasso estimator is optimized for the first rolling window and remains constant over time, which is shown in column 2 of Table 4.

D. Stability of the Estimated Inverse Covariance Matrix $\hat{\Omega}_{SN}$

The condition number of the covariance matrix measures the numerical stability of the inverse (Won et al. (2013)). Note that the condition number of the covariance matrix and the condition number of the inverse covariance matrix are identical. If the condition number is high, say, greater than 10^3 as a rule of thumb, then the matrix is said to be ill-conditioned. Practically, such a matrix is almost singular, and the computation of its inverse is prone to large numerical errors. A matrix that is not invertible has a condition number equal to infinity. In the mean-variance portfolio-optimization problem, the condition number of $\hat{\Sigma}$ or $\hat{\Omega}$ quantifies the sensitivity of the portfolio weights to small changes in the estimated inputs, such as mean and covariance estimates.

FIGURE 1

Time-Varying Regularization Parameters

Figure 1 displays the time-varying regularization parameter of $\hat{\Omega}_{SN}$ over the rolling windows based on the 100FF data set: Graph A SN_1, Graph B SN_2-L, and Graph D SN_∞ are based on minimizing the cross-validated negative log-likelihood; Graph C SN_2-Var is based on minimizing the cross-validated out-of-sample variance.

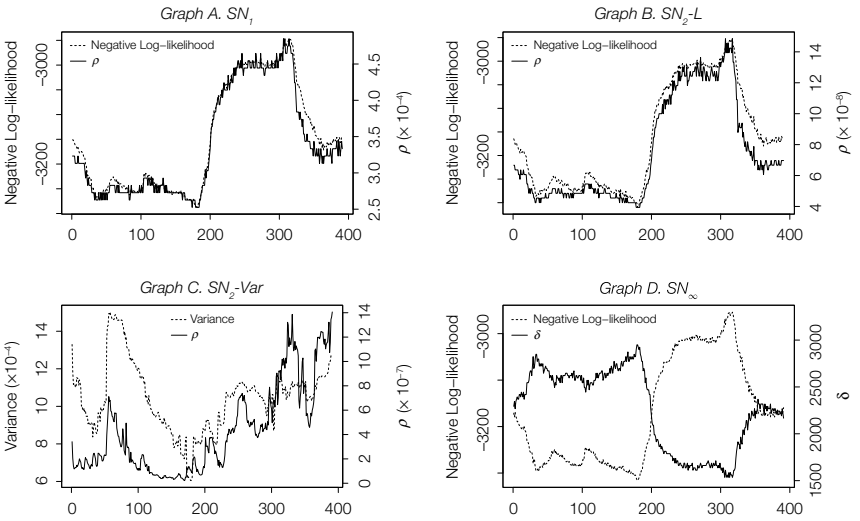


TABLE 4

Regularization Parameter of the Glasso Estimator and Sparsity for Inverse Covariance Matrix Estimators

For each data set, Table 4 reports the regularization parameter ρ chosen from the first rolling window (column 2) for the Glasso estimator and degree of sparsity for different estimators. For the Glasso estimator, the degree of sparsity is measured by the average percentage of 0 off-diagonal elements in the estimated inverse covariance matrix. For the Schatten-norm-regularized estimator, $\hat{\Omega}_{SN}$ is not sparse. The average percentage of the near-to-zero (NTZ) off-diagonal elements in $\hat{\Omega}_{SN}$ is reported, computed by $|\hat{\Omega}_{i,j}|/\sqrt{\hat{\Omega}_{i,i}\hat{\Omega}_{j,j}} < 0.01$.

	Glasso		Percentage of NTZ Off-Diagonal Elements in $\hat{\Omega}_{SN}$				
Data Set	$\rho\,(\times 10^{-5})$	Sparsity	SN_1	SN_∞	SN_2-L	SN_2-Var	SN_2-SR
17Ind	2.857	0.196	0.065	0.059	0.066	0.073	0.085
30Ind	6.050	0.373	0.090	0.100	0.095	0.118	0.153
49Ind	9.596	0.496	0.114	0.146	0.129	0.167	0.236
100FF	8.081	0.717	0.160	0.235	0.190	0.272	0.248
100SP	66.634	0.776	0.166	0.254	0.201	0.369	0.423
132S	8.571	0.781	0.195	0.270	0.227	0.270	0.237
200S	7.895	0.814	0.242	0.331	0.280	0.424	0.428

Table 5 reports the condition numbers for different estimators of the inverse covariance matrix, including S^{-1}, $\hat{\Omega}_{SN}$, the Glasso estimator $\hat{\Omega}_{Glasso}$, the inverse of the Ledoit–Wolf shrunk covariance matrix denoted by $\hat{\Omega}_{LW\text{-}I}$ (i.e., sample covariance matrix shrunk to the identity matrix), and the condition-number-regularized estimator denoted by $\hat{\Omega}_{ConReg}$. Both the mean and standard deviation of condition numbers across the rolling windows are presented.

Table 5 shows that S^{-1} is ill-conditioned for the 49Ind, 100FF, and 100SP data sets and noninvertible for the 132S and 200S data sets. The other estimators,

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

TABLE 5

Condition Numbers for Inverse Covariance Matrix Estimates

For each data set, Table 5 reports the condition numbers of different inverse covariance matrix estimators: S^{-1} is the inverse of the sample covariance matrix; $\hat{\Omega}_{SN}$ is the proposed estimator based on the Schatten norm; $\hat{\Omega}_{Glasso}$ is the Glasso estimator; $\hat{\Omega}_{LW\text{-}I}$ is the inverse of the Ledoit–Wolf shrunk covariance matrix; $\hat{\Omega}_{ConReg}$ is the condition-number-regularized estimator (Won et al. (2013)). Both the mean and standard deviation (Std. Dev.) of the condition numbers across the rolling windows are presented under each data set. — indicates that MVP-S cannot be constructed because of the singularity of the sample covariance matrix.

	17Ind		30Ind		49Ind		100FF		100SP		132S		200S	
Estimators	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
S^{-1}	290.8	137.3	540.7	211.2	1,743.2	898.1	82,812	26,311	10,202	5,404	—	—	—	—
$\hat{\Omega}_{SN_1}$	149.0	65.9	183.2	76.4	240.9	108.3	773.4	180.7	105.7	27.3	1,115.7	297.1	1,380.3	334.2
$\hat{\Omega}_{SN_\infty}$	120.7	27.7	125.3	53.3	153.8	65.3	574.7	130.7	75.2	19.3	881.7	212.5	1205.8	294.3
$\hat{\Omega}_{SN_2\text{-}L}$	130.6	52.9	166.7	55.4	200.9	88.4	693.4	160.3	91.9	23.6	1,026.4	262.0	1,321.8	321.8
$\hat{\Omega}_{SN_2\text{-}Var}$	95.3	61.7	105.6	84.8	116.4	52.3	368.1	166.9	31.7	15.4	789.1	381.0	680.8	508.9
$\hat{\Omega}_{SN_2\text{-}SR}$	36.6	39.3	53.0	79.5	68.4	154.4	685.1	549.1	36.2	80.6	1,532.8	1,014.1	1,229.1	1,419.4
$\hat{\Omega}_{Glasso}$	191.3	84.8	244.3	82.0	412.9	171.5	1,296.5	268.2	204.3	48.3	2,599.8	498.7	3,789.2	456.4
$\hat{\Omega}_{LW\text{-}I}$	228.5	94.6	421.2	149.4	1,079.6	442.7	7,636.4	860.5	2,074.5	604.9	11,395.8	1,107.1	17,833.8	1,548.0
$\hat{\Omega}_{CondReg}$	99.8	109.8	46.2	27.1	21.0	10.2	18.0	4.5	3.0	0.4	17.1	5.0	11.6	2.9

$\hat{\Omega}_{SN}$, $\hat{\Omega}_{Glasso}$, and $\hat{\Omega}_{ConReg}$, can help stabilize S^{-1}. The estimator $\hat{\Omega}_{ConReg}$ achieves the best stability, and $\hat{\Omega}_{SN}$ has better stability than $\hat{\Omega}_{Glasso}$.

It is also interesting to observe from Table 5 that the Glasso estimator is not always well conditioned. For data sets 100FF, 132S, and 200S, $\hat{\Omega}_{Glasso}$ has a high condition number larger than 1,000, indicating that it is ill-conditioned. In contrast, the condition number of $\hat{\Omega}_{SN_2\text{-}Var}$ is just moderate for these data sets. The improved stability of the proposed estimator over the Glasso estimator is due to shrinkage of the sample eigenvalues adopted by the proposed estimator. Although the Glasso estimator can improve its stability by imposing sparsity, it has no control over the condition number. Therefore, there is no guarantee that it is always well conditioned. In general, shrinkage of the sample eigenvalues helps reduce the condition number and thus improves the stability of the estimator. Some evidence can also be found in Daniels and Kass (2001) and Won et al. (2013).

V. Out-of-Sample Evaluation

A. Out-of-Sample Performance

The primary interest of this article is in the ability of $\hat{\Omega}_{SN}$ to reduce the out-of-sample portfolio risk. Table 6 reports the out-of-sample portfolio variances among various portfolios for various data sets. SN_2-Var serves as the benchmark for comparing the out-of-sample performance. To test the statistical significance of the difference in the out-of-sample portfolio variance, we compute the p-values using the stationary bootstrap methodology proposed by Ledoit and Wolf (2011). We choose bootstrap samples of $B = 1{,}000$ and block size $b = 5$. The values in parentheses denote the p-values of testing the null hypothesis that there is no difference in the out-of-sample variance between the SN_2-Var and other portfolios. The smallest out-of-sample variance for each data set is denoted in bold.

MVP-S and EW belong to the class of portfolios without regularization, whereas the other portfolios belong to the class with regularization. Overall, the

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

TABLE 6
Out-of-Sample Portfolio Risk (monthly)

Table 6 reports the monthly out-of-sample variances of the studied portfolios for various data sets. The variances are reported in square percentages (%^{2}). The values in parentheses denote the significance level of testing the null hypothesis that there is no difference in the out-of-sample variance between SN_2-Var and other portfolios. The p-values are computed using the stationary bootstrap (Ledoit and Wolf (2011)) with bootstrap samples of $B=1,000$ and block size $b=5$. — indicates that MVP-S cannot be constructed because of the singularity of the sample covariance matrix.

Strategies	17Ind	30Ind	49Ind	100FF	100SP	132S	200S
MVP-S	12.24	13.92	16.02	61.36	39.66	—	—
	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	—	—
SN_1	**11.58**	12.09	11.82	12.57	12.06	10.54	11.01
	(0.72)	(0.96)	(0.06)	(0.30)	(0.11)	(0.28)	(0.10)
SN_∞	11.85	12.24	11.89	12.48	12.29	10.56	10.62
	(0.31)	(0.51)	(0.11)	(0.46)	(0.03)	(0.33)	(0.42)
SN_2-L	11.60	12.17	11.73	12.45	12.04	**10.52**	10.80
	(0.80)	(0.75)	(0.14)	(0.49)	(0.11)	(0.22)	(0.28)
SN_2-Var	11.63	12.08	**11.48**	**12.24**	11.38	10.78	**10.34**
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
SN_2-SR	12.33	13.93	13.14	13.54	**10.34**	10.78	11.30
	(0.13)	(0.16)	(0.05)	(0.01)	(0.23)	(0.98)	(0.01)
Glasso	11.76	**12.03**	11.76	12.38	11.29	10.98	10.80
	(0.52)	(0.89)	(0.49)	(0.71)	(0.87)	(0.52)	(0.34)
LW-I	11.96	13.40	14.72	21.46	20.78	14.49	18.45
	(0.19)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
JM	11.97	12.43	12.50	17.28	10.51	15.77	15.58
	(0.38)	(0.53)	(0.16)	(0.00)	(0.37)	(0.00)	(0.00)
EW	20.68	21.24	21.68	24.04	21.07	23.83	24.10
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PC	12.29	13.41	13.16	13.93	16.59	12.91	11.74
	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
CondReg	12.53	14.30	16.50	21.52	18.41	21.27	22.49
	(0.09)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

MVPs with regularized covariance matrices achieve further risk reduction compared with those without regularization.

Compared with LW-I, JM, PC, and CondReg, the out-of-sample risk reduction from SN_2-Var is evident and statistically significant, especially for the last four data sets with a large number of stocks. Compared with Glasso, SN_2-Var achieves competitive out-of-sample risk. The difference is negligible. However, the latter possesses the advantages of being computationally efficient and structure-free.

It is important to note that although $\hat{\Omega}_{\text{CondReg}}$ has better stability than $\hat{\Omega}_{\text{SN}_2\text{-Var}}$, CondReg is outperformed by SN_2-Var. This is a result of overshrinkage of the sample eigenvalues by the CondReg portfolio. For data sets 100FF, 100SP, 132S, and 200S, the condition number of $\hat{\Omega}_{\text{CondReg}}$ is restricted to be a low level (less than 18 in Table 5). Clearly, the sample eigenvalues are overshrunk, which induces a large bias in the estimation of eigenvalues. These results suggest that constraining the condition number indeed improves the stability but, meanwhile, may overshrink the sample eigenvalues. By contrast, the Schatten norm constraint, especially the Schatten 2-norm of the sample eigenvalues, turns out to be particularly effective in suppressing fluctuations of the sample eigenvalues.

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

It is important to note that the relative performance of a portfolio may depend on the underlying data sets, estimation windows, and performance measures, as well as the regularization parameters. The comparisons here are limited. There is no intention to claim that the proposed portfolio possesses general superiority compared with the existing portfolios. However, the empirical studies are sufficient for us to show that the proposed portfolio has a formidable ability to reduce the out-of-sample portfolio risk, especially for high-dimensional portfolios with certain highly correlated assets. Its favorable out-of-sample performance in many real situations suggests another feasibility for improving the MVP risk with uncertainty in the covariance matrix.

B. Out-of-Sample Sharpe Ratio and Turnover

Table 7 reports the monthly out-of-sample Sharpe ratios of different strategies for various data sets. As Goto and Xu (2015) point out, detecting reliable differences in the Sharpe ratio is difficult because of the large estimation errors in mean returns. Similar to Goto and Xu (2015), we employ the bootstrap method of Ledoit and Wolf (2008) to test the null hypothesis of no difference in the Sharpe ratio. The highest Sharpe ratio for each data set is denoted in bold.

TABLE 7

Out-of-Sample Sharpe Ratio (monthly)

Table 7 reports the monthly out-of-sample Sharpe ratios of the studied portfolios for various data sets. The values in parentheses denote the significance level of testing the null hypothesis that there is no difference in the Sharpe ratio between SN_2-Var and other portfolios. The p-values are computed using the stationary bootstrap (Ledoit and Wolf (2008)) with bootstrap samples of B = 1,000 and block size b = 5. — indicates that MVP-S cannot be constructed because of the singularity of the sample covariance matrix.

Strategies	17Ind	30Ind	49Ind	100FF	100SP	132S	200S
MVP-S	0.229	0.249	0.190	0.158	0.155	—	—
	(0.00)	(0.02)	(0.00)	(0.00)	(0.06)	(—)	(—)
SN_1	0.259	0.291	0.255	0.378	0.262	0.375	0.365
	(0.14)	(0.19)	(0.01)	(0.80)	(0.08)	(0.25)	(0.44)
SN_∞	0.253	0.30	0.258	0.374	0.257	0.361	0.361
	(0.14)	(0.92)	(0.06)	(0.95)	(0.04)	(0.70)	(0.23)
SN_2-L	0.261	0.293	0.259	**0.378**	0.262	0.371	0.365
	(0.29)	(0.40)	(0.03)	(0.84)	(0.07)	(0.47)	(0.47)
SN_2-Var	0.267	**0.300**	**0.273**	0.375	0.286	0.365	0.381
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
SN_2-SR	0.279	0.279	0.264	0.363	**0.320**	0.376	0.365
	(0.47)	(0.40)	(0.64)	(0.42)	(0.36)	(0.47)	(0.08)
Glasso	0.258	0.281	0.253	0.370	0.293	**0.386**	**0.386**
	(0.31)	(0.16)	(0.16)	(0.73)	(0.72)	(0.14)	(0.84)
LW-I	0.239	0.260	0.201	0.316	0.214	0.363	0.305
	(0.01)	(0.04)	(0.00)	(0.17)	(0.16)	(0.97)	(0.09)
JM	0.271	0.261	0.253	0.265	0.320	0.277	0.274
	(0.85)	(0.09)	(0.45)	(0.00)	(0.40)	(0.04)	(0.02)
EW	0.216	0.216	0.211	0.215	0.271	0.210	0.216
	(0.25)	(0.06)	(0.14)	(0.00)	(0.76)	(0.01)	(0.01)
PC	**0.288**	0.298	0.221	0.311	0.205	0.323	0.343
	(0.55)	(0.39)	(0.53)	(0.02)	(0.24)	(0.27)	(0.13)
CondReg	0.246	0.272	0.233	0.232	0.287	0.224	0.224
	(0.17)	(0.33)	(0.26)	(0.00)	(0.98)	(0.00)	(0.00)

SN_2-Var attains a higher Sharpe ratio than LW-I, JM, EW, and CondReg for nearly all data sets. The difference is significant for the first three data sets as compared with LW-I, and it is significant for the last two data sets as compared with JM, EW, and CondReg. Also, SN_2-Var outperforms PC for the 100FF data set and performs better than PC for the remaining data sets, except the 17Ind data set. Compared with Glasso, there is no clear winner.

SN_2-Var selects the regularization parameter based on minimizing the out-of-sample variance, whereas SN_2-SR selects it based on maximizing the out-of-sample Sharpe ratio. The performance of SN_2-SR and SN_2-Var is comparable in terms of the Sharpe ratio. SN_2-SR attains the highest Sharpe ratio for the data set 100SP, whereas SN_2-Var attains the highest for the 30Ind and 49Ind data sets.

Table 8 reports the monthly turnover of different portfolios for different data sets. The turnover can reflect the stability of each portfolio. Consistent with the result of DeMiguel et al. (2009), EW dominates other portfolios in the sense that it provides the lowest turnover for all the data sets considered. This is not surprising, because the EW portfolio only involves small contrarian rebalancing but does not involve any hedge trades (Goto and Xu (2015)). The no-short-sale constraint and the condition-number constraint also keep the turnover of JM and CondReg relatively low, respectively.

TABLE 8
Portfolio Turnover (monthly)

Table 8 reports the monthly turnover of the studied portfolios for various data sets. The smallest turnover for each data set is denoted in bold. — indicates that MVP-S cannot be constructed because of the singularity of the sample covariance matrix.

Strategies	17Ind	30Ind	49Ind	100FF	100SP	132S	200S
MVP-S	0.262	0.454	0.782	7.853	2.089	—	—
SN_1	0.198	0.265	0.321	0.695	0.287	0.812	0.963
SN_∞	0.225	0.261	0.310	0.642	0.270	0.744	0.877
SN_2-L	0.195	0.260	0.303	0.648	0.268	0.756	0.893
SN_2-Var	0.173	0.220	0.241	0.492	0.172	0.732	0.648
SN_2-SR	0.133	0.184	0.209	0.850	0.203	0.992	0.901
Glasso	0.207	0.264	0.320	0.646	0.202	0.678	0.828
LW-I	0.236	0.402	0.645	2.204	0.969	2.300	2.602
JM	0.059	0.071	0.074	0.113	0.120	0.107	0.127
EW	**0.026**	**0.030**	**0.032**	**0.023**	**0.060**	**0.022**	**0.022**
PC	0.331	0.600	0.793	2.047	1.160	2.150	2.150
CondReg	0.205	0.176	0.125	0.092	0.069	0.087	0.066

SN_2-Var outperforms LW-I, PC, and MVP-S and achieves slightly lower turnover than Glasso but higher turnover than EW, JM, and CondReg. Not surprisingly, the turnover of MVP-S is the highest for the high-dimensional data sets. The large estimation errors in the sample covariance matrix lead to unstable portfolio weights and thus extremely high turnover.

C. The Behavior of Optimized Portfolio Weights

One serious issue in portfolio risk minimization is that estimation errors in the covariance matrix could yield unstable weights that take extreme values. It is very interesting to compare the behavior of the optimized portfolio weights generated from different strategies. Table 9 reports the minimum, the 5th and 95th percentiles, the maximum, the percentage of negative weights (Neg), the Herfindahl index for each portfolio, and the Manhattan distance (MD) between

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

TABLE 9
Distribution of Portfolio Weights

Table 9 reports the distributions of weights of the studied portfolios for various data sets. For each rolling window, the minimum value, 5th percentile, 95th percentile, and the maximum value of the monthly weights are calculated. Also, the Manhattan distance (MD) between the estimated portfolio and the $1/N$ portfolio, the percentage of negative weights ("Neg."), and the Herfindahl index (%^{2}) of the monthly weights are calculated. The average values of these summary statistics over all the rolling windows are reported.

Data Set	Strategies	Minimum	5%	95%	Maximum	MD	Neg.	Herfindahl
100FF	MVP-S	−0.844	−0.521	0.564	0.993	26.297	0.497	118,235
	SN_1	−0.175	−0.119	0.145	0.210	6.179	0.455	6,402
	SN_∞	−0.182	−0.121	0.140	0.209	6.177	0.450	6,336
	SN_2-L	−0.174	−0.119	0.143	0.206	6.100	0.454	6,245
	SN_2-Var	−0.146	−0.094	0.114	0.169	4.865	0.435	4,150
	SN_2-SR	−0.075	−0.047	0.092	0.167	1.758	0.307	1,534
	Glasso	−0.164	−0.111	0.146	0.241	6.129	0.473	6,323
	LW-I	−0.383	−0.233	0.278	0.468	12.567	0.488	25,859
	JM	0.000	0.000	0.052	0.342	1.852	0.000	2,397
	PC	−0.224	−0.138	0.150	0.236	6.999	0.441	8,136
	CondReg	−0.014	−0.007	0.026	0.035	0.786	0.143	202
200S	MVP-S	—	—	—	—	—	—	—
	SN_1	−0.124	−0.073	0.087	0.135	7.922	0.470	4,972
	SN_∞	−0.124	−0.071	0.082	0.130	7.674	0.466	4,637
	SN_2-L	−0.122	−0.071	0.084	0.130	7.728	0.469	4,717
	SN_2-Var	−0.095	−0.054	0.065	0.098	5.977	0.450	3,036
	SN_2-SR	−0.099	−0.059	0.071	0.107	6.426	0.447	3,903
	Glasso	−0.124	−0.068	0.084	0.146	7.386	0.476	4,452
	LW-I	−0.238	−0.143	0.159	0.274	14.746	0.488	17,284
	JM	0.000	0.000	0.002	0.377	1.926	0.000	2,565
	PC	−0.135	−0.075	0.081	0.128	7.728	0.452	4,768
	CondReg	−0.004	−0.000	0.010	0.014	0.521	0.063	72

each portfolio and the EW portfolio. The Herfindahl index measures the degree of concentration, computed by $\frac{1}{T-n}\sum_{t=n}^{T-1}\sum_{i=1}^{p} w_{i,t}^2$, where $w_{i,t}$ denotes the optimized weight on asset i at period t. The Manhattan distance measures the deviation of the underlying portfolio weights from the $1/N$ portfolio weights, computed by $\frac{1}{T-n}\sum_{t=n}^{T-1}\sum_{i=1}^{p}|w_{it}-1/p|$. In portfolio theory, the effective number of positions held in a portfolio is the reciprocal of the Herfindahl index. A low Herfindahl index implies a very diversified portfolio. To conserve space, Table 9 reports results based on only the two data sets for which MVP-S is inappropriate, 100FF and 200S. The results for other data sets are also available, with similar observations.

In the 100FF data set, MVP-S takes very large portfolio weights. For example, the minimum and maximum weights from MVP-S are as extreme as −0.844 and 0.993, respectively. The extremely large value of the Herfindahl index (118,235%^{2}) is also consistent with the presence of extreme portfolio weights. Due to singularity, MVP-S cannot be constructed for the 200S data set.

We can observe a general tendency from Table 9 that CondReg has a smaller Herfindahl index than SN_2-Var and Glasso. All these portfolios use both long and short positions of many assets. The relatively small Herfindahl index implies that the underlying portfolio is more aggressive in reducing extreme weights and that the underlying portfolio is more diversified.

D. Economic Gains

An ideal portfolio is expected to have a high mean of returns, low variance, and low turnover. However, there is often a need to make a trade-off between portfolio risk and turnover. In order to reduce the portfolio risk, frequent rebalancing is usually required, which in turn leads to higher turnover and a higher transaction

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

cost. To assess the economic significance of portfolio risk reduction, the effects of the transaction cost need to be taken into account. For this reason, it is also of interest to compare the certainty-equivalent excess return of each portfolio after subtracting the transaction cost, defined as

$$\text{CER_ADJUSTED} = \hat{\mu}_w - \frac{\gamma}{2}\hat{\sigma}_w^2 - \text{TO}_w \times C, \tag{33}$$

where γ is the risk-aversion coefficient, and C is the transaction cost per trade. In this article, we set γ to be the popular choice of $\gamma = 5$ and C to be 50 basis points (bps) per trade.

Table 10 reports the CER_ADJUSTED values of different portfolios for various data sets. Clearly, the family of SN_q portfolios can achieve positive economic gains over MVP-S for all data sets. LW-I and PC have relatively low CER_ADJUSTED values, especially for the data sets with a relatively large number of assets, such as 100FF, 100SP, 132S, and 200S. This is a result of their high turnover in these situations.

TABLE 10
Net Certainty Equivalent Return (monthly)

Table 10 reports the values of CER_ADJUSTED of the studied portfolios for various data sets. The transaction cost of each is calculated as 50 basis points per trade. For each data set, the highest value of CER_ADJUSTED is denoted in bold. — means "not applicable."

Strategies	17Ind	30Ind	49Ind	100FF	100SP	132S	200S
MVP-S	0.364	0.354	−0.031	−4.227	−1.057	—	—
SN_1	0.493	0.576	0.421	0.680	0.466	0.549	0.454
SN_∞	0.462	0.616	0.438	0.690	0.460	0.537	0.473
SN_2-L	0.503	0.586	0.442	0.697	0.475	0.561	0.483
SN_2-Var	0.534	**0.631**	0.517	**0.761**	0.595	0.561	**0.643**
SN_2-SR	0.605	0.601	0.522	0.573	0.669	0.467	0.494
Glasso	0.488	0.543	0.414	0.668	0.601	**0.666**	0.584
LW-I	0.410	0.417	0.080	−0.176	−0.030	−0.131	−0.452
JM	**0.608**	0.575	**0.543**	0.615	0.718	0.654	0.628
EW	0.450	0.450	0.424	0.440	0.688	0.416	0.446
PC	0.538	0.456	0.076	−0.210	−0.160	−0.239	−0.195
CondReg	0.455	0.583	0.471	0.494	**0.738**	0.457	0.468

The economic gains from the proposed portfolio also compare favorably with JM, Glasso, and CondReg. As Table 10 shows, SN_2-Var can achieve the highest economic gains from portfolio risk reduction for three data sets (30Ind, 100FF, and 200S), whereas JM can achieve the highest economic gains for two data sets only (17Ind and 49Ind). Glasso and CondReg can achieve the highest economic gains for one data set only (132S and 100SP, respectively).

E. Under the Randomized Samples of Asset Returns

The previous comparison under each data set is limited to a single sample. To reduce the effect of randomness on the out-of-sample performance, we randomly choose 100 assets out of the 200S data set at a time to form 100 test samples. The comparison results across the 100 samples would be more reliable to assess the ability of each portfolio strategy. This exercise also allows us to examine the relationship between the condition number of an estimator and the magnitude of out-of-sample portfolio risk reduction achieved by it across the 100 samples.

In Table 11, we first report the summary statistics of the estimation results for the out-of-sample risk, Sharpe ratio, and turnover, including monthly mean,

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

TABLE 11

Distribution of Out-of-Sample Risk, Sharpe Ratio and Turnover, and Risk Reduction Versus Condition-Number Ratio for 100 Randomized Samples Selected from the 200S Data Set

Table 11 compares the SN_q and Glasso portfolios in terms of distributions of out-of-sample variance (%^{2}) (Panel A), out-of-sample Sharpe ratio (Panel B), and turnover (Panel C) over 100 randomized samples, each consisting of 100 assets randomly selected from the 200S data set. For each panel, we compute the mean, standard deviation (Std. Dev.), minimum, and maximum of the realized monthly out-of-sample variance, Sharpe ratio, and turnover. "No. of Glasso > ALT" counts the number of times that Glasso has achieved a higher realized monthly variance, Sharpe ratio, and turnover than each alternative portfolio in the 100 randomized samples. Panel D reports the results regressing the magnitude of out-of-sample portfolio risk reduction ($\sigma_{\text{Glasso}} - \sigma_{\text{SN}_2\text{-Var}}$) on the condition-number ratio, $\text{cond}(\hat{\Omega}_{\text{Glasso}})/\text{cond}(\hat{\Omega}_{\text{SN}_2\text{-Var}})$, across the 100 randomized samples. Panel E groups the 100 randomized samples by the condition-number ratio into bottom-30, middle-40, and top-30 subsamples and then reports the average risk reduction within each group.

	Glasso	MPV-S	SN_1	SN_∞	SN_2-L	SN_2-Var	SN_2-SR
Panel A. Out-of-Sample Variance (monthly; %^{2})							
Mean	11.56	53.11	11.62	11.57	11.62	11.35	11.43
Std. dev.	0.52	6.91	0.54	0.57	0.54	1.44	1.72
Maximum	12.74	97.10	13.11	13.12	13.07	22.79	25.51
Minimum	10.38	41.84	10.39	10.20	10.37	9.88	10.00
No. of Glasso > ALT		0	38	49	39	89	83
No. of Glasso < ALT		100	62	51	61	11	17
Panel B. Out-of-Sample Sharpe Ratio (monthly)							
Mean	0.366	0.182	0.351	0.351	0.351	0.360	0.360
Std. dev.	0.022	0.047	0.021	0.021	0.021	0.025	0.026
Maximum	0.425	0.345	0.402	0.399	0.402	0.406	0.405
Minimum	0.302	0.071	0.284	0.290	0.283	0.257	0.238
No. of Glasso > ALT		100	95	92	95	58	58
No. of Glasso < ALT		0	5	8	5	42	42
Panel C. Portfolio Turnover (monthly)							
Mean	0.651	7.353	0.664	0.652	0.665	0.483	0.434
Std. dev.	0.017	0.370	0.017	0.019	0.018	0.264	0.318
Maximum	0.685	8.444	0.710	0.714	0.703	2.294	2.697
Minimum	0.609	6.576	0.624	0.600	0.618	0.254	0.206
No. of Glasso > ALT		0	10	43	11	89	90
No. of Glasso < ALT		100	90	57	89	11	10

Panel D. Risk Reduction (monthly % $\sigma_{Glasso} - \sigma_{SN_2\text{-}Var}$) Regressed on cond($\hat{\Omega}_{Glasso}$)/cond($\hat{\Omega}_{SN_2\text{-}Var}$)

Intercept	*t*-Statistic	*p*-Value	Ratio	*t*-Statistic	*p*-Value	R^2
3.215×10^{-4}	(2.610)	0.010	9.475×10^{-5}	(3.937)	<0.001	13.66%

Panel E. Risk Reduction (monthly % $\sigma_{Glasso} - \sigma_{SN_2\text{-}Var}$) across Condition-Number-Ratio Groups

Condition-Number-Ratio Group	Bottom 30%	Middle 40%	Top 30%
Risk reduction	−0.077	0.083	0.087

standard deviation, and minimum and maximum values. Because SN_q dominates CondReg in the preceding comparisons, the present study mainly focuses on the comparison between SN_q and Glasso. Also, we report the fraction of runs for which SN_q yields higher and lower portfolio variance, Sharpe ratio, and turnover as compared with Glasso.

Panel A of Table 11 shows that SN_2-Var can achieve lower out-of-sample portfolio risk than Glasso in 89 runs out of 100. This is clear evidence of the capability of reducing portfolio risks through SN_2-Var. Moreover, Panel B shows that SN_2-Var has lower turnover than Glasso in 89 runs out of 100. Yet Panel C shows that the former has a lower Sharpe ratio than the latter in a majority of runs.

In Panels D and E of Table 11, we then report the relation between the out-of-sample portfolio risk reduction ($\sigma_{\text{Glasso}} - \sigma_{\text{SN}_2\text{-Var}}$) and the condition-number ratio across the 100 randomized samples, defined by $\text{cond}(\hat{\Omega}_{\text{Glasso}})/\text{cond}(\hat{\Omega}_{\text{SN}_2\text{-Var}})$. Panel D shows the regression results when regressing $\sigma_{\text{Glasso}} - \sigma_{\text{SN}_2\text{-Var}}$ on the

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

condition-number ratio $\text{cond}(\hat{\Omega}_{\text{Glasso}})/\text{cond}(\hat{\Omega}_{\text{SN}_2\text{-Var}})$. The slope coefficient is 9.475×10^{-5} (t-statistic = 3.937). That is, when the condition-number ratio is 1% higher in a sample than in other samples, SN_2-Var tends to achieve 9.475×10^{-5}% more portfolio risk reduction (per month) than Glasso. In Panel E, we sort these randomized samples by the condition-number ratio into three groups: bottom-30, middle-40, and top-30 subsamples. The average risk reductions in the bottom-30, middle-40, and top-30 groups are –0.077%, 0.083%, and 0.087% per month, respectively. This shows a tendency that SN_2-Var achieves greater out-of-sample portfolio risk reduction in groups with a higher condition-number ratio. These results suggest that our approach tends to achieve greater portfolio risk reduction relative to Glasso in samples for which the Glasso estimate has a high condition number.

VI. Detailed Comparisons between the SN_q and Glasso Portfolios under Hypothesized Structures

Note that the Glasso estimate is a structured estimate. Intuitively, the structured estimator works well when the hypothesized structure is correct or close, but it may perform poorly when the hypothesized structure is incorrect. On the contrary, the proposed estimate is an unstructured estimate without any assumption of matrix structure. In reality, the true covariance matrix of returns or its inverse is rarely known. It is likely that there exists a mismatch between the hypothesized structure and the true one. It is interesting to compare the out-of-sample performance between the Glasso and proposed portfolios under misspecification of matrix structure.

A number of models with different covariance structures are considered. The first two models belong to the case in which the inverse covariance matrix is not sparse and include the following:

- Model 1: The covariance matrix is a symmetric Toeplitz matrix of the form $\Sigma = (c_{i,j})$, where $c_{i,j} = c_{i+1,j+1} = c_{j-i} = c_{i-j}$. For simplicity, we set the difference between c_i and c_{i+1} to be a constant (i.e., $c_i - c_{i+1} = 0.2/(p-2)$ for $i = 1, 2, \ldots, p-2$).

- Model 2: The covariance matrix is $\Sigma = [\text{diag}(R)]^{-1/2} R [\text{diag}(R)]^{-1/2}$, where $R = X'X + D$. Matrix X is a $p \times p$ random matrix whose elements are taken from a uniform distribution $U(0, a)$, and the error matrix D is a $p \times p$ diagonal matrix whose diagonal entries are drawn from a uniform distribution $U(0, 1)$.

The remaining models belong to the case when the inverse covariance matrix is sparse. Following Cai, Liu, and Xia (2014), we consider four models of the following structures:

- Model 3: A diagonal covariance matrix $\Sigma = \text{diag}(c_{1,1}, \ldots, c_{p,p})$. In this case, all the assets are independent, and the inverse covariance matrix is also diagonal.

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

- Model 4: a block diagonal covariance matrix of the form $\Sigma=(c_{i,j})$, where $c_{i,i}=1$, $c_{i,j}=c$ $(-1<c<1)$ for $2(k-1)+1\leq i\neq j\leq 2k$, where $k=1,\ldots,[p/2]$, and $c_{ij}=0$ otherwise. In this case, the inverse covariance matrix is also block diagonal.

- Model 5: An exponential decay covariance matrix of the form $\Sigma=(c_{i,j})$, where $c_{ij}=\text{sign}(c)|c|^{|i-j|}$ for $1\leq i,j\leq p$.

- Model 6: A banded inverse covariance matrix of the form $\Omega=(\omega_{i,j})$, where $\omega_{i,i}=2$ for $i=1,\ldots,p$; $\omega_{i,i+1}=4a$ for $i=1,\ldots,p-1$; $\omega_{i,i+2}=2a$ for $i=1,\ldots,p-2$; $\omega_{i,i+3}=2a$ for $i=1,\ldots,p-3$; $\omega_{i,i+4}=a$ for $i=1,\ldots,p-4$; $\omega_{i,j}=\omega_{j,i}$ for $i,j=1,\ldots,p$; and $\omega_{i,j}=0$ otherwise. The corresponding covariance matrix is given by $\Sigma=\Omega^{-1}$.

Similar to settings of the FF100 data set with 510 observations, independent samples are generated with a sample size of 510 from a multivariate normal distribution with 0 mean and covariance structure Σ of dimension 100×100. Table 12 compares the out-of-sample portfolio variances between the SN_q and Glasso portfolios based on models 1 and 2. Under these two models, the hypothesized structure of sparsity is incorrect for the Glasso estimator. Not surprisingly, SN_2-Var has much smaller out-of-sample variances than Glasso. The differences are statistically significant at 1% in most cases.

TABLE 12
Out-of-Sample Portfolio Risk for Models 1 and 2

Table 12 reports the out-of-sample variances of the SN_q and Glasso portfolios for models 1 and 2. The null hypothesis is that there is no difference in the out-of-sample variance between SN_2-Var and the alternatives. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Model 1	$c_1=0.3$	$c_1=0.4$	$c_1=0.5$	$c_1=0.6$	$c_1=0.7$	$c_1=0.8$
SN_1	0.2697***	0.3202***	0.5655***	0.6998***	0.7999***	0.8557***
SN_∞	0.2679***	0.3208***	0.5665***	0.7040***	0.8105***	0.8542***
SN_2-L	0.2681***	0.3193***	0.5653***	0.7011***	0.8033***	0.8539***
SN_2-Var	0.2363	0.2860	0.4644	0.5526	0.6699	0.7251
Glasso	0.2761***	0.3304***	0.5723***	0.7280***	0.8420***	0.8922***
Model 2	$a=0.05$	$a=0.1$	$a=0.3$	$a=0.5$	$a=0.7$	$a=1$
SN_1	0.0575*	0.1275	0.5583	0.8005	0.8351	0.8555
SN_∞	0.0545***	0.1218*	0.5802*	0.8333***	0.8827*	0.9132**
SN_2-L	0.0560	0.1241	0.5646	0.8113*	0.8539	0.8786
SN_2-Var	0.0566	0.1261	0.5618	0.7552	0.8127	0.8327
Glasso	0.0637***	0.1332*	0.5863*	0.8565***	0.9336**	0.9524***

To gain more insight into the misspecification of matrix structure, Table 13 reports the time-series average of the off-diagonal elements of the true covariance matrix and its inverse, sparsity of $\hat{\Omega}_{\text{Glasso}}$, the average Manhattan distance between the portfolio weights and true weights, and the average condition numbers of the estimated inverse covariance matrices. For brevity, only the results based on model 1 are presented. It is surprising to observe that the Glasso estimator shrinks more than 75% of the off-diagonal elements toward 0, although the true inverse covariance matrix is not sparse. This leads to a great degree of mismatch between the Glasso estimate and the true inverse covariance matrix.

Table 14 compares the out-of-sample variances between the SN_q and Glasso portfolios based on model 3. This is an extreme case in which all the assets are

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

TABLE 13

Mean of Off-Diagonal Elements of the True Covariance Matrix and Its Inverse, Manhattan Distance, and Condition Number for Model 1

For model 1, Table 13 reports the mean of off-diagonal elements of Ω, $\bar{\varphi}_{\text{off}}=(1/p(p-1))\sum_{i\neq j}|\Omega_{i,j}|$; the mean of off-diagonal elements of Σ, $\bar{c}_{\text{off}}=(1/p(p-1))\sum_{i\neq j}|\Sigma_{i,j}|$; the sparsity of the Glasso estimator; the Manhattan distances between the estimated and true portfolio weights; and the condition numbers of the inverse covariance matrix estimates. The average values of these summary statistics over all the rolling windows are reported.

Model 1	$c_1=0.3$	$c_1=0.4$	$c_1=0.5$	$c_1=0.6$	$c_1=0.7$	$c_1=0.8$
$\bar{c}_{\text{off}}$	0.233	0.333	0.433	0.533	0.633	0.733
$\bar{\varphi}_{\text{off}}$	0.014	0.016	0.019	0.024	0.032	0.047
Sparsity of $\hat{\Omega}_{\text{Glasso}}$	0.757	0.760	0.756	0.762	0.757	0.753
Manhattan Distances between the Estimated and True Portfolio Weights						
SN_1	2.151	2.632	3.355	4.166	5.245	6.921
SN_∞	2.056	2.511	3.343	4.143	5.301	6.837
SN_2-L	2.097	2.574	3.339	4.150	5.263	6.888
SN_2-Var	0.924	1.062	0.866	0.951	1.220	1.825
Glasso	2.067	2.618	3.584	4.415	5.827	7.264
Condition Numbers for Estimated Ω						
SN_1	46.51	70.87	129.45	185.17	301.48	474.01
SN_∞	32.70	50.31	91.55	132.00	215.44	339.64
SN_2-L	40.70	62.45	114.41	164.27	268.30	421.85
SN_2-Var	7.36	11.51	6.18	2.63	12.67	40.65
Glasso	44.38	70.82	135.70	204.54	337.84	526.08

TABLE 14

Out-of-Sample Portfolio Risk, Manhattan Distance, and Condition Number for Model 3

For model 3, Table 14 reports the out-of-sample variances of the SN_q and Glasso portfolios, the Manhattan distances between the estimated and true portfolio weights, and the condition numbers of the inverse covariance matrix estimates. The null hypothesis is that there is no difference in the out-of-sample variance between SN_2-Var and the alternatives. Different variances in the diagonal covariance matrix are considered, including i) variances generated from two uniform distributions U(0.003, 0.005) and U(0.002, 0.006), ii) variances corresponding to the diagonal entries taken from the sample covariance matrix of the FF100 data set, iii) variances of 1 (i.e., $\Sigma=I$), and iv) variances of 0.004. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Strategies	U(0.003, 0.005)	U(0.002, 0.006)	FF100	I	0.004
Panel A. Out-of-Sample Portfolio Variance					
SN_1	0.0000549***	0.0000499***	0.0000383***	0.0116831***	0.0000478***
SN_∞	0.0000509***	0.0000445***	0.0000352	0.0106562**	0.0000431***
SN_2-L	0.0000523***	0.0000469***	0.0000363*	0.0110788***	0.0000450***
SN_2-Var	0.0000451	0.0000405	0.0000346	0.0100034	0.0000397
Glasso	0.0000442	0.0000369**	0.0000314***	0.0101402	0.0000393
Panel B. Manhattan Distances between the Estimated and True Portfolio Weights					
SN_1	0.413	0.392	0.365	0.385	0.379
SN_∞	0.280	0.302	0.257	0.253	0.233
SN_2-L	0.344	0.338	0.308	0.319	0.307
SN_2-Var	0.128	0.254	0.213	0.060	0.025
Glasso	0.130	0.122	0.123	0.119	0.126
Panel C. Condition Numbers for Estimated Ω					
SN_1	6.23	6.80	7.28	6.11	6.04
SN_∞	3.74	4.19	4.51	3.67	3.60
SN_2-L	4.87	5.37	5.77	4.78	4.71
SN_2-Var	1.16	1.40	2.41	1.41	1.19
Glasso	2.88	4.55	4.71	2.14	2.10

uncorrelated. Different variances in the diagonal covariance matrix are considered, including i) variances generated from two uniform distributions U(0.003, 0.005) and U(0.002, 0.006), ii) variances corresponding to the diagonal entries taken from the sample covariance matrix of the FF100 data set, iii) variances of

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

1 (i.e., $\Sigma = I$), and iv) variances of 0.004. The former three cases have unequal variances, and the last two cases have equal variances.

In Panel A of Table 14, we can see that Glasso has smaller out-of-sample variances than SN_2-Var under unequal variances. This observation is not surprising because the hypothesized structure of the Glasso estimate coincides with the true structure in this case. Under uncorrelated assets with equal variances, the difference in the out-of-sample variance between Glasso and SN_2-Var becomes negligible. In this case, both of them converge to the $1/N$ portfolio.

Models 4–6 represent less extreme cases where the inverse covariance matrix is sparse but not diagonal, compared with model 3. Table 15 compares the out-of-sample portfolio variances between the SN_q and Glasso portfolios based on models 4–6. The results show that the SN_2-Var portfolio has nearly the same out-of-sample variance as Glasso when the true inverse covariance matrix is relatively sparse (model 4), but it can have smaller out-of-sample variances than the latter when the true inverse covariance matrix is less sparse (models 5 and 6).

TABLE 15
Out-of-Sample Portfolio Risk for Models 4–6

Table 15 reports the out-of-sample portfolio variances of the SN_q and Glasso portfolios for models 4–6. The null hypothesis is that there is no difference in the out-of-sample variance between SN_2-Var and the alternatives. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Model 4	$c=-0.5$	$c=-0.3$	$c=-0.1$	$c=0.1$	$c=0.3$	$c=0.5$
SN_1	0.0060***	0.0081***	0.0127***	0.0121***	0.0162***	0.0213***
SN_∞	0.0051*	0.0070**	0.0115***	0.0106**	0.0146***	0.0198***
SN_2-L	0.0056***	0.0076***	0.0120***	0.0113***	0.0153***	0.0202***
SN_2-Var	0.0048	0.0066	0.0104	0.0099	0.0133	0.0163
Glasso	0.0050	0.0067	0.0106	0.0100	0.0136	0.0177***
Model 5	$c=-0.3$	$c=-0.1$	$c=0.1$	$c=0.3$	$c=0.5$	$c=0.7$
SN_1	0.0019***	0.0108***	0.0146***	0.0224***	0.0366***	0.0626***
SN_∞	0.0016	0.0097***	0.0134***	0.0213***	0.0363***	0.0641***
SN_2-L	0.0017**	0.0102***	0.0138***	0.0215***	0.0359***	0.0627***
SN_2-Var	0.0016	0.0089	0.0121	0.0190	0.0302	0.0503
Glasso	0.0017**	0.0091*	0.0125*	0.0198	0.0324*	0.0570***
Model 6	$a=0.05$	$a=0.1$	$a=0.15$	$a=0.2$	$a=0.25$	$a=0.3$
SN_1	0.0045***	0.0031***	0.0028***	0.0023***	0.0024***	0.0015***
SN_∞	0.0040*	0.0027***	0.0026**	0.0019	0.0020***	0.0013***
SN_2-L	0.0042**	0.0029***	0.0026***	0.0021***	0.0022***	0.0014***
SN_2-Var	0.0038	0.0025	0.0024	0.0019	0.0018	0.0012
Glasso	0.0040*	0.0026*	0.0026***	0.0020***	0.0020***	0.0013***

These comparisons demonstrate that the out-of-sample performance of Glasso critically depends on the sparsity of the true inverse covariance matrix. When there is no or little sparsity, Glasso performs worse than the proposed portfolio. Only when the true inverse covariance matrix becomes sparser may Glasso perform better. However, how likely the difference is to occur depends on the degree of structural misspecification of the inverse covariance matrix. In reality, the true covariance matrix and its inverse are rarely known. The proposed portfolio with a structure-free estimator, capable of providing a robust out-of-sample performance, would be intuitively appealing.

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

VII. Conclusion

To cope with the problem of overdispersion of sample eigenvalues, we propose a general framework to estimate the inverse covariance matrix based on maximum likelihood with a constraint on the Schatten q-norm. The proposed estimator is structure-free and also computationally efficient. The optimal shrinkage density of the sample eigenvalues can be analytically determined for a given regularization parameter. This property substantially reduces the computational load involved in portfolio risk optimization, making it especially attractive when the underlying portfolio has a large number of assets. The empirical comparisons show that SN_2-Var based on the Schatten 2-norm of the sample eigenvalues turns out to be effective in suppressing fluctuations of the sample eigenvalues. SN_2-Var behaves reasonably well in terms of reducing portfolio risk or turnover and increasing out-of-sample performance as compared with existing portfolio strategies.

Our proposed estimator differs from the Glasso estimator in two aspects, even though both schemes rest on shrinking the inverse covariance matrix directly. First, the proposed estimator is a structure-free estimator, whereas the Glasso estimator is a structure-based estimator. Second, the proposed estimator improves its stability by shrinking the sample eigenvalues, whereas the Glasso estimator achieves this by imposing the sparsity structure. Because the Glasso estimator has no control of the condition number, it is not always well conditioned, especially when some underlying assets are strongly correlated within high-dimensional portfolios. As corroborated by our empirical and simulation studies, these two different properties enable our proposed portfolio to achieve smaller out-of-sample variances than Glasso in the cases with no or little sparsity in the inverse covariance matrix and with highly correlated assets in high-dimensional portfolios (i.e., high condition number). However, when the inverse covariance matrix becomes sparser, Glasso works better than the proposed method.

Directions for future research include taking into account dependency across time, such as autoregressive conditional heteroscedasticity (ARCH)/generalized autoregressive conditional heteroscedasticity (GARCH) effects, and a detailed investigation of the Schatten norm constraint under other frameworks apart from maximum likelihood. The analytical solution to the optimal shrinkage density of the sample eigenvalues is derived based on maximum likelihood from a multivariate normal model. In practice, the asset returns may not follow a multivariate normal distribution. It is worth investigating the proposed method under nonnormal distributions or unknown distributions.

Appendix A. Derivation of the Regularized Eigenvalue Estimates When $q = \infty$

When $q = \infty$, the Schatten q-norm reduces to $m_{\max}(\Omega) \leq \delta$. The eigenvalue-regularized estimator can be obtained by solving the following optimization problem:

$$\begin{array}{ll} \min_{\Omega} & \operatorname{tr}(\Omega S) - \log|\Omega| \\ \text{s.t.} & m_{\max}(\Omega) \quad \leq \quad \delta. \end{array} \tag{A-1}$$

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

The spectral decomposition of the sample covariance matrix is $S = QLQ'$, where $L = \text{diag}(l_1, \ldots, l_p)$. The inverse covariance Ω has the spectral decomposition UMU', where $M = \text{diag}(m_1, \ldots, m_p)$ and $m_1 \geq \ldots \geq m_p$. It follows that

$$\text{(A-2)} \qquad \begin{aligned} \text{tr}(\Omega S) - \log|\Omega| &= \text{tr}(UMU'QLQ') - \log|\Omega| \\ &\geq \text{tr}(ML) - \log|M|, \end{aligned}$$

where equality holds when $U = Q$ (Farrell 1985, chapter 14). Therefore, the problem in equation (A-1) can be rewritten as

$$\text{(A-3)} \qquad \begin{aligned} &\min_{m_1,\ldots,m_p} \sum_{i=1}^{p} (l_i m_i - \log(m_i)) \\ &\text{s.t.} \quad 0 \leq m_p \leq \cdots \leq m_1 \leq \delta. \end{aligned}$$

Ignoring the order constraint among the eigenvalues, the problem in equation (A-3) becomes separable in $m_1, \ldots, m_p$. For each summand of the objective function in the problem in equation (A-3) without the order constraint, the regularized eigenvalue estimate is given by

$$\text{(A-4)} \qquad \hat{m}_i = \arg\min_{m_i \leq \delta} (l_i m_i - \log(m_i)) = \min\left\{\frac{1}{l_i}, \delta\right\}.$$

Note that the minimizer in equation (A-4) satisfies the order constraint.

Appendix B. Proofs of Proposition 3 and Theorem 3

1. Proof of Proposition 3

The estimate of the inverse covariance matrix is given by

$$\text{(B-1)} \qquad \hat{\Omega}_{\text{SN}} = Q\hat{M}Q' = \sum_{i=1}^{p} \hat{m}_i Q_i Q_i',$$

where Q_i is the ith column of matrix Q. Let w_h and w_k be the optimal weights on assets h and k, respectively. The ratio of w_h to w_k is given by

$$\text{(B-2)} \qquad \frac{w_h}{w_k} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} \hat{m}_i Q_{h,i} Q_{j,i}}{\sum_{i=1}^{p}\sum_{j=1}^{p} \hat{m}_i Q_{k,i} Q_{j,i}}.$$

For $q = 1$,

$$\text{(B-3)} \qquad \lim_{\rho\to\infty} \frac{\hat{m}_i}{\hat{m}_s} = \lim_{\rho\to\infty} \frac{l_s + \rho}{l_i + \rho} = 1.$$

For $q = 2$,

$$\text{(B-4)} \qquad \lim_{\rho\to\infty} \frac{\hat{m}_i}{\hat{m}_s} = \lim_{\rho\to\infty} \frac{\sqrt{l_s^2 + 8\rho} + l_s}{\sqrt{l_i^2 + 8\rho} + l_i} = 1.$$

For $q = \infty$,

$$\text{(B-5)} \qquad \lim_{\delta\to 1/l_{\max}} \frac{\hat{m}_i}{\hat{m}_s} = \lim_{\delta\to 1/l_{\max}} \frac{\min(1/l_i, \delta)}{\min(1/l_s, \delta)} = \frac{l_{\max}}{l_{\max}} = 1.$$

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

https://doi.org/10.1017/S0021090919000899 Published online by Cambridge University Press

Then

$$\lim_{\rho\to\infty}\frac{w_h}{w_k} = \lim_{\rho\to\infty}\frac{\sum_{i=1}^{p}\sum_{j=1}^{p}\hat{m}_i Q_{h,i} Q_{j,i}}{\sum_{i=1}^{p}\sum_{j=1}^{p}\hat{m}_i Q_{k,i} Q_{j,i}} \tag{B-6}$$

$$= \lim_{\rho\to\infty}\frac{\sum_{i=1}^{p}\sum_{j=1}^{p}\hat{m}_i/\hat{m}_s Q_{h,i} Q_{j,i}}{\sum_{i=1}^{p}\sum_{j=1}^{p}\hat{m}_i/\hat{m}_s Q_{k,i} Q_{j,i}}$$

$$= \frac{\sum_{i=1}^{p}\sum_{j=1}^{p}\lim_{\rho\to\infty}\hat{m}_i/\hat{m}_s Q_{h,i} Q_{j,i}}{\sum_{i=1}^{p}\sum_{j=1}^{p}\lim_{\rho\to\infty}\hat{m}_i/\hat{m}_s Q_{k,i} Q_{j,i}}$$

$$= \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} Q_{h,i} Q_{j,i}}{\sum_{i=1}^{p}\sum_{j=1}^{p} Q_{k,i} Q_{j,i}}.$$

Note that $QQ' = I$ and $\sum_{i=1}^{p}\sum_{j=1}^{p} Q_{h,i} Q_{j,i} = \sum_{i=1}^{p}\sum_{j=1}^{p} Q_{k,i} Q_{j,i} = 1$. It follows that

$$\lim_{\rho\to\infty}\frac{w_h}{w_k} = 1. \tag{B-7}$$

Similarly, we can prove that

$$\lim_{\delta\to 1/l_{\max}}\frac{w_h}{w_k} = 1. \tag{B-8}$$

2. Proof of Theorem 3

Let $S^{(k)} = Q^{(k)} L^{(k)} (Q^{(k)})'$ and $\hat{\Omega}^{(-k)} = Q^{(-k)} \hat{M}^{(-k)} (Q^{(-k)})'$, where $Q^{(-k)}$ and $\hat{M}^{(-k)}$ denote the sample eigenvectors and the diagonal matrix whose entries are the regularized eigenvalues obtained from the deletion of the kth group of observations in a sample, respectively. The K-fold cross-validated negative log-likelihood can be written as

$$\sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)}) = \sum_{k=1}^{K}\left[\operatorname{tr}(\hat{\Omega}^{(-k)} S^{(k)}) - \log|\hat{\Omega}^{(-k)}|\right] \tag{B-9}$$

$$= \sum_{k=1}^{K}\left[\operatorname{tr}(Q^{(-k)}\hat{M}^{(-k)}(Q^{(-k)})' S^{(k)}) - \log|Q^{(-k)}\hat{M}_{-k}(Q^{(-k)})'|\right]$$

$$= \sum_{k=1}^{K}\left[\operatorname{tr}(\hat{M}^{(-k)}(Q^{(-k)})' S^{(k)} Q^{(-k)}) - \sum_{i=1}^{p}\log(\hat{m}_i^{(-k)})\right],$$

where $\hat{m}_i^{(-k)}$ is the ith diagonal element of $\hat{M}^{(-k)}$. As $n_k \to \infty$, we have $S^{(k)} \xrightarrow{p} \Sigma$, $Q^{(-k)} \xrightarrow{p} U$, $Q^{(k)} \xrightarrow{p} U$, $L \xrightarrow{p} \Lambda$, $(Q^{(-k)})' S^{(k)} Q^{(-k)} \xrightarrow{p} \Lambda$, and $\hat{M}^{(-k)} \xrightarrow{p} \hat{M}$, where $\hat{M}$ is the diagonal matrix

whose entries are the modified eigenvalues. It follows that

$$\text{(B-10)} \qquad \begin{aligned} \sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)}) &\xrightarrow{p} K\left[\text{tr}(\hat{M}\Lambda) - \sum_{i=1}^{p}\log(\hat{m}_i)\right] \\ &= K\left[\sum_{i=1}^{p}(\hat{m}_i\lambda_i - \log(\hat{m}_i))\right]. \end{aligned}$$

For $q=1$, $\hat{m}_i = 1/(\lambda_i+\rho)$. Differentiating $\sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)})$ with respect to ρ leads to

$$\text{(B-11)} \qquad \begin{aligned} \frac{\partial\left(\sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)})\right)}{\partial\rho} &\xrightarrow{p} K\sum_{i=1}^{p}\frac{\partial\left(\frac{\lambda_i}{\lambda_i+\rho} - \log\left(\frac{1}{\lambda_i+\rho}\right)\right)}{\partial\rho} \\ &= K\sum_{i=1}^{p}\frac{\rho}{(\lambda_i+\rho)^2}, \end{aligned}$$

which is positive for $\rho>0$ and $K>0$. If Σ has a full rank, then $\lambda_i>0\ \forall i$. Therefore, the negative log-likelihood function in equation (B-10) has a unique minimum at $\rho^*=0$ when $q=1$.

For $q=2$, $\hat{m}_i = 2/(\sqrt{\lambda_i^2+8\rho}+\lambda_i)$. Differentiating $\sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)})$ with respect to ρ leads to

$$\text{(B-12)} \qquad \begin{aligned} &\frac{\partial\left(\sum_{k=1}^{K} -\log L(\hat{\Omega}^{(-k)}, S^{(k)})\right)}{\partial\rho} \\ &\xrightarrow{p} K\sum_{i=1}^{p}\frac{\partial\left(\frac{2\lambda_i}{\sqrt{\lambda_i^2+8\rho}+\lambda_i} - \log\left(\frac{2}{\sqrt{\lambda_i^2+8\rho}+\lambda_i}\right)\right)}{\partial\rho} \\ &= K\sum_{i=1}^{p}\frac{4\left(\sqrt{\lambda_i^2+8\rho}-\lambda_i\right)}{\sqrt{\lambda_i^2+8\rho}\left(\sqrt{\lambda_i^2+8\rho}+\lambda_i\right)^2}, \end{aligned}$$

which is positive for $\rho>0$ and $K>0$. Therefore, the negative log-likelihood function in equation (B-10) has a unique minimum at $\rho^*=0$.

For $q=\infty$, $\hat{m}_i = \min(1/\lambda_i, \delta)$. The negative log-likelihood function in equation (B-10) has a minimum when $\hat{m}_i = 1/\lambda_i$, implying that $\delta \geq 1/\lambda_{\min}$.

References

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

Best, M. J., and R. 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.

Cai, T. T.; W. Liu; and Y. Xia. "Two-Sample Test of High Dimensional Means under Dependence." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76 (2014), 349–372.

Candelon, B.; C. Hurlin; and S. Tokpavi. "Sampling Error and Double Shrinkage Estimation of Minimum Variance Portfolios." *Journal of Empirical Finance*, 19 (2012), 511–527.

Daniels, M. J., and R. E. Kass. "Shrinkage Estimators for Covariance Matrices." *Biometrics*, 57 (2001), 1173–1184.

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press

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

Farrell, R. H. "Multivariate Inequalities." In *Multivariate Calculation*. New York, NY: Springer (1985).

Gillen, B. J. "An Empirical Bayesian Approach to Stein-Optimal Covariance Matrix Estimation." *Journal of Empirical Finance*, 29 (2014), 402–420.

Goto, S., and Y. Xu. "Improving Mean Variance Optimization through Sparse Hedging Restrictions." *Journal of Financial and Quantitative Analysis*, 50 (2015), 1415–1441.

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. "Estimation for Markowitz Efficient Portfolios." *Journal of the American Statistical Association*, 75 (1980), 544–554.

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

Johnstone, I. M., and A. Y. Lu. "On Consistency and Sparsity for Principal Components Analysis in High Dimensions." *Journal of the American Statistical Association*, 104 (2009), 682–693.

Jorion, P. "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis*, 21 (1986), 279–292.

Kan, R., and G. Zhou. "Optimal Portfolio Choice with Parameter Uncertainty." *Journal of Financial and Quantitative Analysis*, 42 (2007), 621–656.

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." *J. Multivariate Analysis*, 88 (2004), 365–411.

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

Ledoit, O., and M. Wolf. "Robust Performances Hypothesis Testing with the Variance." Working Paper No. 516, University of Zurich Institute for Empirical Research in Economics (2011).

Levy, H., and M. Levy. "The Benefits of Differential Variance-Based Constraints in Portfolio Optimization." *European Journal of Operational Research*, 234 (2014), 372–381.

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

Markowitz, H. "Portfolio Selection." *Journal of Finance*, 7 (1952), 77–91.

Scherer, B., and R. D. Martin. *Modern Portfolio Optimization with NuOPT™, S-PLUS®, and S+ Bayes™*. New York, NY: Springer Science & Business Media (2005).

Stein, C. "Estimation of a Covariance Matrix." Rietz Lecture, 39th Annual IMS Meeting, Atlanta, GA, (1975).

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

Warton, D. I. "Penalized Normal Likelihood and Ridge Regularization of Correlation and Covariance Matrices." *Journal of the American Statistical Association*, 103 (2008), 340–349.

Witten, D. M., 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.

Won, J.-H.; J. Lim; S.-J. Kim; and B. Rajaratnam. "Condition-Number-Regularized Covariance Estimation." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75 (2013), 427–450.

Xing, X.; J. Hu; and Y. Yang. "Robust Minimum Variance Portfolio with L-Infinity Constraints." *Journal of Banking and Finance*, 46 (2014), 107–117.

https://doi.org/10.1017/S0022109019000899 Published online by Cambridge University Press