Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks

Olivier Ledoit
University of Zurich, AlphaCrest Capital Management

Michael Wolf
University of Zurich

Markowitz (1952) portfolio selection requires an estimator of the covariance matrix of returns. To address this problem, we promote a nonlinear shrinkage estimator that is more flexible than previous linear shrinkage estimators and has just the right number of free parameters (i.e., the *Goldilocks* principle). This number is the same as the number of assets. Our nonlinear shrinkage estimator is asymptotically optimal for portfolio selection when the number of assets is of the same magnitude as the sample size. In backtests with historical stock return data, it performs better than previous proposals and, in particular, it dominates linear shrinkage. (*JEL* C13, C58, G11)

Received January 21, 2014; editorial decision January 25, 2017 by Editor Geert Bekaert.

Markowitz's portfolio selection requires estimates of the vector of expected returns and the covariance matrix of returns. Green et al. (2013) list over 300 papers that have been written on the first estimation problem. By comparison, much less has been written about the covariance matrix. The one thing we do know is that the textbook estimator, the sample covariance matrix, is inappropriate. This is a simple degrees-of-freedom argument. The number of degrees of freedom in the sample covariance matrix is of order N^2, where N is the number of investable assets. In finance, the sample size T can be of the

---

We thank Editor Geert Bekaert and three anonymous referees for helpful comments that greatly enhanced the value of the paper. We also thank seminar participants at Aarhus University, Nanyang Technological University, Singapore Management University, Tilburg University, The University of Mannheim, and UCLA and conference participants at the Second NUS-Stanford Workshop in Quantitative Finance (Singapore), the Workshop on Financial Econometrics: Challenges and Directions for Future Research (Rio de Janeiro), the Ninth Annual SoFiE Conference (Hong Kong), and the Third Vietnam International Conference in Finance (Da Nang) for additional comments. Finally, we thank Stefan Bruder for careful proof-reading and editing. All remaining errors are ours. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Michael Wolf, Department of Economics, University of Zurich, Zürichbergstrasse 14, 8032 Zurich, Switzerland; telephone: +41 44 634 5096. E-mail: michael.wolf@econ.uzh.ch.

doi:10.1093/rfs/hhx052 Advance Access publication June 8, 2017

same order of magnitude as N.[1] Then the number of points in the historical data base is also of order N^2. We cannot possibly estimate $O(N^2)$ free parameters from a data set of order N^2. The number of degrees of freedom has to be an order of magnitude smaller than N^2, or else portfolio selection inevitably turns into what Michaud (1989) calls "error maximization".

Recent proposals by Ledoit and Wolf (2003, 2004a,b), Kan and Zhou (2007), Brandt et al. (2009), DeMiguel et al. (2009, 2013), Frahm and Memmel (2010), and Tu and Zhou (2011), among others, show that this topic is currently gathering a significant amount of attention. All these articles resolve the problem by going from $O(N^2)$ degrees of freedom to $O(1)$ degrees of freedom. They look for estimators of the covariance matrix, its inverse, or the portfolio weights that are optimal in a space of dimension one, two, or three. For example, the linear shrinkage approach of Ledoit and Wolf (2004b) finds a covariance matrix estimator that is optimal in the one-dimensional space of convex linear combinations of the sample covariance matrix with the (properly scaled) identity matrix. Another important class of models with $O(1)$ degrees of freedom, which has a long tradition in finance, is the class of factor models; for example, see Bekaert et al. (2009) and the references therein.

Given a data set of size $O(N^2)$, estimating $O(1)$ parameters is easy. The point of the present paper is that we can push this frontier. From a data set of size $O(N^2)$, we should be able, using sufficiently advanced technology, to estimate $O(N)$ free parameters consistently instead of merely $O(1)$. The sample covariance matrix with its $O(N^2)$ degrees of freedom is too loose, but the existing literature with only $O(1)$ degrees of freedom is too tight. $O(N)$ degrees of freedom is "just right" for a data set of size $O(N^2)$: It is the *Goldilocks* order of magnitude.[2]

The class of estimators we consider was introduced by Stein (1975, 1986) and is called *nonlinear shrinkage*. This means that the small eigenvalues of the sample covariance matrix are pushed up and the large ones pulled down by an amount that is determined *individually* for each eigenvalue. Since there are N eigenvalues, this gives N degrees of freedom, as required. The challenge is to determine the optimal shrinkage intensity for each eigenvalue. It cannot be optimal in the general sense of the word: It can only be optimal with respect to a particular loss function. We propose to use a loss function that captures the objective of an investor or researcher using portfolio selection and has been previously considered by Engle and Colacito (2006, Equation (3)) and Kan and Smith (2008). Our first theoretical contribution is to prove that this loss function has a well-defined limit under large-dimensional asympotics, i.e.,

[1] Such is the case when one invests in single stocks. There are other settings, such as strategic asset allocation, in which the dimension of the universe is much smaller.

[2] The Goldilocks principle refers to the classic fairy tale *The Three Bears*, where young Goldilocks finds a bed that is neither too soft nor too hard but "just right". In economics, this term describes a monetary policy that is neither too accomodative nor too restrictive but "just right".

when the dimension N goes to infinity along with the sample size T, and to compute its limit in closed form. Our second theoretical contribution is to characterize the nonlinear shrinkage formula that minimizes the limit of the loss. This original work results in an estimator of the covariance matrix that is asymptotically optimal for portfolio selection in the N-dimensional class of nonlinear shrinkage estimators when the number of investable assets, N, is large. We also prove uniqueness in the sense that all optimal estimators are asymptotically equivalent to one another, up to multiplication by a positive scalar.

To put this result in perspective, the statistics literature has only obtained nonlinear shrinkage formulas that are optimal with respect to some generic loss functions renowned for their tractability; see Ledoit and Wolf (2017). Note that all the work must be done anew for every loss function. Thus, our analytical results bridge the gap between theory and practice. Interestingly, a technology called *beamforming* (Capon, 1969) is essential for radars, wireless communication and other areas of signal processing, and is mathematically equivalent to portfolio selection. This equivalence implies that our covariance matrix estimator is also optimal for beamforming. The same is true for *fingerprinting*, the technique favored by the Intergovernmental Panel on Climate Change (IPCC, 2001, 2007) to measure the change of temperature on Earth. Thus, the applicability of our optimality result reaches beyond finance.

One caveat is that the optimal estimator we obtain is only an oracle, meaning that it depends on a certain unobservable quantity, which happens to be a complex-valued function tied to the limiting distribution of sample eigenvalues. The only way to make the nonlinear shrinkage approach usable in practice (i.e., bona fide) is to find a consistent estimator of this unobservable function. Fortunately this problem has been solved before (Ledoit and Wolf, 2012, 2015), so the transition from oracle to bona fide is completely straightforward in our case and requires no extra work.

Our optimal nonlinear shrinkage estimator dominates its competitors on historical stock returns data. For $N = 100$ assets, we obtain a global minimum-variance portfolio with 10.99% annualized standard deviation, versus 13.11% for the usual estimator, the sample covariance matrix. The amount of improvement is more pronounced in large dimensions. For example, for a universe comparable to the S&P 500, our global minimum-variance portfolio has an almost 50% lower out-of-sample volatility than the $1/N$ portfolio promoted by DeMiguel et al. (2009). We improve over the linear shrinkage estimator of Ledoit and Wolf (2004b) across the board. Having $O(N)$ free parameters chosen optimally confers a decisive advantage over having only $O(1)$ free parameters. We also demonstrate superior out-of-sample performance for portfolio strategies that target a certain exposure to an exogenously specified proxy for the vector of expected returns (also called a signal). This has implications for research on market efficiency, as it improves the power of a test for the ability of a candidate characteristic to explain the cross-section of stock returns.[3]

1. Loss Function for Portfolio Selection

The number of assets in the investable universe is denoted by N. Let m denote an $N \times 1$ cross-sectional signal or combination of signals that proxies for the vector of expected returns. Subrahmanyam (2010) documents at least 50 such signals. Hou et al. (2015) bring the tally up to 80 signals, McLean and Pontiff (2016) to 97 (listing them in their online Appendix) and Green et al. (2013) to 333 signals (extended bibliography available on request). Further overviews are provided by Ilmanen (2011) and Harvey et al. (2016).

It is not a goal of our paper to contribute to this strand of literature, i.e., to come up with an improved cross-sectional signal m. Our focus is only on the estimation of the covariance matrix.

1.1 Out-of-sample variance

The goal of researchers and investors alike is to put together a portfolio strategy that loads on the vector m, however decided upon. Let Σ denote the $N \times N$ population covariance matrix of asset returns; note that Σ is unobservable. Portfolio selection seeks to maximize the reward-to-risk ratio

$$\max_{w} \frac{w'm}{\sqrt{w'\Sigma w}} , \tag{1}$$

where w denotes an $N \times 1$ vector of portfolio weights. This optimization problem abstracts from leverage and short-sales constraints in order to focus on the core of Markowitz (1952) portfolio selection: the trade-off between reward and risk. A vector w is a solution to (1) if and only if there exists a strictly positive scalar a such that $w = a \times \Sigma^{-1} m$. This claim can be easily verified from the first-order condition of (1). The scale of the vector of portfolio weights can be set by targeting a certain level of expected returns, say b, in which case we get

$$w = \frac{b}{m'\Sigma^{-1}m} \times \Sigma^{-1} m . \tag{2}$$

In general, the weights will not sum up to one. Thus, $1 - \sum_{i=1}^{N} w_i$ is the share in the risk-free asset; the same reasoning can be found in Engle and Colacito (2006, Section 2).

Note that expression (2) is not scale-invariant with respect to b and m: if we double b, the portfolio weights double; and if we replace m by $2m$, the portfolio weights are halved. Scale dependence can be eliminated simply by setting $b := \sqrt{m'm}$. In practice, the covariance matrix Σ is not known and needs to be

[3] For example, see Bell et al. (2014) for such a test.

estimated from historical data. Let $\widehat{\Sigma}$ denote a generic (invertible) estimator of the covariance matrix. The plug-in estimator of the optimal portfolio weights is

$$\widehat{w} := \frac{\sqrt{m'm}}{m'\widehat{\Sigma}^{-1}m} \times \widehat{\Sigma}^{-1}m \,. \tag{3}$$

All investing takes place *out of sample* by necessity. Since the population covariance matrix Σ is unknown and the covariance matrix estimator $\widehat{\Sigma}$ is not equal to it, out-of-sample performance is different from in-sample performance. We want the portfolio with the best possible behavior out of sample. This is why we define the loss function for portfolio selection as the out-of-sample variance of portfolio returns conditional on $\widehat{\Sigma}$ and m.

Definition 1. The loss function is

$$\mathcal{L}\left(\widehat{\Sigma}, \Sigma, m\right) := \widehat{w}'\Sigma\widehat{w} = m'm \times \frac{m'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}m}{\left(m'\widehat{\Sigma}^{-1}m\right)^2} \,. \tag{4}$$

In terms of assessing the broad scientific usefulness of this line of research, it is worth pointing out that the loss function in Definition 1 also handles the problems known as optimal beamforming in signal processing and optimal fingerprinting in climate change research, because they are mathematically equivalent to optimal portfolio selection, as evidenced in Du et al. (2010) and Ribes et al. (2009), respectively.

Remark 1. The usual approach would be to minimize the risk function which is defined as the expectation of the loss function (4). However, in our asymptotic framework, the loss function converges almost surely to a nonstochastic limit, as we will show in Theorem 1. Therefore, there is no need to take the expectation. □

1.2 Out-of-sample Sharpe ratio

An alternative objective of interest to financial investors is the Sharpe ratio. The vector m represents the investor's best proxy for the vector of expected returns given the information available, so we use m to evaluate the numerator of the Sharpe ratio. From the weights in Equation (3), we deduce the Sharpe ratio as

$$\begin{aligned}
\frac{\widehat{w}'m}{\sqrt{\widehat{w}'\Sigma\widehat{w}}} &= \frac{\sqrt{m'm}}{m'\widehat{\Sigma}^{-1}m} \times m'\widehat{\Sigma}^{-1}m \times \frac{m'\widehat{\Sigma}^{-1}m}{\sqrt{m'm}} \times \frac{1}{\sqrt{m'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}m}} \\
&= \frac{m'\widehat{\Sigma}^{-1}m}{\sqrt{m'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}m}} \\
&= \sqrt{\frac{m'm}{\mathcal{L}\left(\widehat{\Sigma}, \Sigma, m\right)}} \,.
\end{aligned}$$

Since we do not optimize over the Euclidian norm of the vector m, which is a given, the objective of maximizing the Sharpe ratio is therefore equivalent to minimizing the loss function of Definition 1.

Note that this equivalence only holds *conditional* on the expected returns proxy m. As indicated at the beginning of Section 2, we do not delve into the issue of how to choose m.

1.3 Quadratic objective function

Yet another way to control the trade-off between risk and return is to maximize the expectation of a quadratic utility function of the type $W-\gamma W^2$, where W represents the final wealth and γ is a risk aversion parameter. As above, we use the cross-sectional return predictive signal m in lieu of the unavailable first moment because it is the investor's proxy of future expected returns conditional on her information set. Plugging in the weights of Equation (3) yields

$$
\begin{aligned}
&W_0+W_0\widehat{w}'m-\gamma\left[\left(W_0+W_0\widehat{w}'m\right)^2+W_0^2\widehat{w}'\Sigma\widehat{w}\right]\\
&\quad=W_0+W_0\frac{\sqrt{m'm}}{m'\Sigma^{-1}m}\times m'\Sigma^{-1}m\\
&\qquad-\gamma W_0^2\left[\left(1+\frac{\sqrt{m'm}}{m'\Sigma^{-1}m}\times m'\Sigma^{-1}m\right)^2+\frac{m'm}{\left(m'\Sigma^{-1}m\right)^2}\times m'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}m\right]\\
&\quad=W_0+W_0\sqrt{m'm}-\gamma W_0^2\left[1+2\sqrt{m'm}+m'm+\mathcal{L}\left(\widehat{\Sigma},\Sigma,m\right)\right],
\end{aligned}
$$

where W_0 stands for the agent's initial wealth. Since we do not optimize over the Euclidian norm of the vector m, which is a given, and we do not optimize over the initial wealth W_0 either, this objective is also equivalent to minimizing the loss function of Definition 1. This equivalence is further confirmation that the loss function $\mathcal{L}\left(\widehat{\Sigma},\Sigma,m\right)$ is the right quantity to look at in the context of portfolio selection.

2. Large-Dimensional Asymptotic Limit of the Loss Function

The framework defined in Assumptions 1–4 below is standard in the literature on covariance matrix estimation under large-dimensional asymptotics; see Bai and Silverstein (2010) for an authoritative and comprehensive monograph on this subject. These assumptions have to be formulated explicitly here for completeness' sake, and as they may not be so familiar to finance audiences we have interspersed additional explanations whenever warranted. The remainder of the section from Remark 3 onwards focuses on finance-related issues. Some of the assumptions made in Section 1 are restated below in a manner more suitable for the large-dimensional framework, and the subscript T will be affixed to the quantities that require it.

Assumption 1 (Dimensionality). Let T denote the sample size and $N := N(T)$ the number of variables. It is assumed that the ratio N/T converges, as $T \to \infty$, to a limit $c \in (0,1) \cup (1,+\infty)$ called the *concentration (ratio)*. Furthermore, there exists a compact interval included in $(0,1) \cup (1,+\infty)$ that contains N/T for all T large enough.

Quantities introduced in Section 1 will henceforth be indexed by the subscript T so that we can study their asymptotic behavior. Unlike the proposals by Kan and Zhou (2007), Frahm and Memmel (2010), and Tu and Zhou (2011), our method can also handle the case $c > 1$, where the sample covariance matrix is not invertible. The case $c = 1$ is ruled out in the theoretical treatment for technical reasons, but the empirical results in Section 4 indicate that our method works well in practice even in this challenging case.

Assumption 2 (Population Covariance Matrix).

(a) The population covariance matrix Σ_T is a nonrandom symmetric positive-definite matrix of dimension N.
(b) Let $\boldsymbol{\tau}_T := (\tau_{T,1}, \ldots, \tau_{T,N})'$ denote a system of eigenvalues of Σ_T sorted in increasing order. The empirical distribution function (e.d.f.) of population eigenvalues is defined as: $\forall x \in \mathbb{R}$, $H_T(x) := N^{-1} \sum_{i=1}^{N} \mathbb{1}_{[\tau_{T,i},\infty)}(x)$, where $\mathbb{1}$ denotes the indicator function of a set. It is assumed that H_T converges weakly to a limit law H, called the *limiting spectral distribution (function)*.
(c) $\mathsf{Supp}(H)$, the support of H, is the union of a finite number of closed intervals, bounded away from zero and infinity.
(d) There exists a compact interval $[\underline{h}, \overline{h}] \subset (0,\infty)$ that contains $\mathsf{Supp}(H_T)$ for all T large enough.

The existence of a limiting population spectral distribution is a usual assumption in the literature on large-dimensional asymptotics, but given that it is relatively new in finance, it is worth providing additional explanations. In item (a) the population covariance matrix harbors the subscript T to signify that it depends on the sample size: it changes as T goes to infinity because its dimension N is a function of T, as stated in Assumption 1. Item (b) defines the cross-sectional distribution of population eigenvalues H_T as the nondecreasing function that returns the proportion of eigenvalues to the left of any given number. H_T converges to some limit H, which can be interpreted as the signature of the population covariance matrix: it says what proportion of eigenvalues are big, small, etc. Items (c) and (d) are technical assumptions requiring the supports of H and H_T to be well-behaved.

Assumption 3 (Data Generating Process). X_T is a $T \times N$ matrix of i.i.d. random variables with mean zero, variance one, and finite 12th moment.

The matrix of observations is $Y_T := X_T \times \sqrt{\Sigma_T}$, where $\sqrt{\Sigma_T}$ denotes the symmetric positive-definite square root of Σ_T. Neither $\sqrt{\Sigma_T}$ nor X_T are observed on their own: only Y_T is observed.

Remark 2. The matrix $\sqrt{\Sigma_T}$ is not obtained by the Cholesky decomposition of the population covariance matrix Σ_T, since then it would be lower triangular. Instead $\sqrt{\Sigma_T}$ is the symmetric positive-definite matrix obtained by keeping the same eigenvectors as Σ_T, but recombining them with the square roots of the population eigenvalues, namely $\left(\sqrt{\tau_{T,1}}, \ldots, \sqrt{\tau_{T,N}}\right)'$. $\square$

If the asset returns have nonzero means, as is usually the case, then it is possible to remove the sample means, and our results still go through because it only constitutes a rank-one perturbation for the large-dimensional matrices involved, as shown in Theorem 11.43 of Bai and Silverstein (2010). Whereas the bound on the 12th moment simplifies the mathematical proofs, numerical simulations (not reported here) indicate that a bounded fourth moment would be sufficient in practice.

The sample covariance matrix is defined as $S_T := T^{-1} Y_T' Y_T = T^{-1} \sqrt{\Sigma_T} X_T' X_T \sqrt{\Sigma_T}$. It admits a spectral decomposition $S_T = U_T \Lambda_T U_T'$, where Λ_T is a diagonal matrix, and U_T is an orthogonal matrix: $U_T U_T' = U_T' U_T = \mathbb{I}_T$, where $\mathbb{I}_T$ (in slight abuse of notation) denotes the identity matrix of dimension $N \times N$. Let $\Lambda_T := \mathsf{Diag}(\boldsymbol{\lambda}_T)$, where $\boldsymbol{\lambda}_T := (\lambda_{T,1}, \ldots, \lambda_{T,N})'$. We can assume without loss of generality that the sample eigenvalues are sorted in increasing order: $\lambda_{T,1} \leq \lambda_{T,2} \leq \cdots \leq \lambda_{T,N}$. Correspondingly, the ith sample eigenvector is $u_{T,i}$, the ith column vector of U_T. The e.d.f. of the sample eigenvalues is given by: $\forall x \in \mathbb{R}$, $F_T(x) := N^{-1} \sum_{i=1}^{N} \mathbb{1}_{[\lambda_{T,i}, \infty)}(x)$, where $\mathbb{1}$ denotes the indicator function of a set.

The literature on sample covariance matrix eigenvalues under large-dimensional asymptotics is based on a foundational result due to Marčenko and Pastur (1967). It has been strengthened and broadened by subsequent authors reviewed in Bai and Silverstein (2010). Under Assumptions 1–3, there exists a limiting sample spectral distribution F continuously differentiable on $(0, +\infty)$ such that

$$\forall x \in (0, +\infty) \qquad F_T(x) \xrightarrow{\text{a.s.}} F(x). \tag{5}$$

In addition, the existing literature has unearthed important information about the limiting spectral distribution F, including an equation that relates F to H and c. This means that, asymptotically, one knows the average number of sample eigenvalues that fall in any given interval. Another useful result concerns the support of the distribution of the sample eigenvalues. Theorem 6.3 of Bai and Silverstein (2010) and Assumptions 1–3 imply that the support of F, $\mathsf{Supp}(F)$, is the union of a finite number $\kappa \geq 1$ of compact intervals $\bigcup_{k=1}^{\kappa} [a_k, b_k]$, where $0 < a_1 < b_1 < \cdots < a_\kappa < b_\kappa < \infty$, with the addition of the singleton $\{0\}$ in the case $c > 1$.

Assumption 4 (Class of Estimators). We consider positive-definite covariance matrix estimators of the type $\widehat{\Sigma}_T := U_T \widehat{\Delta}_T U_T'$, where $\widehat{\Delta}_T$ is a diagonal matrix: $\widehat{\Delta}_T := \mathsf{Diag}\big(\widehat{\delta}_T(\lambda_{T,1})\ldots, \widehat{\delta}_T(\lambda_{T,N})\big)$, and $\widehat{\delta}_T$ is a real univariate function which can depend on S_T. We assume that there exists a nonrandom real univariate function $\widehat{\delta}$ defined on $\mathsf{Supp}(F)$ and continuously differentiable such that $\widehat{\delta}_T(x) \xrightarrow{\text{a.s}} \widehat{\delta}(x)$, for all $x \in \mathsf{Supp}(F)$. Furthermore, this convergence is uniform over $x \in \bigcup_{k=1}^{\kappa}[a_k+\eta, b_k-\eta]$, for any small $\eta > 0$. Finally, for any small $\eta > 0$, there exists a finite nonrandom constant $\widehat{K}$ such that almost surely, over the set $x \in \bigcup_{k=1}^{\kappa}[a_k-\eta, b_k+\eta]$, $\widehat{\delta}_T(x)$ is uniformly bounded by $\widehat{K}$ from above and by $1/\widehat{K}$ from below, for all T large enough. In the case $c > 1$ there is the additional constraint $1/\widehat{K} \leq \widehat{\delta}_T(0) \leq \widehat{K}$ for all T large enough.

This is the class of *rotation-equivariant* estimators introduced by Stein (1975, 1986): rotating the original variables results in the same rotation being applied to the covariance matrix estimator. Rotation equivariance is appropriate in the general case where the statistician has no a priori information about the orientation of the eigenvectors of the covariance matrix.

The financial interpretation of rotating the original variables is to repackage the N individual stocks listed on the exchange into an equal number N of funds that span the same space of attainable investment opportunities. The assumption of rotation equivariance simply means that the covariance matrix estimator computed from the N individual stocks must be consistent with the one computed from the N funds.

The fact that we keep the sample eigenvectors does not mean that we assume they are close to the population eigenvectors. It only means that we do not know how to improve upon them. If we believed that the sample eigenvectors were close to the population eigenvectors, then the optimal covariance matrix estimator would have eigenvalues very close to the population eigenvalues. As we will see below, this is not at all what we do, because it is not optimal. Our nonlinear shrinkage formula differs from the population eigenvalues precisely because it needs to minimize the impact of sample eigenvectors estimation error.

We call $\widehat{\delta}_T(\cdot)$ the *shrinkage function* (or at times the *shrinkage formula*) because, in all applications of interest, its effect is to shrink the set of sample eigenvalues by reducing its dispersion around the mean, pushing up the small ones and pulling down the large ones. Shrinkage functions need to be as well-behaved asymptotically as spectral distribution functions, except possibly on a finite number of arbitrarily small regions near the boundary of the support. The large-dimensional asymptotic properties of a generic rotation-equivariant estimator $\widehat{\Sigma}_T$ are fully characterized by its limiting shrinkage function $\widehat{\delta}(\cdot)$.

Remark 3. The linear shrinkage estimator of Ledoit and Wolf (2004b) also belongs to this class of rotation-equivariant estimators, with the shrinkage function given by

$$\widehat{\delta}_T(\lambda_{T,i}) := (1-\widehat{k}) \cdot \lambda_{T,i} + \widehat{k} \cdot \bar{\lambda}_T \quad \text{where} \quad \bar{\lambda}_T := \frac{1}{N} \sum_{j=1}^{N} \lambda_{T,j} . \tag{6}$$

Here, the shrinkage intensity $\widehat{k} \in [0,1]$ is determined in an asymptotically optimal fashion; see Ledoit and Wolf (2004b, Section 3.3). □

Remark 4. If rotation equivariance is lost, this means that our method can be improved further still by taking into account a priori information about the orientation of the underlying data structure. Although this line of research is not the main thrust of the paper, we describe how to implement such an extension in Section 4.2. □

Estimators in the class defined by Assumption 4 are evaluated according to the limit as T and N go to infinity together (in the manner of Assumption 1) of the loss function defined in Equation (4). For this limit to exist, some assumption on the return predictive signal is required.[4] The assumption that we make below is coherent with the rotation-equivariant framework we have built so far.

Assumption 5 (Return Predictive Signal). m_T is distributed independently of S_T and its distribution is rotation invariant.

Remark 5. Assumption 5 may be hard to uphold for models that link expected returns to covariances. However, our methodology does not take a stance on the origin of the return predictive signal. It is designed to work well over the widest range, as opposed to being custom-tailored to a specific signal. Hence features of specific signals, such as expected returns being linked to covariances, are not accomodated by our methodology. Presumably, the performance of our method can be further improved by taking such linkages into account, but then the optimal nonlinear shrinkage formula would have to be derived anew for every different model of expected returns. Doing so lies beyond the scope of the present paper, but constitutes a promising avenue for future research. □

Rotation invariance means that the normalized return predictive signal $m_T/\sqrt{m_T' m_T}$ is uniformly distributed on the unit sphere. This setting favors covariance matrix estimators that work well across the board, without indicating a preference about the orientation of the vector of expected returns. Furthermore, it implies that m_T is distributed independently of any $\widehat{\Sigma}_T$ that belongs to the rotation-equivariant class of Assumption 4. The limit of the

[4] The return predictive signal can be interpreted as an estimator of the vector of expected returns, which is not available in practice.

loss function defined in Section 1 is given by the following theorem, where $\mathbb{C}^+ := \{a + i \cdot b : a \in \mathbb{R}, b \in (0,\infty)\}$ denotes the strict upper half of the complex plane; here, $i := \sqrt{-1}$.

Theorem 1. Under Assumptions 1–5,

$$m_T' m_T \times \frac{m_T' \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} m_T}{\left(m_T' \widehat{\Sigma}_T^{-1} m_T\right)^2} \xrightarrow{\text{a.s.}} \frac{\displaystyle\sum_{k=1}^{\kappa} \int_{a_k}^{b_k} \frac{dF(x)}{x\,|s(x)|^2 \widehat{\delta}(x)^2} + \mathbb{1}_{\{c>1\}} \frac{1}{c s(0) \widehat{\delta}(0)^2}}{\left[\displaystyle\int \frac{dF(x)}{\widehat{\delta}(x)}\right]^2}, \tag{7}$$

where, for all $x \in (0,\infty)$, and also for $x = 0$ in the case $c > 1$, $s(x)$ is defined as the unique solution $s \in \mathbb{R} \cup \mathbb{C}^+$ to the equation

$$s = -\left[x - c \int \frac{\tau}{1 + \tau s} dH(\tau)\right]^{-1}. \tag{MP}$$

The important message of the theorem is that there is a nonstochastic limit of the loss function. This means that we do not have to take expectations, since all randomness vanishes in the large-dimensional limit. Then the line of attack will be to find the nonlinear shrinkage function that minimizes the limiting loss. The formulas themselves are not particularly intuitive, especially because much of the action takes place in the complex plane, but since they are relatively easy to compute and implement, this is not much of a handicap.

All proofs are in Appendix A. Although Equation (MP) may appear daunting at first sight, it comes from the original article by Marčenko and Pastur (1967) that spawned the literature on large-dimensional asymptotics. Broadly speaking, the complex-valued function $s(x)$ can be interpreted as the Stieltjes (1894) transform of the limiting empirical distribution of sample eigenvalues; see Appendix A.1 for more specific mathematical details. By contrast, Equation (7) is one of the major mathematical innovations of this paper.

3. Optimal Covariance Matrix Estimator for Portfolio Selection

An *oracle* estimator is one that depends on unobservable quantities. It constitutes an important stepping stone towards the formulation of a bona fide estimator, i.e., an estimator usable in practice, provided that the unobservable quantities can be estimated consistently.

3.1 Oracle estimator

Equation (7) enables us to characterize the optimal limiting shrinkage function in the following theorem, which represents the culmination of the new theoretical analysis developed in the present paper.

Theorem 2. Define the oracle shrinkage function

$$\forall x \in \mathsf{Supp}(F) \qquad d^*(x) := \begin{cases} \dfrac{1}{x\,|s(x)|^2} & \text{if } x > 0, \\ \dfrac{1}{(c-1)s(0)} & \text{if } c > 1 \text{ and } x = 0\,, \end{cases} \tag{8}$$

where $s(x)$ is the complex-valued Stieltjes transform defined by Equation (MP). If Assumptions 1–5 are satisfied, then the following statements hold true.

(a) The oracle estimator of the covariance matrix

$$S_T^* := U_T D_T^* U_T' \quad \text{where} \quad D_T^* := \mathsf{Diag}\big(d^*(\lambda_{T,1}), \ldots, d^*(\lambda_{T,N})\big) \tag{9}$$

minimizes in the class of rotation-equivariant estimators defined in Assumption 4 the almost sure limit of the portfolio-selection loss function introduced in Section 1,

$$\mathcal{L}_T\left(\widehat{\Sigma}_T, \Sigma_T, m_T\right) := m_T' m_T \frac{m_T' \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} m_T}{\left(m_T' \widehat{\Sigma}_T^{-1} m_T\right)^2}\,, \tag{10}$$

as T and N go to infinity together in the manner of Assumption 1.

(b) Conversely, any covariance matrix estimator $\widehat{\Sigma}_T$ that minimizes the a.s. limit of the portfolio-selection loss function $\mathcal{L}_T$ is asymptotically equivalent to S_T^* up to scaling, in the sense that its limiting shrinkage function is of the form $\widehat{\delta} = \alpha d^*$ for some positive constant α.

S_T^* is an oracle estimator because it depends on c and $s(x)$ which are both unobservable. Nonetheless, deriving a bona fide counterpart to S_T^* will be easy because consistent estimators for c and $s(x)$ are readily available, as demonstrated in Section 3.4 below; therefore, we shall keep our attention on S_T^* for the time being.

As is apparent from part (b) of Theorem 2, minimizing the loss function characterizes the optimal shrinkage formula only up to an arbitrary positive scaling factor α. This is inherent to the problem of portfolio selection: Equation (2) shows that two covariance matrices that only differ by the scaling factor α yield the same vector of portfolio weights. The point of the following proposition is to control the trace of the estimator in order to pick the most natural scaling coefficient.

Proposition 1. Let Assumptions 1–5 hold, and let the limiting shrinkage function of the estimator $\widehat{\Sigma}_T$ be of the form $\widehat{\delta}(\cdot) = \alpha d^*(\cdot)$ for some positive constant α as per item (b) of Theorem 2. Then,

$$\frac{1}{N}\mathsf{Tr}\big(\widehat{\Sigma}_T\big) - \frac{1}{N}\mathsf{Tr}\big(\Sigma_T\big) \xrightarrow{\text{a.s.}} 0 \tag{11}$$

if and only if $\alpha = 1$.

Since it is desirable for an estimator to have the same trace as the population covariance matrix, from now on we will focus exclusively on the scaling coefficient $\alpha = 1$ and the oracle estimator S_T^*.

The optimal oracle estimator S_T^* does not depend on the vector of return signals m_T. This is because it is designed to work well across the board for all m_T, as evidenced by Assumption 5. In subsequent Monte Carlo simulations (Section 4.3 and Appendix D), we will make specific choices for m_T (the unit vector and the momentum factor, respectively), but that is only for the purpose of illustration.

One of the basic features of the optimal nonlinear transformation $d^*(\cdot)$ is that it preserves the grand mean of the eigenvalues, as evidenced by Proposition 1. The natural follow-up question is whether the cross-sectional dispersion of eigenvalues about their grand mean expands or shrinks. The answer can be found by combining Theorem 1.4 of Ledoit and Péché (2011) with Section 2.3 of Ledoit and Wolf (2004b). The former provides a heuristic interpretation of $d^*(\lambda_{T,i})$ as an approximation to $u_{T,i}' \Sigma_T u_{T,i}$, whereas the latter shows that $\left(u_{T,i}' \Sigma_T u_{T,i}\right)_{i=1,\dots,N}$ are less dispersed than $\left(\lambda_{T,i}\right)_{i=1,\dots,N}$. Together they imply that the transformation d^* does indeed deserve to be called a "shrinkage" because it reduces cross-sectional dispersion.

Further information regarding Theorem 2 can be gathered by comparing Equation (8) with the two shrinkage formulas obtained earlier by Ledoit and Wolf (2012). These authors use a generic loss function renowned for its tractability, based on the *Frobenius* norm. The Frobenius norm of a quadratic matrix A is defined as $\|A\|_F := \sqrt{\mathsf{Tr}(AA')}$, so it is essentially quadratic in nature. Ledoit and Wolf (2012) use a Frobenius-norm-based loss function in two different ways, once with the covariance matrix and then again with its inverse:

$$\mathcal{L}_T^1\left(\widehat{\Sigma}_T, \Sigma_T\right) := \frac{1}{N} \left\| \widehat{\Sigma}_T - \Sigma_T \right\|_F^2 \quad \text{and} \tag{12}$$

$$\mathcal{L}_T^{-1}\left(\widehat{\Sigma}_T, \Sigma_T\right) := \frac{1}{N} \left\| \widehat{\Sigma}_T^{-1} - \Sigma_T^{-1} \right\|_F^2 , \tag{13}$$

leading to two different optimal shrinkage formulas.

The first unexpected result is that Equation (8) is the same as one of the two shrinkage formulas obtained by Ledoit and Wolf (2012), even though the loss functions are completely different. We consider this result to be reassuring because it is easier to trust a shrinkage formula that is optimal with respect to multiple loss functions than one which is intimately tied to just one particular loss function.

The second unexpected result is that, of the two shrinkage formulas of Ledoit and Wolf (2012), $d^*(\cdot)$ is equal to the "wrong" one. Indeed, Equation (2) makes it clear that Markowitz (1952) portfolio selection involves not the covariance matrix itself but its *inverse*. Thus we might have expected that $d^*(\cdot)$ is equal to the shrinkage formula obtained by minimizing the loss function $\mathcal{L}_T^{-1}$, which penalizes estimation error in the inverse covariance matrix. It turns out that the exact opposite is true: $d^*(\cdot)$ is optimal with respect to $\mathcal{L}_T^1$ instead. This insight could not have been anticipated without the analytical developments achieved in Theorems 1 and 2. In particular, there have been several papers recently concerned with the direct estimation of the inverse covariance matrix[5] using a loss function of the type (13), with Markowitz (1952) portfolio selection listed as a major motivation; for example, see Frahm and Memmel (2010), Bodnar et al. (2014), and Wang et al. (2015). But our result shows that, unexpectedly, this approach is suboptimal in the context of portfolio selection.

Remark 6. To the extent that some intuition can be gleaned, it goes as follows. A bit of linear algebra reveals that our loss function $\mathcal{L}\left(\widehat{\Sigma}_T, \Sigma_T, m_T\right)$ involves the diagonal of $U_T' \Sigma_T U_T$. This is the critical ingredient that will condition the shape of the final result. $\mathcal{L}_T^1\left(\widehat{\Sigma}_T, \Sigma_T\right)$ can also be rewritten in terms of the diagonal of $U_T' \Sigma_T U_T$. This is why they both end up with the same shrinkage formula. But $\mathcal{L}_T^{-1}\left(\widehat{\Sigma}_T, \Sigma_T\right)$ involves the diagonal of $U_T' \Sigma_T^{-1} U_T$, which is different from the diagonal of $U_T' \Sigma_T U_T$, resulting in a different shrinkage formula.

This is easy to say in hindsight, after having done the mathematical derivations in detail. Intuition alone can be misleading; there is no short cut to bypass the hard work of going through all the necessary calculations. □

3.2 Portfolio decomposition

In the end, the best way to gain comfort with this mathematical result may be to seek a portfolio-decomposition interpretation of it. Starting from Section 1.1, we can express the vector of optimal portfolio weights as

$$w_T^* := a_T \times (S_T^*)^{-1} m_T = a_T \times U_T (D_T^*)^{-1} U_T' m_T = a_T \times \sum_{i=1}^{N} \frac{u_{T,i}' m_T}{d^*(\lambda_{T,i})} u_{T,i}$$

$$\approx a_T \times \sum_{i=1}^{N} \frac{u_{T,i}' m_T}{u_{T,i}' \Sigma_T u_{T,i}} u_{T,i} \ ,$$

where a_T is a suitably chosen scalar coefficient, and the last approximation comes from Theorem 1.4 of Ledoit and Péché (2011), as mentioned earlier. Thus, the mean-variance efficient portfolio can be decomposed into a linear combination of sample eigenvector portfolios, with the ith sample eigenvector portfolio assigned a weight approximately proportional to $u_{T,i}' m_T / \left(u_{T,i}' \Sigma_T u_{T,i}\right)$. This weighting scheme is intuitively appealing because it represents the out-of-sample reward-to-risk ratio of the ith sample eigenvector portfolio. (By "out-of-sample", we mean the true risk, which is determined by the population covariance matrix Σ_T, and not any estimator of it.) Thus we can be confident that the proposed nonlinear shrinkage formula makes sense economically.

[5] The inverse covariance matrix is also called the *precision matrix* in the statistics literature.

3.3 Why shrinkage can be useful even for small N

The amount of bias in the sample eigenvalues induced by having a non-vanishing concentration ratio $c = N/T$ depends on the whole shape of the spectral distribution, and is not generally available in closed form. However a particular case in which a closed-form solution is known can serve for illustration purposes: it is when all population eigenvalues are equal to one another. The resulting cross-sectional distribution of the sample eigenvalues is known as the Marčenko-Pastur law. If all population eigenvalues are equal to, say, τ, then the support of the Marčenko-Pastur law is the interval $[\tau(1-\sqrt{c})^2, \tau(1+\sqrt{c})^2]$. Thus the maximum relative bias is of the order of $2\sqrt{c}$ for small c. Suppose that we are willing to tolerate a relative error in the allocation of our portfolio weights across sample eigenvectors of 5% maximum. This means that we need to have $2\sqrt{N/T} = 0.05$. For a small portfolio of $N = 30$ stocks, which corresponds to the number of constituents in a narrow-based index such as the Dow Jones, this requires 192 years of daily data already. Thus, for all practical purposes, even for small portfolios of only $N = 30$ stocks, using our shrinkage formula is not a luxury.

The problem of correcting the eigenvalues is highly nonlinear in nature, even for what people may consider to be fairly low values of the concentration ratio $c = N/T$. Let us say, for example, that we have five times more observations than the number of stocks. This corresponds, for example, to one year of daily data on a portfolio of $N = 50$ stocks, which is probably towards the lower end for a quantitative equity portfolio manager. At first sight, it would appear that $T = 5N$ is sufficient to escape from the singularity problems that arise when $T > N$. Yet, even when $c = 1/5$, a highly nonlinear correction is needed. This correction, of course, depends on the actual shape of the spectral distribution. For the sake of illustration, we consider a broad selection of distributions from the Beta family, linearly shifted so that the support is [1,10]; their densities are plotted in Figure 7 of Ledoit and Wolf (2012). The corresponding oracle shrinkage functions (8) are displayed in Figure 1.

Visually, one can verify that the nonlinearities are quite pronounced, even though N is small relative to T. Also bear in mind that all the c.d.f.s from the beta family are nicely continuous; the nonlinear effects would be even more striking if we had used discontinuous population spectral c.d.f.s as in Bai and Silverstein (1998), Johnstone (2001), or Mestre (2008).

The improvement we get by correctly shrinking the sample eigenvalues in a nonlinear fashion compensates for the fact that we do not seek to improve over the estimator of the mean vector. It may be possible to cumulate the improvements of the two strands of literature by combining our method for the estimation of the covariance matrix with some other method for the reduction in the estimation risk of the vector of expected returns. This topic is an interesting avenue for future research but it lies outside the scope of the present paper.

Given that, by a mathematical accident, we end up with the same shrinkage formula as Ledoit and Wolf (2012), we can recycle their Monte Carlo

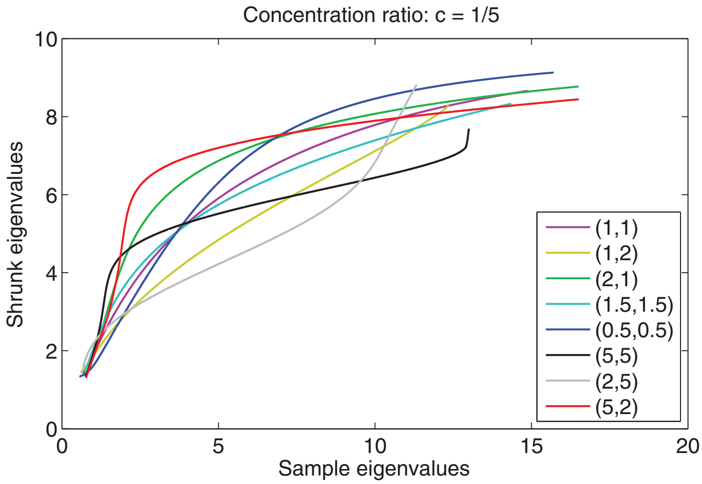

Figure 1
Oracle shrinkage functions (8), mapping sample eigenvalues to shrunk eigenvalues, for a variety of (shifted) Beta (α, β) distributions governing the population eigenvalues.

simulations. When the universe dimension N is ten times smaller than the sample size T, nonlinear shrinkage improve by up to 90% over the sample covariance matrix. Even when N is 100 times smaller than T, there is still up to 60% improvement.

3.4 Bona fide estimator

Transforming our optimal oracle estimator S_T^* into a bona fide one is a relatively straightforward affair thanks to the solutions provided by Ledoit and Wolf (2012, 2015). These authors develop an estimator $\widehat{s}(x)$ for the unobservable Stieltjes transform $s(x)$, and demonstrate that replacing $s(x)$ with $\widehat{s}(x)$ and replacing the limiting concentration ratio c with its natural estimator $\widehat{c}_T := N/T$ is done at no loss asymptotically. Given that the estimator $\widehat{s}(x)$ is not novel, a restatement of its definition for the sake of convenience is relegated to Appendix B. The following corollary gives the bona fide covariance matrix estimator that is optimal for portfolio selection in the N-dimensional class of nonlinear shrinkage estimators.

Corollary 1. Suppose that Assumptions 1–5 are satisfied, and let $\widehat{s}(x)$ denote the estimator of the Stieltjes transform $s(x)$ introduced by Ledoit and Wolf (2015) and reproduced in Appendix B. Then the covariance matrix estimator

$$\widehat{S}_T := U_T \widehat{D}_T U_T' \quad \text{where} \quad \widehat{D}_T := \mathsf{Diag}\big(\widehat{d}_T(\lambda_{T,1}), \dots, \widehat{d}_T(\lambda_{T,N})\big) \tag{14}$$

$$\text{and} \quad \forall i=1,\ldots,N \qquad \widehat{d}_T(\lambda_{T,i}) := \begin{cases} \dfrac{1}{\lambda_{T,i}\left|\widehat{s}(\lambda_{T,i})\right|^2} & \text{if } \lambda_{T,i}>0, \\ \dfrac{1}{\left(\frac{N}{T}-1\right)\widehat{s}(0)} & \text{if } N>T \text{ and } \lambda_{T,i}=0 \end{cases} \tag{15}$$

minimizes in the class of rotation-equivariant estimators defined in Assumption 4 the almost sure limit of the portfolio-selection loss function $\mathcal{L}_T$ as T and N go to infinity together in the manner of Assumption 1.

This corollary is given without proof, as it is an immediate consequence of Theorem 2 above, via Ledoit and Wolf (2012, Proposition 4.3; 2015, Theorem 2.2).

3.5 Alternative covariance matrix estimators

Estimation of a large-dimensional covariance matrix has become a large and active field of research in recent times. A comprehensive review of the entire literature is clearly beyond the scope of the present paper. Nevertheless, we can make some remarks to put our contribution into perspective.

There are two broad avenues for estimating a covariance matrix when the number of variables is of the same magnitude as the sample size: structure-based estimation and structure-free estimation.

Structure-based estimation makes the problem more amenable by assuming additional structure on the covariance matrix. The three most commonly used sorts of structure in this avenue are sparsity, graph models, and (approximate) factor models. Sparsity assumes that most entries in the covariance matrix are (near) zero; graph models assume that most entries in the inverse covariance matrix are (near) zero. Neither assumption is generally realistic for a covariance matrix of financial returns. Factor models, on the other hand, have a long history in finance; for example, see Campbell et al. (1997, Chapter 6) and Ahn et al. (2009). An exact factor model assumes a known number of factors and a diagonal covariance matrix of the error terms. Weaker forms assume an unknown number of factors and/or sparsity of the covariance matrix of the error terms.[6] Since the number of factors is always assumed to be small and fixed, exact factor models have $O(1)$ degrees of freedom. If the number of factors is estimated from the data, there is one additional degree of freedom. If the covariance matrix of the error terms is only assumed to be sparse rather than diagonal (i.e., an approximate factor model), the additional degrees of freedom depend on the thresholding scheme applied to the sample covariance matrix of the residuals of the estimated factor model. To this end, one simply uses a scheme from the literature on estimating a sparse covariance matrix and most

[6] Note that a diagonal matrix is a special case of a sparse matrix.

such schemes only have one degree of freedom; for example, see Bickel and Levina (2008), Cai and Liu (2011), and Fan et al. (2013). As a result, even approximate factor models generally only have $O(1)$ degrees of freedom.

Structure-free estimation typically falls in our rotation-equivariant framework. As we have explained, the method of Ledoit and Wolf (2004b) has $O(1)$ degrees of freedom, whereas our new proposal has $O(N)$ degrees of freedom. Another recent method is the one of Won et al. (2013) which has $O(1)$ degrees of freedom.[7]

4. Empirical Results

The goal of this section is to examine the out-of-sample properties of Markowitz portfolios based on our newly suggested covariance matrix estimator. In particular, we make comparisons to other popular investment strategies in the finance literature; some of these are based on an alternative covariance matrix estimator while others avoid the problem of estimating the covariance matrix altogether.

For compactness of notation, similarly to Section 1, we do not use the subscript T in denoting the covariance matrix itself, an estimator of the covariance matrix, or a return predictive signal that proxies for the vector of expected returns.

4.1 Data and general portfolio-formation rules

We download daily data from the Center for Research in Security Prices (CRSP) starting in 1 January 1972 and ending in 31 December 2011. For simplicity, we adopt the common convention that 21 consecutive trading days constitute one "month". The out-of-sample period ranges from 19 January 1973 through 31 December 2011, resulting in a total of 480 months (or 10,080 days). All portfolios are updated monthly.[8] We denote the investment dates by $h = 1, \dots, 480$. At any investment date h, a covariance matrix is estimated using the most recent $T = 250$ daily returns, corresponding roughly to one year of past data.

We consider the following portfolio sizes: $N \in \{30, 50, 100, 250, 500\}$. This range covers the majority of the important stock indexes, from the Dow Jones Industrial Average to the S&P 500. For a given combination (h, N), the investment universe is obtained as follows. We first determine the 500 largest stocks, as measured by their market value on the investment date h, that have a complete return history over the most recent $T = 250$ days as well as

[7] The method winsorizes the sample eigenvalues and the two degrees of freedom are the two points of winsorization, at the lower end and at the upper end of the spectrum.

[8] Monthly updating is common practice to avoid an unreasonable amount of turnover, and thus transaction costs.

a complete return "history" over the next 21 days.[9] Out of these 500 stocks, we then select N at random: These N randomly selected stocks constitute the investment universe for the upcoming 21 days. As a result, there are 480 different investment universes over the out-of-sample period.

4.2 Rotation equivariance and preconditioning

The focus of our paper is mainly on rotation-equivariant estimators, but in the empirical study we also include some other estimators for the sake of comparison and completeness.

What rotation-equivariant estimation really means is that the researcher does not have any a priori beliefs about the orientation of the population eigenvectors. It is thus the most general and neutral approach, which is why we favor it. There have been several recent proposals for the estimation of optimized portfolios that fall in the rotation-equivariant framework, and we shall compare our nonlinear shrinkage estimator to them.

In order to make a comparison with factor models, which are not in the class of rotation-equivariant estimators, we have to come up with an adaptation of our method that breaks rotation equivariance. The way we do it is by preconditioning the data according to a simple model. We choose an exact factor model with a single factor: the return on the equal-weighted portfolio of the stocks in the investment universe.[10] Let $\widehat{\Sigma}_F$ denote the covariance matrix estimator that comes from fitting this exact factor model. We precondition the data by right-multiplying the observation matrix Y_T by $\widehat{\Sigma}_F^{-1/2}$. Doing so removes the structure contained in the factor matrix $\widehat{\Sigma}_F$. We then apply our nonlinear shrinkage technology to the preconditioned data $Y_T \times \widehat{\Sigma}_F^{-1/2}$, which yields an output $\widehat{\Sigma}_C$. The final estimator is then obtained by reincorporating the structure from the factor model:

$$\widehat{\Sigma} := \widehat{\Sigma}_F^{1/2} \times \widehat{\Sigma}_C \times \widehat{\Sigma}_F^{1/2} \,. \tag{16}$$

This approach can accomodate other a priori beliefs about the orientation of the population eigenvectors simply by changing the preconditioning matrix $\widehat{\Sigma}_F$.

4.3 Global minimum-variance portfolio

We consider the problem of estimating the global minimum-variance (GMV) portfolio in the absence of short-sales constraints.[11] The problem is

---

[9] The latter, forward-looking restriction is not a feasible one in real life but is commonly applied in the related finance literature on the out-of-sample evaluation of portfolios.

[10] The motivation for using the equal-weighted portfolio of the stocks in the investment universe as (single) factor is that no outside information is needed. As a result, the method is entirely self-contained and can be applied by anyone to any universe of assets.

[11] The problem of estimating a full Markowitz portfolio with momentum signal is considered in Appendix D.

formulated as

$$\min_{w} w'\Sigma w \tag{17}$$

$$\text{subject to} \quad w'\mathbf{1}=1 \,, \tag{18}$$

where $\mathbf{1}$ denotes a vector of ones of dimension $N \times 1$. It has the analytical solution

$$w = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \,. \tag{19}$$

The natural strategy in practice is to replace the unknown Σ by an estimator $\widehat{\Sigma}$ in formula (19), yielding a feasible portfolio

$$\widehat{w} := \frac{\widehat{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}} \,. \tag{20}$$

Alternative strategies, motivated by estimating the optimal w of (19) directly, as opposed to indirectly via the estimation of Σ, have been proposed recently by Frahm and Memmel (2010).

Estimating the GMV portfolio is a clean problem in terms of evaluating the quality of a covariance matrix estimator, since it abstracts from having to estimate the vector of expected returns at the same time. In addition, researchers have established that estimated GMV portfolios have desirable out-of-sample properties not only in terms of risk but also in terms of reward-to-risk (i.e., in terms of the Sharpe ratio); for example, see Haugen and Baker (1991), Jagannathan and Ma (2003), and Nielsen and Aylursubramanian (2008). As a result, such portfolios have become an addition to the large array of products sold by the mutual fund industry.

The following 11 portfolios are included in the study.

- **$1/N$**: The equal-weighted portfolio promoted by DeMiguel et al. (2009), among others. This portfolio can be viewed as a special case of (20) where $\widehat{\Sigma}$ is given by the $N \times N$ identity matrix. This strategy avoids any parameter estimation whatsoever.
- **Sample:** The portfolio (20) where $\widehat{\Sigma}$ is given by the sample covariance matrix; note that this portfolio is not available when $N > T$, since the sample covariance matrix is not invertible in this case.
- **FM:** The dominating portfolio of Frahm and Memmel (2010). The particular version we use is defined in their Equation (10), where the reference portfolio w_R is given by the equal-weighted portfolio. This portfolio is a convex linear combination of the two previous portfolios $1/N$ and Sample. Therefore, it is also not available when $N > T$.
- **FYZ:** The GMV portfolio with gross-exposure constraint of Equation (2.6) of Fan et al. (2012). As suggested in their paper, we take the sample covariance matrix as an estimator of Σ and set the gross-exposure constraint parameter equal to $c=2$.

- **Lin:** The portfolio (20) where the matrix $\widehat{\Sigma}$ is given by the linear shrinkage estimator of Ledoit and Wolf (2004b).
- **NonLin:** The portfolio (20) where $\widehat{\Sigma}$ is given by the estimator $\widehat{S}$ of Corollary 1.
- **NL-Inv:** The portfolio (20) where $\widehat{\Sigma}^{-1}$ is given by the direct nonlinear shrinkage estimator of Σ^{-1} based on generic a Frobenius-norm loss. This estimator was first suggested by Ledoit and Wolf (2012) for the case $N<T$; the extension to the case $T \geq N$ can be found in Ledoit and Wolf (2017).
- **SF:** The portfolio (20) where $\widehat{\Sigma}$ is given by the single-factor covariance matrix $\widehat{\Sigma}_F$ used in the construction of the single-factor-preconditioned nonlinear shrinkage estimator (16).
- **FF:** The portfolio (20) where $\widehat{\Sigma}$ is given by the covariance matrix estimator based on the (exact) three-factor model of Fama and French (1993).[12]
- **POET:** The portfolio (20) where $\widehat{\Sigma}$ is given by the POET covariance matrix estimator of Fan et al. (2013). This method uses an approximate factor model where the factors are taken to be the principal components of the sample covariance matrix and thresholding is applied to covariance matrix of the principal orthogonal complements.[13]
- **NL-SF:** The portfolio (20) where $\widehat{\Sigma}$ is given by the single-factor-preconditioned nonlinear shrinkage estimator (16).

We report the following three out-of-sample performance measures for each scenario. (All measures are annualized and in percent for ease of interpretation.)

- **AV:** We compute the average of the 10,080 out-of-sample returns in excess of the risk-free rate and then multiply by 250 to annualize.
- **SD:** We compute the standard deviation of the 10,080 out-of-sample returns in excess of the risk-free rate and then multiply by $\sqrt{250}$ to annualize.
- **SR:** We compute the (annualized) Sharpe ratio as the ratio AV/SD.

Our stance is that in the context of the GMV portfolio, the most important performance measure is the out-of-sample standard deviation, SD. The true (but unfeasible) GMV portfolio is given by (19). It is designed to minimize the variance (and thus the standard deviation) rather than to maximize the expected return or the Sharpe ratio. Therefore, any portfolio that implements the GMV portfolio should be primarily evaluated by how successfully it achieves this goal.

12 Data on the three Fama-French factors were downloaded from Ken French's Data Library.

13 In particular, we use $K=5$ factors, soft thresholding, and the value of $C=1.0$ for the thresholding parameter. Among several specifications we tried, this one worked best on average.

Table 1
Performance measures for various estimators of the GMV portfolio

Period: January 19, 1973 to December 31, 2011

	1/*N*	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	11.14	8.64	8.65	8.68	8.52	8.71	8.72	8.22	9.39	8.29	8.88
SD	20.05	14.21	14.11	14.11	14.16	**14.08***	14.08	14.08	14.59	14.49	**14.00**
SR	0.56	0.61	0.61	0.62	0.60	0.62	0.62	0.56	0.66	0.57	0.63
						$N=50$					
AV	9.54	4.65	4.99	4.70	5.10	5.21	5.22	5.22	5.44	5.59	5.44
SD	19.78	13.15	13.01	12.83	12.75	**12.68*****	12.68	13.04	12.51	12.60	**12.28**
SR	0.48	0.35	0.38	0.37	0.40	0.41	0.41	0.40	0.43	0.44	0.44
						$N=100$					
AV	10.53	4.74	5.31	4.96	4.99	5.10	5.12	4.81	5.80	4.83	5.07
SD	19.34	13.11	12.71	11.75	11.79	**11.52*****	11.55	11.96	11.30	11.31	**10.99**
SR	0.54	0.36	0.42	0.42	0.42	0.44	0.44	0.40	0.51	0.43	0.46
						$N=250$					
AV	9.57	275.02	275.02	6.73	5.81	6.26	6.43	5.95	6.60	5.90	5.71
SD	18.95	3,542.90	3,542.90	10.69	10.91	**10.34*****	10.49	11.30	10.47	9.93	**9.46**
SR	0.50	0.08	0.08	0.63	0.53	0.61	0.61	0.52	0.63	0.59	0.60
						$N=500$					
AV	9.78	*NA*	*NA*	5.90	5.03	5.34	5.41	5.30	5.87	5.32	5.53
SD	18.95	*NA*	*NA*	10.21	10.20	**9.65*****	9.75	11.07	10.06	9.25	**8.61**
SR	0.52	*NA*	*NA*	0.58	0.49	0.55	0.55	0.48	0.58	0.57	0.64

AV, average; SD, standard deviation; SR, Sharpe ratio; NA, not available. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

We also consider the question of whether one portfolio delivers a lower out-of-sample standard deviation than another portfolio at a level that is statistically significant. Since we consider 11 portfolios, there are 55 pairwise comparisons. To avoid a multiple-testing problem and since a major goal of this paper is to show that nonlinear shrinkage improves upon linear shrinkage in portfolio selection, we restrict attention to the single comparison between the two portfolios Lin and NonLin. For a given scenario, a two-sided p-value for the null hypothesis of equal standard deviations is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2011, Section 3.1).[14]

Table 1 reports the results, which can be summarized as follows.

- The standard deviation of the true GMV portfolio (19) decreases in N. So the same should be true for any good estimator of the GMV portfolio. As N increases from $N=30$ to $N=500$, the standard deviation of $1/N$ decreases by only 1.1 percentage points. On the other hand, the standard

[14] Since the out-of-sample size is very large at 10,080, there is no need to use the computationally more involved bootstrap method described in Ledoit and Wolf (2011, Section 3.2), which is preferred for small sample sizes.

deviations of Lin and Nonlin decrease by 3.9 and 4.4 percentage points, respectively. Therefore, sophisticated estimators of the GMV portfolio are successful in overcoming the increased estimation error for a larger number of assets and indeed deliver a markedly better performance.

- $1/N$ is consistently outperformed in terms of the standard deviation by all other portfolios with the exception of Sample and FM for $N=250$, when the sample covariance matrix is nearly singular.
- FM improves upon Sample but, in turn, is outperformed by the other sophisticated rotation-equivariant portfolios. It is generally also outperformed by the factor-based portfolios, with the exception of FF and POET for $N=30$.
- The performance of FZY and Lin is comparable.
- NonLin has the uniformly best performance among the rotation-equivariant portfolios and the outperformance over Lin is statistically significant at the 0.1 level for $N=30$ and statistically significant at the 0.01 level for $N=50, 100, 250, 500$.
- In terms of economic significance, for $N = 250$ and 500, we get Sharpe ratio gains of 0.08 and 0.06, respectively. In relative terms, this corresponds to boosts of 15% and 12%, respectively. (Note that even stronger gains are realized in the full Markowitz portfolio with momentum signal; see Appendix D).
- NonLin also outperforms NL-Inv, though the differences are always small.
- Among the four factor-based portfolios, NL-SF is uniformly the best; in particular, NL-SF outperforms the three-factor model FF which in return outperforms the single-factor model SF. NL-SF outperforms NonLin in terms of the standard deviation and, generally, also in terms of the Sharpe ratio (except for $N=250$).

Summing up, in the global minimum-variance portfolio problem, NonLin dominates the other six rotation-equivariant portfolios portfolios in terms of the standard deviation and, in addition, dominates Lin in terms of the Sharpe ratio. NL-SF constitutes a further improvement over NonLin.

Remark 7. It is true that the economic significance of our improvements is stronger for larger values of N. When academics research anomalies in the cross-sectional of stock returns, forming low-volatility portfolios that load on the candidate characteristic produces a more powerful test than looking at the top decile minus the bottom decile. This requires the estimated covariance matrix for all the stocks in the CRSP universe that are alive at a given point in time, which reaches well into the thousands. Bell et al. (2014) show how such an approach can be implemented. In view of the heightened t-statistic thresholds advocated by Harvey et al. (2016) to deal with the multiple testing problem, we are going to need a more powerful test.

A second mission of finance professors is to forge tools that can be used by practictioners to implement investment methodologies that are scientifically correct, and in the simplest sense this means: Markowitz portfolio selection. Quantitative asset managers specializing in single-stock equities commonly use large values of N so that they benefit from a cross-sectional law of large numbers. For example, in the USA, there is decent liquidity in the constituents of the Russell 3000 Index. In Europe, it is easy to get a 600-stock universe with sufficient liquidity, and the same again in Japan. This is also true of hedge fund strategies such as Statistical Arbitrage, even though they tend to have a higher turnover than classic long-only fund managers. As for pure technical players, who display a strong preference for liquid stocks, they can still find more than 500 names worthy of being traded in the USA. Therefore, demand for covariance matrix estimators that excel in the large-N domain is strong. □

4.4 Analysis of weights

We also provide some summary statistics on the vectors of portfolio weights $\widehat{w}$ over time. In each month, we compute the following four characteristics:

- **Min:** Minimum weight.
- **Max:** Maximum weight.
- **SD:** Standard deviation of weights.
- **MAD-EW:** Mean absolute deviation from equal-weighted portfolio computed as
$$\frac{1}{N}\sum_{i=1}^{N}\left|\widehat{w}_i - \frac{1}{N}\right|.$$

For each characteristic, we then report the average outcome over the 480 portfolio formations.

Table 2 reports the results. Not surprisingly, the most dispersed weights among the rotation-equivariant portfolios are found for Sample, followed by FM and FZY. The three shrinkage methods have generally the least dispersed weights, with NonLin and NL-Inv being more dispersed than Lin for $N = 30, 50$ and less dispersed than Lin for $N = 100, 250, 500$.

There is no clear ordering among the four factor-based portfolios and the dispersion of their weights is comparable to the rotation-equivariant shrinkage portfolios.

4.5 Robustness checks

The goal of this section is to examine whether the outperformance of NonLin over Lin is robust to various changes in the empirical analysis.

Subperiod analysis. The out-of-sample period comprises 480 months (or 10,080 days). It might be possible that the outperformance if NonLin over

Table 2
Average characteristics of the weight vectors of various estimators of the GMV portfolio

Period: January 19, 1973 to December 31, 2011

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
					$N=50$						
Min	0.0333	−0.0729	−0.0653	−0.0690	−0.0522	−0.0584	−0.0582	−0.0463	−0.0533	−0.0658	−0.0540
Max	0.0333	0.2720	0.2562	0.2698	0.1902	0.2210	0.2228	0.2440	0.2472	0.2361	0.2536
SD	0.0000	0.0737	0.0686	0.0726	0.0562	0.0614	0.0616	0.0675	0.0669	0.0678	0.0680
MAD-EW	0.0000	0.0515	0.0479	0.0507	0.0423	0.0444	0.0444	0.0491	0.0476	0.0495	0.0481
					$N=50$						
Min	0.0200	−0.0762	−0.0680	−0.0695	−0.0531	−0.0537	−0.0536	−0.0393	−0.0497	−0.0496	−0.0483
Max	0.0200	0.2219	0.2055	0.2211	0.1461	0.1556	0.1589	0.1861	0.1945	0.1790	0.1952
SD	0.0000	0.0552	0.0506	0.0533	0.0411	0.0415	0.0418	0.0458	0.0471	0.0451	0.0471
MAD-EW	0.0000	0.0386	0.0353	0.0367	0.0309	0.0306	0.0306	0.0327	0.0329	0.0323	0.0330
					$N=100$						
Min	0.0100	−0.0837	−0.0733	−0.0598	−0.0499	−0.0423	−0.0424	−0.0281	−0.0390	−0.0346	−0.0397
Max	0.0100	0.1776	0.1595	0.1684	0.0989	0.0846	0.0890	0.1208	0.1307	0.1208	0.1354
SD	0.0000	0.0407	0.0362	0.0333	0.0270	0.0234	0.0237	0.0258	0.0276	0.0258	0.0279
MAD-EW	0.0000	0.0288	0.0256	0.0218	0.0205	0.0180	0.0181	0.0182	0.0191	0.0182	0.0192
					$N=250$						
Min	0.0040	−7.2464	−7.2464	−0.0438	−0.0362	−0.0260	−0.0263	−0.0151	−0.0228	−0.0210	−0.0251
Max	0.0040	6.7094	6.7094	0.1225	0.0530	0.0357	0.0362	0.0628	0.0716	0.0684	0.0791
SD	0.0000	1.9296	1.9296	0.0172	0.0150	0.0108	0.0108	0.0113	0.0127	0.0121	0.0134
MAD-EW	0.0000	1.4159	1.4159	0.0098	0.0118	0.0085	0.0086	0.0079	0.0087	0.0084	0.0092
					$N=500$						
Min	0.0020	NA	NA	−0.0359	−0.0232	−0.0167	−0.0164	−0.0089	−0.0140	−0.0169	−0.0166
Max	0.0020	NA	NA	0.0998	0.0293	0.0199	0.0203	0.0364	0.0430	0.0446	0.0505
SD	0.0000	NA	NA	0.0106	0.0083	0.0059	0.0059	0.0058	0.0067	0.0068	0.0075
MAD-EW	0.0000	NA	NA	0.0052	0.0066	0.0046	0.0047	0.0041	0.0046	0.0047	0.0052

Min, mininum weight; Max, maximum weight; SD, standard deviation of the weights; MAD-EW, mean absolute deviation from the equal-weighted portfolio (i.e., from $1/N$); NA, not available. All measures reported are the averages of the corresponding characteristic over the 480 portfolio formations.

Lin is driven by certain subperiods but does not hold universally. We address this concern by dividing the out-of-sample period into three subperiods of 160 months (or 3,360 days) each and repeating the above exercises in each subperiod.

Tables 3–5 report the results. It can be seen that among the rotation-equivariant portfolios, NonLin has the best performance in terms of the standard deviation in 15 out of the 15 cases and that the outperformance over Lin is generally with statistical significance for $N \geq 50$. Among the factor-based portfolios, NL-SF has the best performance in 14 out of the 15 cases and it constitutes a further improvement over NonLin.

Therefore, this analysis demonstrates that the outperformance of NonLin over Lin is consistent over time and not due to a subperiod artifact.

Table 3
Performance measures for various estimators of the GMV portfolio

Period: January 19, 1973 to May 8, 1986

	1/*N*	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	7.64	4.97	5.00	4.97	5.49	5.40	5.39	5.12	6.53	5.46	5.91
SD	15.43	11.63	11.54	11.63	11.42	**11.39**	11.39	**11.21**	11.30	11.46	11.32
SR	0.50	0.43	0.43	0.43	0.48	0.47	0.47	0.46	0.58	0.48	0.52
						$N=50$					
AV	7.20	3.63	3.93	3.57	4.58	4.68	4.68	5.28	4.95	4.93	4.84
SD	14.83	10.91	10.74	10.88	10.54	**10.47****	10.47	10.40	10.35	10.44	**10.28**
SR	0.49	0.33	0.37	0.33	0.43	0.45	0.45	0.51	0.48	0.47	0.47
						$N=100$					
AV	8.62	6.06	6.09	6.54	6.82	6.92	6.90	6.99	5.90	6.57	6.60
SD	14.50	10.39	10.03	9.61	9.43	**9.18*****	9.18	9.03	8.88	8.89	**8.77**
SR	0.59	0.58	0.61	0.68	0.72	0.75	0.75	0.77	0.67	0.74	0.75
						$N=250$					
AV	6.16	−527.27	−527.27	4.78	5.08	6.09	5.96	5.79	3.85	4.24	4.61
SD	14.18	2,009.60	2,009.60	8.41	8.51	**7.81*****	7.86	7.84	7.53	7.39	**7.14**
SR	0.43	−0.26	−0.26	0.57	0.60	0.78	0.76	0.74	0.51	0.57	0.65
						$N=500$					
AV	6.91	NA	NA	3.35	4.20	5.45	5.41	6.14	3.99	3.90	4.57
SD	14.14	NA	NA	7.82	7.50	**7.03*****	7.14	7.58	7.10	6.82	**6.49**
SR	0.49	NA	NA	0.43	0.56	0.77	0.76	0.81	0.56	0.57	0.70

AV, average; SD, standard deviation; SR, Sharpe ratio; NA, not available. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Longer estimation window. Generally, at any investment date h, a covariance matrix is estimated using the most recent $T=250$ daily returns, corresponding roughly to one year of past data. As a robustness check, we alternatively use the most recent $T=500$ daily returns, corresponding roughly to two years of past data.

Table 6 reports the results, which are similar to the results in Table 1. In particular, NonLin has the uniformly best performance among the rotation-equivariant portfolios in terms of the standard deviation and the outperformance over Lin is statistically significant for $N=100,250,500$. Again, NL-SF constitutes a further improvement.

Winsorization of past returns. Financial return data frequently contain unusually large (in absolute value) observations. In order to mitigate the effect of such observations on an estimated covariance matrix, we employ a winsorization technique, as is standard with quantitative portfolio managers; the details can be found in Appendix C. Of course, we always use the actual, non-winsorized data in the computation of the out-of-sample portfolio returns.

Table 7 reports the results. It can be seen that the relative performance of the various portfolios is similar to that in Table 1, although in absolute terms, the

Table 4
Performance measures for various estimators of the GMV portfolio

Period: May 9, 1986 to August 25, 1999

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	12.56	12.06	12.01	12.04	12.23	12.08	12.10	12.43	11.09	10.78	11.76
SD	16.24	12.91	12.81	12.81	12.78	**12.77**	12.77	12.77	12.71	13.14	**12.67**
SR	0.77	0.93	0.94	0.94	0.96	0.95	0.95	0.97	0.87	0.82	0.93
						$N=50$					
AV	12.94	7.22	7.69	7.58	7.20	7.62	7.60	7.78	8.50	8.40	8.24
SD	15.82	12.04	11.94	11.90	11.86	**11.85**	11.85	11.56	11.41	11.62	**11.33**
SR	0.82	0.60	0.64	0.64	0.61	0.64	0.64	0.67	0.75	0.72	0.73
						$N=100$					
AV	12.45	6.68	7.23	6.92	6.84	6.56	6.61	5.24	6.22	6.28	6.17
SD	15.37	11.42	11.10	10.63	10.39	**10.26*****	10.28	10.09	9.75	9.87	**9.75**
SR	0.81	0.58	0.65	0.65	0.66	0.64	0.64	0.52	0.64	0.64	0.63
						$N=250$					
AV	12.04	652.67	652.67	7.92	7.41	6.72	6.70	5.86	6.78	6.48	7.15
SD	15.09	2,126.98	2,126.98	10.07	9.78	**9.55*****	9.64	9.37	9.11	8.93	**8.73**
SR	0.80	0.31	0.31	0.79	0.76	0.70	0.70	0.63	0.74	0.73	0.82
						$N=500$					
AV	12.12	NA	NA	9.45	7.96	6.93	7.29	5.47	6.86	6.61	7.45
SD	15.01	NA	NA	9.57	9.21	**8.93*****	9.06	8.95	8.44	8.11	**7.86**
SR	0.81	NA	NA	0.99	0.86	0.78	0.80	0.61	0.81	0.81	0.95

AV, average; SD, standard deviation; SR, Sharpe ratio; NA, not available. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

performance is somewhat worse. Among the rotation-equivariant portfolios, NonLin is no longer uniformly best but it is always either best or a (very close) second-best after NL-Inv. Again, NL-SF constitutes a further improvement.

No-short-sales constraint. Since some fund managers face a no-short-sales constraint, we now impose a lower bound of zero on all portfolio weights.

Table 8 reports the results. Note that Sample is now available for all N whereas FM and FZY are not available at all. It can be seen that Sample is uniformly best among the rotation-equivariant portfolios in terms of the standard deviation, although the differences to Lin, NonLin, and NL-SF are always small. Comparing the results to those of Table 1 shows that disallowing short sales helps Sample but hurts Lin, NonLin, and NL-SF. These findings are consistent with Jagannathan and Ma (2003) who demonstrate theoretically that imposing a no-short-sales constraint corresponds to an implicit shrinkage of the sample covariance matrix in the context of estimating the global minimum-variance portfolio.

There is no clear winner among the factor-based portfolios: FF is best once, POET is best twice, and NL-SF is best twice. (On the other hand, there is a clear

Table 5
Performance measures for various estimators of the GMV portfolio

Period: August 26, 1999 to December 31, 2011

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	13.22	8.90	8.96	9.03	7.84	8.65	8.66	7.09	10.56	8.64	8.98
SD	26.55	17.43	17.32	17.27	17.54	**17.38**	17.38	18.71	17.89	18.06	**17.29**
SR	0.50	0.51	0.52	0.52	0.45	0.50	0.50	0.38	0.59	0.48	0.52
						$N=50$					
AV	8.49	3.10	3.34	2.95	3.52	3.34	3.38	2.60	2.88	3.46	3.25
SD	26.53	15.97	15.80	15.30	15.36	**15.25****	15.25	16.39	15.25	15.25	**14.79**
SR	0.32	0.19	0.21	0.19	0.23	0.22	0.22	0.15	0.19	0.23	0.22
						$N=100$					
AV	10.52	1.47	2.61	1.43	1.32	1.82	1.86	2.19	5.27	1.65	2.43
SD	25.99	16.65	16.16	14.46	14.84	**14.46*****	14.49	15.68	14.46	14.39	**13.80**
SR	0.40	0.09	0.16	0.10	0.09	0.13	0.13	0.14	0.36	0.11	0.18
						$N=250$					
AV	10.50	699.67	699.67	7.50	4.95	5.97	6.63	6.20	9.17	7.00	5.36
SD	25.47	5,394.22	5,394.22	13.07	13.76	**12.99*****	13.24	15.29	13.75	12.72	**11.90**
SR	0.41	0.13	0.13	0.57	0.36	0.46	0.50	0.41	0.67	0.55	0.45
						$N=500$					
AV	10.30	NA	NA	4.91	2.93	3.65	3.53	4.29	6.76	5.45	4.57
SD	25.54	NA	NA	12.64	13.07	**12.24*****	12.34	15.16	13.49	12.03	**10.89**
SR	0.40	NA	NA	0.39	0.22	0.30	0.29	0.28	0.50	0.45	0.42

AV, average; SD, standard deviation; SR, Sharpe ratio; NA, not available. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

loser, namely SF which is always worst.) Overall, the factor-based portfolios have somewhat better performance than the rotation-equivariant portfolios.

Remark 8. Our method can still be useful for long-only managers of a certain type; namely, those who are benchmarked against a passive index (such as the S&P 500) and manage their active risk. The active portfolio of such managers is really a long-short dollar-neutral overlay on top of the passive benchmark weights. If the active short position is well diversified, the overall no-short-sales constraint is not very binding. This is when having a good estimator of the covariance matrix of the active positions can really pay off again. □

Transaction costs. An important consideration in any practical implementation of portfolio rules are transaction costs. None of our results so far take transaction costs into account. In our setting, transaction costs would arise due to two unrelated causes: (1) the investment universe changes from month to month and (2) for the stocks that belong to successive investment universes, the portfolio weights change.

As described in Section 4.1, at the beginning of every month, the portfolio universe is determined by selecting N stocks at random from the 500 largest

Table 6
Performance measures for various estimators of the GMV portfolio

Period: January 19, 1973 to December 31, 2011

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	11.14	8.68	8.73	8.61	8.80	8.78	8.77	8.56	9.39	8.29	8.88
SD	20.05	14.15	14.11	14.12	14.11	**14.09**	14.09	14.82	14.25	14.49	**14.00**
SR	0.56	0.61	0.62	0.61	0.62	0.62	0.62	0.58	0.66	0.57	0.63
						$N=50$					
AV	9.54	6.20	6.32	6.06	6.21	6.25	6.26	5.49	5.74	5.60	5.58
SD	19.78	12.87	12.82	12.70	12.67	**12.67**	12.67	13.34	12.74	12.79	**12.51**
SR	0.48	0.48	0.49	0.48	0.49	0.49	0.49	0.41	0.45	0.44	0.45
						$N=100$					
AV	10.53	4.74	5.07	5.23	5.17	5.31	5.30	5.04	6.11	4.63	5.09
SD	19.34	11.95	11.85	11.51	11.59	**11.48*****	11.48	12.31	11.60	11.50	**11.19**
SR	0.54	0.40	0.43	0.45	0.45	0.46	0.46	0.41	0.53	0.40	0.45
						$N=250$					
AV	9.57	7.43	7.47	6.69	6.77	6.66	6.79	6.51	6.14	5.08	5.24
SD	18.95	11.83	11.55	10.21	10.51	**10.09*****	10.12	11.67	10.87	10.26	**9.63**
SR	0.50	0.63	0.65	0.66	0.64	0.66	0.67	0.56	0.56	0.50	0.54
						$N=500$					
AV	9.78	1,200.31	1,200.31	6.83	6.66	6.25	6.33	5.99	6.73	5.58	6.21
SD	18.95	8,551.27	8,551.26	9.54	10.33	**9.43*****	9.87	11.50	10.52	9.39	**8.70**
SR	0.52	0.14	0.14	0.72	0.64	0.66	0.64	0.52	0.64	0.59	0.71

The past window to estimate the covariance matrix is taken to be of length $T=500$ days instead of $T=250$ days. AV, average; SD, standard deviation; SR, Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

stocks (as measured by their market value) that have a complete return history. So unless $N=500$, there will be high transaction costs due these drastic changes in the investment universe alone. Such a rule would not be of interest in any practical implementation; instead, the investment manager would ensure that the investment universe changes only slowly over time.

In our rules, the portfolio selection at the beginning of a month is unconstrained and does not pay any attention to the weights of various stocks at the end of the previous month.[15] In a practical implementation, it might be preferred to use constrained portfolio selection to actively limit portfolio turnover by taking into account the portfolio weights at the end of the previous month; for example, see Yoshimoto (1996).

A detailed empirical study of real-life constrained portfolio selection that actively limits portfolio turnover (and thus transaction costs) from one month to the next is beyond the scope of the present paper.

[15] Note that the weights at the end of a given month are not equal to the weights at the beginning of that month; this is because during a month, the number of shares are held fixed rather than the portfolio weights (which would require daily rebalancing).

Table 7
Performance measures for various estimators of the GMV portfolio

Period: January 19, 1973 to December 31, 2011

	1/*N*	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	11.14	8.27	8.30	8.31	8.21	8.42	8.44	6.26	8.93	9.03	8.83
SD	20.05	14.57	14.42	14.46	14.38	14.38	**14.37**	14.82	14.35	14.79	**14.11**
SR	0.56	0.57	0.58	0.57	0.57	0.59	0.59	0.42	0.62	0.61	0.63
						$N=50$					
AV	9.54	4.60	4.94	4.55	5.00	5.29	5.31	5.52	5.60	5.24	5.36
SD	19.78	13.56	13.34	13.09	13.13	12.97***	**12.95**	13.53	12.77	13.18	**12.47**
SR	0.48	0.34	0.37	0.35	0.38	0.41	0.41	0.41	0.44	0.40	0.43
						$N=100$					
AV	10.53	4.84	5.29	5.71	4.89	5.04	5.08	6.40	6.14	5.37	5.08
SD	19.34	13.86	13.29	12.20	12.54	11.95***	**11.93**	12.43	11.56	11.46	**11.26**
SR	0.54	0.35	0.40	0.47	0.39	0.42	0.43	0.51	0.53	0.47	0.45
						$N=250$					
AV	9.57	−2,498.27	−2,498.27	6.99	6.72	6.70	6.75	5.61	6.80	5.96	5.96
SD	18.95	12,130.15	12,130.15	10.81	11.75	**10.58***	10.58	11.67	10.60	9.84	**9.61**
SR	0.50	−0.21	−0.21	0.65	0.57	0.63	0.64	0.48	0.64	0.61	0.62
						$N=500$					
AV	9.78	NA	NA	5.86	5.44	5.56	5.68	5.28	6.03	5.47	5.64
SD	18.95	NA	NA	10.33	10.83	**9.71***	9.78	11.36	10.18	8.91	**8.70**
SR	0.52	NA	NA	0.57	0.50	0.57	0.58	0.47	0.59	0.61	0.65

In the estimation of a covariance matrix, the past returns are winsorized as described in Appendix C. AV, average; SD, standard deviation; SR, Sharpe ratio; NA, not available. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Instead, we provide some limited results for unconstrained portfolio selection with $N=500$ only (to limit the contribution due to cause (1), changing investment universes). We assume a bid-ask spread ranging from three to fifty basis points. This number three is rather low by academic standards but can actually be considered an upper bound for liquid stocks nowadays; for example, see Avramovic and Mackintosh (2013) and Webster et al. (2015, p.33).

Table 9 reports the results, which are virtually unchanged compared to the results for $N=500$ in Table 1 in terms of the standard deviation, though all portfolios suffer in terms of the average return and the Sharpe ratio. Unsurprisingly, $1/N$ suffers the least and is the only portfolio that still has positive average return for a bid-ask-spred of fifty basis points. Furthermore, it is noteworthy that the nonlinear shrinkage portfolios NonLin, NL-Inv, and NL-SF all have lower average turnover than the linear shrinkage portfolio Lin, and are therefore less affected by trading costs.

Different return frequency. Finally, we change the return frequency from daily to monthly. As there is a longer history available for monthly returns, we download data from CRSP from January 1945 through December 2011. We use the $T=120$ most recent months of previous data to estimate a covariance matrix.

Table 8
Performance measures for various estimators of the GMV portfolio

Period: January 19, 1973 to December 31, 2011

	1/*N*	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	11.14	9.75	NA	NA	9.87	9.92	9.92	8.22	9.45	9.34	9.54
SD	20.05	**14.21**	NA	NA	14.41	14.26	14.26	14.59	14.25	14.15	**14.13**
SR	0.56	0.69	NA	NA	0.69	0.70	0.70	0.56	0.66	0.66	0.67
						$N=50$					
AV	9.54	6.97	NA	NA	6.99	7.10	7.08	5.22	7.48	7.26	7.09
SD	19.78	**12.96**	NA	NA	13.06	12.99	12.99	13.04	12.81	12.78	**12.77**
SR	0.48	0.54	NA	NA	0.54	0.55	0.55	0.40	0.58	0.57	0.55
						$N=100$					
AV	10.53	6.93	NA	NA	7.22	7.37	7.35	4.81	7.49	6.86	6.91
SD	19.34	**12.04**	NA	NA	12.14	12.17	12.18	11.96	**11.87**	11.92	11.90
SR	0.54	0.58	NA	NA	0.59	0.61	0.60	0.40	0.63	0.58	0.58
						$N=250$					
AV	9.57	7.40	NA	NA	7.40	7.51	7.58	5.95	7.50	7.19	7.05
SD	18.95	**11.06**	NA	NA	11.09	11.20	11.29	11.30	10.80	**10.75**	10.81
SR	0.50	0.67	NA	NA	0.67	0.67	0.67	0.53	0.69	0.67	0.65
						$N=500$					
AV	9.78	7.21	NA	NA	7.16	7.54	7.56	6.42	7.16	7.05	7.17
SD	18.95	**10.57**	NA	NA	10.59	10.74	10.83	10.68	10.40	**10.28**	10.31
SR	0.52	0.68	NA	NA	0.68	0.70	0.70	0.60	0.69	0.69	0.70

A lower bound of zero is imposed on all portfolio weights, so that short sales are not allowed. AV, average; SD, standard deviation; SR, Sharpe ratio; NA, not available. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Consequently, the out-of-sample investment period ranges from January 1955 through December 2011, yielding 684 out-of-sample returns. The remaining details are as before.

Table 10 reports the results. It can be seen that the relative performance of the various portfolios is similar to that in Table 1. The only difference is that NL-Inv is now sometimes better than NonLin, though the differences are always small. As with daily returns, NonLin is always better than Lin and the outperformance is statistically significant for $N=50, 100, 250, 500$. Again, NL-SF constitutes a further improvement over NonLin.

Different data sets. So far, we have focused on individual stocks as assets, since we believe this is the most relevant case for fund managers. On the other hand, many academics also consider the case in which the assets are portfolios.

To check the robustness of our findings in this regard, we consider three universes of size $N=100$ from Ken French's Data Library:

- 100 portfolios formed on size and book-to-market
- 100 Portfolios formed on size and operating profitability
- 100 Portfolios formed on size and investment

Table 9
Performance measures for various estimators of the GMV portfolio

Period: January 19, 1973 to December 31, 2011

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=500$, BAS = 3 basis points					
AV	9.53	NA	NA	4.90	3.67	4.49	4.54	4.72	5.18	4.56	4.68
SD	18.95	NA	NA	10.20	10.20	**9.65*****	9.75	11.06	10.06	9.26	**8.61**
SR	0.50	NA	NA	0.48	0.36	0.47	0.47	0.42	0.52	0.49	0.54
						$N=500$, BAS = 5 basis points					
AV	9.53	NA	NA	4.23	2.77	3.93	3.96	4.34	4.73	4.05	4.12
SD	18.95	NA	NA	10.21	10.21	**9.65*****	9.75	11.07	10.06	9.26	**8.61**
SR	0.50	NA	NA	0.41	0.27	0.41	0.39	0.39	0.47	0.44	0.48
						$N=500$, BAS = 10 basis points					
AV	8.95	NA	NA	2.56	0.51	2.51	2.49	3.37	3.58	2.78	2.71
SD	18.95	NA	NA	10.23	10.26	**9.67*****	9.77	11.08	10.08	9.28	**8.63**
SR	0.47	NA	NA	0.25	0.05	0.26	0.25	0.30	0.36	0.30	0.31
						$N=500$, BAS = 20 basis points					
AV	8.12	NA	NA	−0.78	−4.02	−0.33	−0.45	1.44	1.30	0.23	−0.10
SD	18.95	NA	NA	10.34	10.49	**9.75*****	9.87	11.11	10.13	9.36	**8.74**
SR	0.43	NA	NA	−0.07	−0.38	−0.03	−0.05	0.13	0.13	0.02	−0.01
						$N=500$, BAS = 50 basis points					
AV	5.66	NA	NA	−10.89	−17.60	−8.84	−9.27	−4.34	−5.55	−7.40	−8.55
SD	18.97	NA	NA	11.16	12.06	**10.42*****	10.57	11.39	10.55	9.95	**9.47**
SR	0.30	NA	NA	−0.97	−1.46	−0.85	−0.88	−0.38	−0.53	−0.74	−0.90
						$N=500$					
AT	0.69	NA	NA	2.81	3.81	2.39	2.47	1.62	1.92	2.14	2.37

AV, average; SD, standard deviation; SR, Sharpe ratio; AT, average turnover (from one month to the next); NA, not available. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate. The returns are adjusted for transaction costs assuming a bid-ask-spread (BAS) that ranges from three to fifty basis points. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

We use daily data. The out-of-sample period ranges from 13 December 1965 through 31 December 2015, resulting in a total of 600 months (or 12,600 days). At any investment date, a covariance matrix is estimated using the most recent $T=250$ daily returns.

Table 11 reports the results. It can be seen that the relative performance of the various portfolios is similar to that in Table 1, although in absolute terms, the performance is much better. The latter fact is not surprising, since portfolios are generally less risky compared to individual stocks. Among the rotation-equivariant portfolios, NonLin is uniformly best (and always significantly better than Lin). Again, NL-SF constitutes a further improvement.

4.6 Illustration of nonlinear versus linear shrinkage

We now use a specific data set to illustrate how nonlinear shrinkage can differ from linear shrinkage. Both estimators, as well as the sample covariance matrix, belong to the class of rotation-equivariant estimators introduced in

Table 10
Performance measures for various estimators of the GMV portfolio

Period: January 1955 through December 2011

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=30$					
AV	7.05	6.43	6.46	6.64	6.82	7.35	7.30	7.95	7.40	7.04	7.13
SD	16.69	14.25	13.68	13.90	13.48	13.45	**13.42**	13.90	13.63	14.19	**13.33**
SR	0.42	0.45	0.47	0.48	0.51	0.55	0.54	0.57	0.54	0.50	0.53
						$N=50$					
AV	6.74	3.30	3.95	3.25	4.37	4.47	4.56	5.16	5.08	4.29	4.18
SD	16.44	15.12	14.06	13.16	12.90	12.61**	**12.56**	13.24	12.59	12.80	**12.27**
SR	0.41	0.22	0.28	0.25	0.34	0.35	0.36	0.39	0.40	0.33	0.34
						$N=100$					
AV	6.29	9.22	8.77	6.30	6.66	6.32	6.45	6.31	5.50	5.66	5.91
SD	16.06	25.98	21.36	12.99	13.67	12.70***	**12.81**	13.67	13.07	12.61	**12.31**
SR	0.39	0.35	0.41	0.48	0.49	0.50	0.50	0.46	0.42	0.45	0.49
						$N=250$					
AV	7.29	NA	NA	4.56	3.24	4.61	4.89	5.12	4.95	4.39	4.44
SD	16.05	NA	NA	11.76	11.92	10.77***	**10.75**	12.88	11.85	10.67	**10.35**
SR	0.45	NA	NA	0.39	0.27	0.43	0.45	0.40	0.42	0.41	0.43
						$N=500$					
AV	7.03	NA	NA	4.98	4.05	4.81	4.88	5.30	4.94	4.42	4.41
SD	15.98	NA	NA	11.35	10.59	**10.23****	10.37	12.73	11.66	10.11	**9.90**
SR	0.44	NA	NA	0.44	0.38	0.47	0.47	0.42	0.42	0.43	0.44

AV, average; SD, standard deviation; SR, Sharpe ratio; NA, not available. All measures are based on 684 monthly out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 11
Performance measures for various estimators of the GMV portfolio

Period: December 13, 1965 to December 31, 2015

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv	SF	FF	POET	NL-SF
						$N=100$ portfolios formed on size and book-to-market					
AV	7.94	10.07	9.89	10.67	10.19	9.83	9.83	10.53	9.30	9.69	9.81
SD	15.98	9.09	8.98	8.77	8.43	**8.14*****	8.14	10.35	8.42	8.20	**7.89**
SR	0.50	1.11	1.10	1.22	1.21	1.21	1.21	1.02	1.10	1.18	1.24
						$N=100$ portfolios formed on size and operating profitability					
AV	7.60	10.45	10.15	8.90	10.01	9.35	9.35	9.00	8.99	9.20	9.40
SD	15.98	9.48	9.36	9.37	8.78	**8.30*****	8.30	10.45	8.57	8.26	**8.14**
SR	0.48	1.10	1.08	0.95	1.14	1.13	1.13	0.86	1.05	1.11	1.16
						$N=100$ portfolios formed on size and investment					
AV	8.02	10.92	10.69	9.45	10.38	10.23	10.23	10.62	9.69	10.65	10.42
SD	15.88	9.15	9.06	9.32	8.48	**8.05*****	8.05	10.23	8.44	8.11	**7.94**
SR	0.50	1.19	1.18	1.01	1.22	1.27	1.27	1.04	1.15	1.31	1.31

AV, average; SD, standard deviation; SR, Sharpe ratio. All measures are based on 12,600 daily out-of-sample returns in excess of the risk-free rate. In the rows labeled SD, the lowest number in each division appears in bold. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Assumption 4. Therefore, they can only differ in their eigenvalues, but not in their eigenvectors.

The specific data set chosen is roughly in the middle of the out-of-sample investment period[16] for an investment universe of size $N = 500$. Figure 2 displays the shrunk eigenvalues (that is, the eigenvalues of linear and nonlinear shrinkage) as a function of the sample eigenvalues (that is, the eigenvalues of the sample covariance matrix). For ease of interpretation, we also include the sample eigenvalues themselves as a function of the sample eigenvalues; this corresponds to the identity function (or the 45-degree line).

Linear shrinkage corresponds to a line that is less steep than the 45-degree line. Small sample eigenvalues are brought up whereas large sample eigenvalues are brought down; the cross-over point is roughly equal to five.

Nonlinear shrinkage also brings up small sample eigenvalues and brings down large sample eigenvalues; the cross-over point is also roughly equal to five. However the functional form is clearly nonlinear. Compared to linear shrinkage, the small eigenvalues are larger, the middle eigenvalues are smaller, and the large eigenvalues are about the same—with the exception of the top eigenvalue, which is larger.

This pattern is quite typical, though there are some other instances where even the middle and the large eigenvalues for nonlinear shrinkage are larger compared to linear shrinkage. What is generally true throughout is that the small eigenvalues as well as the top eigenvalue are larger for nonlinear shrinkage compared to linear shrinkage.

The financial intuition is that linear shrinkage overshrinks the market factor, resulting in insufficient efforts to diversify away market risk and reduce the portfolio beta. Also, linear shrinkage undershrinks the few dimensions that appear to be the safest in sample, resulting in an excessive concentration of money at this end. By contrast, nonlinear shrinkage makes better use of the diversification potential offered by the middle-ranking dimensions.

The quantity of shrinkage applied by the linear method is optimal only *on average* across the whole spectrum, so it can be sub-optimal in certain segments of the spectrum, and it takes the more sophisticated nonlinear correction to realize that.

4.7 Summary of results

We have carried out an extensive backtest analysis, evaluating the out-of-sample performance of our nonlinear shrinkage estimator when used to estimate the global minimum-variance portfolio; in this setting, the primary performance criterion is the standard deviation of the realized out-of-sample returns in excess of the risk-free rate. We have compared nonlinear shrinkage to a number of other strategies to estimate the global mininum-variance portfolio, most of

[16] Specifically, we use month number 250 out of the 480 months in the out-of-sample investment period.

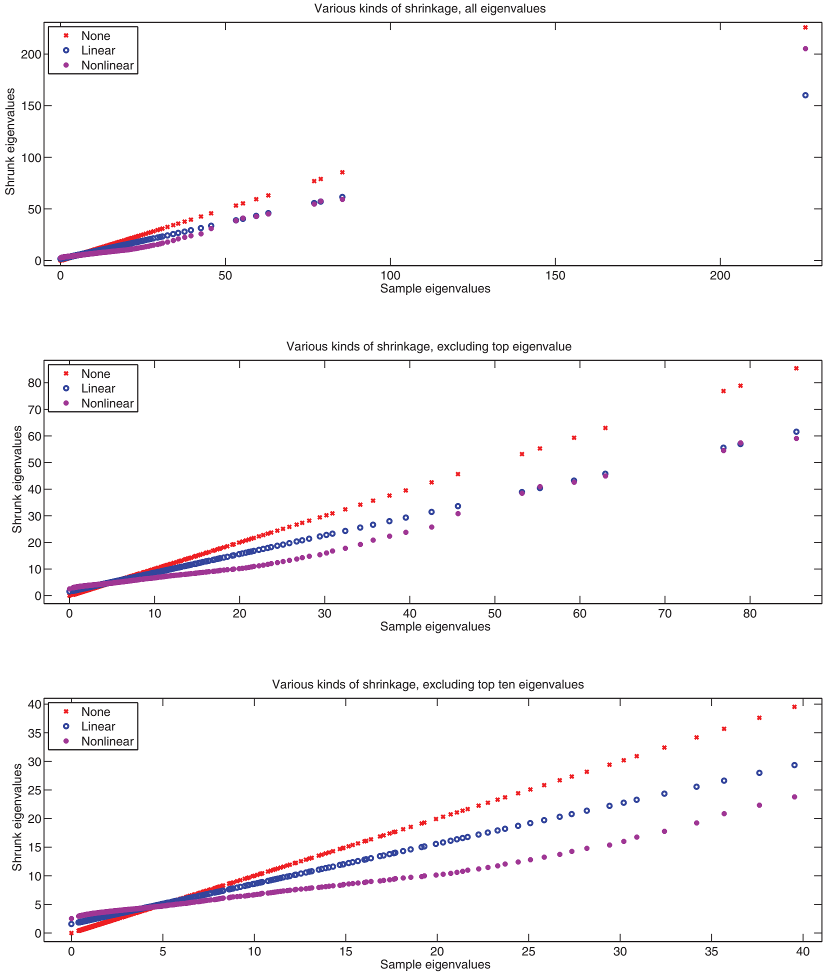

Figure 2
Shrunk eigenvalues as a function of sample eigenvalues for sample covariance matrix, linear shrinkage, and nonlinear shrinkage for an exemplary data set. The size of the investment universe is $N = 500$.

them proposed in the last decade in leading finance and econometrics journals. The portfolios considered can be classified into rotation-equivariant portfolios and portfolios based on factor models.

Our main analysis is based on daily data with an out-of-sample investment period ranging from 1973 through 2011. We have added a large number of robustness checks to study the sensitivity of our findings. Such robustness checks include a subsample analysis, changing the length of the estimation window of past data to estimate a covariance matrix, winsorization of past returns to estimate a covariance matrix, imposing a no-short-sales constraint, and changing the return frequency from daily data to monthly data (where the beginning of the out-of-sample investment period is moved back to 1955).

Among the rotation-equivariant portfolios, nonlinear shrinkage is the clear winner. Linear shrinkage and the gross-exposure constrained portfolio of Fan et al. (2012) have comparable performance and share second place. Last place is generally taken by the equal-weighted portfolio studied by DeMiguel et al. (2009). It is even outperformed by the sample covariance matrix, except when the number of assets is close to (or even equal to) the length of the estimation window. The "dominating" portfolio of Frahm and Memmel (2010) indeed dominates the sample covariance matrix but is generally outperformed by any of the other sophisticated estimators of the global mininum-variance portfolio.

A further improvement over nonlinear shrinkage can be obtained by applying nonlinear shrinkage after preconditioning the data using a single-factor model. This hybrid method also outperforms two other portfolios based on factor models, namely the (exact) three-factor model of Fama and French (1993) and the approximate factor model POET of Fan et al. (2013).

The statements of the two previous paragraphs only apply to unrestricted estimation of the global mininum-variance portfolio when short sales (i.e., negative portfolio weights) are allowed. Consistent with the findings of Jagannathan and Ma (2003), the relative performances change significantly when short sales are not allowed (i.e., when portfolio weights are constrained to be non-negative). In this case, among the rotation-equivariant portfolios, the sample covariance, linear shrinkage, and nonlinear shrinkage have comparable performance (with the sample covariance matrix actually being best by a very slim margin), whereas the equal-weighted portfolio continues to be worst. The portfolios based on factor models generally have a somewhat better performance compared to rotation-equivariant portfolios, with no clear winner among them.

Remark 9 (Alternative Nonlinear Shrinkage). Our nonlinear shrinkage estimator of the covariance matrix is based on a loss function that is tailor-made for portfolio selection; see Section 1. Though nobody could expect this a priori, the mathematical solution is identical to that from a totally different context: namely estimating a covariance matrix under a generic Frobenius-norm-based loss function as previously studied by Ledoit and Wolf (2012, 2015). Since the mathematical formulas for the optimal Markowitz portfolios (when short sales are allowed) actually require the inverse of the covariance matrix, it might appear more intuitive to use a direct estimator of the inverse of the covariance matrix rather than inverting an estimator of the covariance matrix itself. Direct nonlinear shrinkage estimation of the inverse covariance matrix under a generic Frobenius-norm-based loss function is studied by Ledoit and Wolf (2012, 2017). But it turns out that such an approach generally works less well, even though the differences are always small. This somewhat unexpected result demonstrates the potential value of basing the estimation of the covariance matrix on a loss function that is custom-tailored to the problem at hand (here, portfolio selection) rather than on a generic loss function. □

5. Conclusion

Despite its relative simplicity, Markowitz (1952) portfolio selection remains a cornerstone of finance, both for academic researches and fund managers. When applied in practice, it requires two inputs: an estimate of the vector of expected returns and an estimate of the covariance matrix of returns. The focus of this paper has been to address the second problem, having in mind a fund manager who already has a return predictive signal of his own choosing to address the first problem.

Compared with previous methods of estimating the covariance matrix, the key difference of our proposal lies in the number of free parameters to estimate. Let N denote the number of assets in the investment universe. Then previous proposals either estimate $O(1)$ free parameters—a prime example being linear shrinkage advocated by Ledoit and Wolf (2003, 2004a,b)—or estimate $O(N^2)$ free parameters—the prime example being the sample covariance matrix. We take the stance that in a large-dimensional framework, where the number of assets is of the same magnitude as the sample size, $O(1)$ free parameters are not enough, while $O(N^2)$ free parameters are too many. Instead, we have argued that "just the right number"—i.e., the *Goldilocks* principle—is $O(N)$ free parameters.

Our theoretical analysis is based on a stylized version of the Markowitz (1952) portfolio-selection problem under large-dimensional asymptotics, where the number of assets tends to infinity together with the sample size. We derive an estimator of the covariance matrix that is asymptotically optimal in a class of *rotation-equivariant* estimators. Such estimators do not use any a priori information about the orientation of the eigenvectors of the true covariance matrix. In particular, such estimators retain the eigenvectors of the sample covariance matrix but use different eigenvalues. Our contribution has been to work out the asymptotically optimal transformation of the sample eigenvalues to the eigenvalues used by the new estimator of the covariance matrix for the purpose of portfolio selection. We call this transformation *nonlinear shrinkage*.

Having established theoretical optimality properties under a stylized setting, we then put the new estimator to the practical test on historical stock return data. Running backtest exercises for the global minimum-variance portfolio and for a full Markowitz portfolio with a signal,[17] we have found that nonlinear shrinkage outperforms previously suggested estimators and, in particular, that it dominates linear shrinkage.

[17] The results for the latter portfolio are relegated to Appendix D.

Furthermore, we have studied combining nonlinear shrinkage with a simple one-factor model of stock returns. This hybrid approach results in an additional improvement in terms of reducing the out-of-sample volatility of portfolio returns.

Directions for future research include, among others, taking into account dependency across time, such as ARCH/GARCH effects, and a more systematic investigation of non-rotation-equivariant situations for which certain directions in the space of asset returns are privileged.

References

Ahn, D.-H., J. Conrad, and R. F. Dittmar. 2009. Basis assets. *Review of Financial Studies* 22:5133–74.

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

Bai, Z. D., and J. W. Silverstein. 1998. No eigenvalues outside the suppport of the limiting spectral distribution of large-dimensional random matrices. *Annals of Probability* 26:316–45.

———. 2010. *Spectral analysis of large-dimensional Random Matrices*, second edition. New York: Springer.

Bekaert, G., R. J. Hodrick, and X. Zhang. 2009. International stock return comovements. *Journal of Finance* 64:2591–626.

Bell, D. R., O. Ledoit, and M. Wolf. 2014. A new portfolio formation approach to mispricing of marketing performance indicators: An application to customer satisfaction. *Customer Needs and Solutions* 1:263–76.

Bickel, P. J., and E. Levina. 2008. Covariance regularization by thresholding. *Annals of Statistics* 36: 2577–604.

Bodnar, T., A. K. Gupta, and N. Parolya. 2014. Optimal linear shrinkage estimator for large dimensional precision matrix. *Contributions in infinite-dimensional statistics and related topics*, pp. 55–60. Eds. E. G. Bongiorno, A. Goia, E. Salinelli, and P. Vieu. Bologna, Italy: Società Editrice Esculapio.

Brandt, M. W., P. Santa-Clara, and R. Valkanov. 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies* 22:3411–47.

Cai, T., and W. Liu. 2011. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* 106:672–84.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Capon, J. 1969. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE* 57: 1408–18.

Chincarini, L. B., and D. Kim. 2006. *Quantitative equity portfolio management: An active approach to portfolio construction and management*. New York: McGraw-Hill.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal. 2009. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science 55*:798–812.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *Review of Financial Studies* 22:1915–53.

DeMiguel, V., A. Martin-Utrera, and F. J. Nogales. 2013. Size matters: Optimal calibration of shrinkage estimators for for portfolio selection. *Journal of Banking & Finance* 37:3018–34.

Du, L., J. Li, and P. Stoica. 2010. Fully automatic computation of diagonal loading levels for robust adaptive beamforming. *Aerospace and Electronic Systems, IEEE Transactions* 46:449–58.

Engle, R. F., and R. Colacito. 2006. Testing and valuing dynamic correlations for asset allocation. *Journal of Business and Economic Statistics* 24:238–53.

Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.

Fan, J., Y. Liao, and M. Mincheva. 2013. Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* 75:603–80.

Fan, J., J. Zhang, and K. Yu. 2012. Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association* 107:592–606.

Frahm, G. and C. Memmel. 2010. Dominating estimators for minimum-variance portfolios. *Journal of Econometrics* 159:289–302.

Green, J., J. R. M. Hand, and X. F. Zhang. 2013. The supraview of return predictive signals. *Review of Accounting Studies* 18:692–730.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Haugen, R. A. and N. L. Baker. 1991. The efficient market inefficiency of capitalization-weighted stock portfolios. *Journal of Portfolio Management* 17:35–40.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 8:650–705.

Huang, C., and R. Litzenberger. 1988. *Foundations for financial economics*. New York: North-Holland.

Ilmanen, A. 2011. *Expected returns: An investor's guide to harvesting market rewards*. New York: Wiley Finance.

IPCC. 2001. Climate change 2001: The scientific basis. In *Contribution of working group I to the third assessment report of the intergovernmental panel on climate change*. Eds. J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der LINDEN, X. Dai, K. Maskell, and C. Johnson. Cambridge: Cambridge University Press.

IPCC. 2007. Climate change 2007: The scientific basis. In *Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Eds. S. Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. Tignor, H. L. Miller, and Z. Chen. Cambridge: Cambridge University Press.

Jagannathan, R., and T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 54:1651–84.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.

Johnstone, I. M. 2001. On the distribution of the largest eigenvalue in principal component analysis. *Annals of Statistics* 29:295–327.

Kan, R., and D. R. Smith. 2008. The distribution of the sample mininum-variance frontier. *Management Science* 54:1364–80.

Kan, R., and G. Zhou. 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42: 621–56.

Ledoit, O., and S. Péché. 2011. Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields* 150:233–64.

Ledoit, O., and M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10:603–21.

———. 2004a. Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management* 30:110–19.

———. 2004b. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88:365–411.

———. 2008. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance* 15:850–59.

———. 2011. Robust performance hypothesis testing with the variance. *Wilmott Magazine* September:86–89.

———. 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics* 40:1024–60.

———. 2015. Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis* 139:360–84.

———. 2017. Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Working Paper ECON 122, Department of Economics, University of Zurich.

Marčenko, V. A., and L. A. Pastur. 1967. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics* 1:457–83.

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

McLean, R. D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–32.

Mestre, X. 2008. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory* 54:5113–29.

Michaud, R. 1989. The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal* 45:31–42.

Nielsen, F., and R. Aylursubramanian. 2008. Far from the madding crowd—Volatility efficient indices. Research insights, MSCI Barra.

Ribes, A., J.-M. Azaïs, and S. Planton. 2009. Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance matrix estimate. *Climate Dynamics* 33:707–22.

Stein, C. 1975. Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, GA.

———. 1986. Lectures on the theory of estimation of many parameters. *Journal of Mathematical Sciences* 34:1373–403.

Stieltjes, T. J. 1894. Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse* 8:J1–J122.

Subrahmanyam, A. 2010. The cross-section of expected stock returns: What have we learnt from the past twenty-five years of research? *European Financial Management* 16:27–42.

Tu, J., and G. Zhou. 2011. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99:204–15.

Wang, C., G. Pan, T. Tong, and L. Zhu. 2015. Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statistica Sinica* 25:993–1008.

Webster, K., Y. Luo, M.-A. Alvarez, J. Jussa, S. Wang, G. Rohal, A. Wang, D. Elledge, and G. Zhao. 2015. A portfolio manager's guidebook to trade execution: From light rays to dark pools. In *Quantitative strategy: Signal processing*. New York: Deutsche Bank Markets Research.

Won, J.-H., J. Lim, S.-J. Kim, and B. Rajaratnam. 2013. Condition-number regularized covariance estimation. *Journal of the Royal Statistical Society, Series B* 75:427–50.

Yoshimoto, A. 1996. The mean-variance approach to portfolio optimization subject to transaction costs. *Journal of the Operations Research Society of Japan* 39:99–117.