



University of  
Zurich<sup>UZH</sup>

University of Zurich  
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 137

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

Olivier Ledoit and Michael Wolf

Revised version, December 2014

---

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

Olivier Ledoit  
Department of Economics  
University of Zurich  
CH-8032 Zurich, Switzerland  
olivier.ledoit@econ.uzh.ch

Michael Wolf  
Department of Economics  
University of Zurich  
CH-8032 Zurich, Switzerland  
michael.wolf@econ.uzh.ch

December 2014

## Abstract

Markowitz (1952) portfolio selection requires estimates of (i) the vector of expected returns and (ii) the covariance matrix of returns. Many successful proposals to address the first estimation problem exist by now. This paper addresses the second estimation problem. We promote a nonlinear shrinkage estimator of the covariance matrix that is more flexible than previous linear shrinkage estimators and has ‘just the right number’ of free parameters to estimate (that is, the *Goldilocks* principle). It turns out that this number is the same as the number of assets in the investment universe. Under certain high-level assumptions, we show that our nonlinear shrinkage estimator is asymptotically optimal for portfolio selection in the setting where the number of assets is of the same magnitude as the sample size. For example, this is the relevant setting for mutual fund managers who invest in a large universe of stocks. In addition to theoretical analysis, we study the real-life performance of our new estimator using backtest exercises on historical stock return data. We find that it performs better than previous proposals for portfolio selection from the literature and, in particular, that it dominates linear shrinkage.

KEY WORDS: Large-dimensional asymptotics, Markowitz portfolio selection,  
nonlinear shrinkage.

JEL CLASSIFICATION NOS: C13, G11.

# 1 Introduction

Markowitz (1952) portfolio selection requires estimates of (i) the vector of expected returns and (ii) the covariance matrix of returns. Green et al. (2013) list over 300 papers that have been written on the first estimation problem, so it is fair to say we have a useful collection of explanatory signals by now. 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  $T$  is typically of the same order of magnitude as  $N$ .<sup>1</sup> 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. (2009a, 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  $O(1)$ . 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 two-dimensional space spanned by the identity matrix and the sample covariance matrix. 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.<sup>2</sup>

The class of estimators we consider was introduced by Stein (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. Thus our first theoretical contribution is to devise a loss function that captures the objective of an investor or researcher using portfolio selection. Our second contribution is to prove that this loss function has a well-defined limit under large-dimensional asymptotics,

---

<sup>1</sup>Sometimes it can even be smaller: think of ten years of monthly observations on the constituents of the S&P 500. But let us just say it is of the same magnitude.

<sup>2</sup>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’.

that is, when the dimension  $N$  goes to infinity along with the sample size  $T$ , and to compute its limit in closed form. Our third 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 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 (Ledoit and Wolf, 2013a). 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, there is a technology called *beamforming* (Capon, 1969) that is essential for radars, wireless communication and other areas of signal processing, and is mathematically equivalent to portfolio selection. This 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 distribution of sample eigenvalues. The only way to make this covariance matrix estimator usable in practice, that is, *bona fide*, is to find an estimator for the unobservable function. Fortunately this has been done before (Ledoit and Wolf, 2012, 2013b), so the transition from oracle to *bona fide* is completely straightforward in our case and requires no extra work.

Our optimal nonlinear shrinkage estimator fares well on historical stock returns data. 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. (2009b). Due to the highly nonlinear nature of the optimal shrinkage formula derived in this paper, 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, given that  $N$  is large and potentially unbounded. We also demonstrate superior out-of-sample performance for portfolio strategies that target a certain exposure to an exogenously specified vector of expected returns. 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.<sup>3</sup>

The remainder of the paper is organized as follows. Section 2 derives the loss function tailored to portfolio selection. Section 3 finds the limit of this loss function under large-dimensional asymptotics. Section 4 finds a covariance matrix estimator that is asymptotically optimal with respect to the loss function defined in Section 2. Section 5 presents empirical findings supporting the usefulness of the proposed estimator. Section 6 concludes. The Appendix

---

<sup>3</sup>For example, see Bell et al. (2014) for such a test.

contains all tables and mathematical proofs.

## 2 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. It can be drawn from the four categories delineated in the wide-ranging survey by Subrahmanyam (2010):

1. informal Wall Street wisdom (such as “value-investing”),
2. theoretical motivation based on risk-return model variants,
3. behavioral biases or misreaction by cognitively challenged investors, and
4. frictions such as illiquidity or arbitrage constraints.

The author documents at least 50 variables that fall in these categories. McLean and Pontiff (2013) bring the tally up to 82 variables, and Green et al. (2013) to 333 variables. Further overviews are provided by Ilmanen (2011) and Harvey et al. (2013).

In the empirical application of Section 5, we derive  $m$  from the momentum factor of Jegadeesh and Titman (1993) for simplicity and reproducibility, but more sophisticated constructs are clearly possible. All  $m$  needs to have is some power to explain the cross-section of expected asset returns.

### 2.1 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  $\Sigma$  denote the  $N \times N$  population covariance matrix of asset returns; note that  $\Sigma$  is unobservable. Portfolio selection seeks to maximize the reward-to-risk ratio:

$$\max_w \frac{w' m}{\sqrt{w' \Sigma w}}, \quad (2.1)$$

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 (2.1) if and only if there exists a strictly positive scalar  $a$  such that

$$w = a \times \Sigma^{-1} m. \quad (2.2)$$

This claim can be easily seen from the first-order condition of (2.1). 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

$$w = \frac{b}{m' \Sigma^{-1} m} \times \Sigma^{-1} m. \quad (2.3)$$

Note that expression (2.3) 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  $b := \sqrt{m'm}$ , which yields the scale-free solution

$$w = \frac{\sqrt{m'm}}{m'\Sigma^{-1}m} \times \Sigma^{-1}m .$$

In practice, the covariance matrix  $\Sigma$  is not known and needs to be 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

$$\widehat{w} := \frac{\sqrt{m'm}}{m'\widehat{\Sigma}^{-1}m} \times \widehat{\Sigma}^{-1}m . \quad (2.4)$$

All investing takes place *out of sample* by necessity. Since the population covariance matrix  $\Sigma$  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 2.1.** *The loss function is*

$$\mathcal{L}(\widehat{\Sigma}, \Sigma, m) := \widehat{w}'\Sigma\widehat{w} = m'm \times \frac{m'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}m}{(m'\widehat{\Sigma}^{-1}m)^2} . \quad (2.5)$$

In terms of assessing the broad scientific usefulness of this line of research, it is worth pointing out that the loss function in Definition 2.1 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.

## 2.2 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 forecast of expected returns given the information available to her, so we use  $m$  to evaluate the numerator of the Sharpe ratio. From the weights in equation (2.4) we deduce the Sharpe ratio:

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

Since we do not optimize over the Euclidian norm of the vector  $m$ , which is a given, this objective is equivalent to minimizing the loss function of Definition 2.1.

### 2.3 Quadratic Objective Function

Yet another way to control the trade-off between risk and return is to maximize a quadratic objective function with risk aversion parameter  $\gamma$ , in the spirit of Markowitz (1952). As above, we use the cross-sectional return predictive signal  $m$  in lieu of first moment because it is the investor's expectation of future returns conditional on her information set. Plugging in the weights of equation (2.4) yields:

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

Since we do not optimize over the Euclidian norm of the vector  $m$ , which is a given, this objective is also equivalent to minimizing the loss function of Definition 2.1. This is further confirmation that the loss function  $\mathcal{L}(\widehat{\Sigma}, \Sigma, m)$  is the right quantity to look at in the context of portfolio selection.

## 3 Large-Dimensional Asymptotic Limit of the Loss Function

The framework defined in Assumptions 3.1–3.4 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 3.1 onwards focuses on finance-related issues. Some of the assumptions made in Section 2 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 3.1** (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  $T \rightarrow \infty$ , to a limit  $c \in (0, 1) \cup (1, +\infty)$  called the concentration. 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 2 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 also handles 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 reason, but the empirical results in Section 5 indicate our method works in practice even in this challenging case.

**Assumption 3.2** (Population Covariance Matrix).

- a) *The population covariance matrix  $\Sigma_T$  is a nonrandom symmetric positive-definite matrix of dimension  $N$ .*

- b) Let  $\boldsymbol{\tau}_T := (\tau_{T,1}, \dots, \tau_{T,N})'$  denote a system of eigenvalues of  $\Sigma_T$  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)  $\text{Supp}(H)$ , 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  $[\underline{h}, \bar{h}] \subset (0, \infty)$  that contains  $\text{Supp}(H_T)$  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 3.1. 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  $H$  that can be interpreted as the ‘signature’ of the population covariance matrix: it says what proportion of eigenvalues are big, small, etc. Item c) and d) are technical assumptions requiring the supports of  $H$  and  $H_T$  to be well behaved.

**Assumption 3.3** (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  $\Sigma_T$ . Neither  $\sqrt{\Sigma_T}$  nor  $X_T$  are observed on their own: only  $Y_T$  is observed.

If 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). While 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  $\Lambda_T$  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  $\Lambda_T := \text{Diag}(\boldsymbol{\lambda}_T)$  where  $\boldsymbol{\lambda}_T := (\lambda_{T,1}, \dots, \lambda_{T,N})'$ . We can assume without loss of generality that the sample eigenvalues are sorted in increasing order:  $\lambda_{T,1} \leq \lambda_{T,2} \leq \dots \leq \lambda_{T,N}$ . Correspondingly, the  $i$ th sample eigenvector is  $u_{T,i}$ , the  $i$ th column vector of  $U_T$ . The e.d.f. of the sample eigenvalues is given by:  $\forall x \in \mathbb{R}$ ,  $F_T(x) := N^{-1} \sum_{i=1}^N \mathbb{1}_{[\lambda_{T,i}, \infty)}(x)$ .

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 the comprehensive and authoritative monograph by Bai and Silverstein (2010). Under Assumptions 3.1–3.3, there exists a limiting sample spectral distribution  $F$  continuously differentiable on  $(0, +\infty)$  such that

$$\forall x \in (0, +\infty) \quad F_T(x) \xrightarrow{\text{a.s.}} F(x). \quad (3.1)$$

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 3.1–3.3 imply that the support of  $F$ ,  $\text{Supp}(F)$ , is the union of a finite number  $\kappa \geq 1$  of compact intervals  $\bigcup_{k=1}^{\kappa} [a_k, b_k]$ , where  $0 < a_1 < b_1 < \dots < a_{\kappa} < b_{\kappa} < \infty$ , with the addition of the singleton  $\{0\}$  in the case  $c > 1$ .

**Assumption 3.4** (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:  $\widehat{\Delta}_T := \text{Diag}(\widehat{\delta}_T(\lambda_{T,1}), \dots, \widehat{\delta}_T(\lambda_{T,N}))$ , 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  $\text{Supp}(F)$  and continuously differentiable such that  $\widehat{\delta}_T(x) \xrightarrow{\text{a.s.}} \widehat{\delta}(x)$ , for all  $x \in \text{Supp}(F)$ . 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 repack 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.

Just because 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  $\widehat{\delta}_T$  the *shrinkage function* 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  $\widehat{\Sigma}$ .

**Remark 3.1.** The linear shrinkage estimator of Ledoit and Wolf (2004b) also belongs to this class of rotation-equivariant estimators, with the shrinkage function given by

$$\widehat{\delta}_T(\lambda_{T,i}) := (1 - \widehat{k}) \cdot \lambda_{T,i} + \widehat{k} \cdot \bar{\lambda}_T \quad \text{where} \quad \bar{\lambda}_T := \frac{1}{N} \sum_{j=1}^N \lambda_{T,j} . \quad (3.2)$$

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

Estimators in the class defined by Assumption 3.4 are evaluated according to the limit as  $T$  and  $N$  go to infinity together (in the manner of Assumption 3.1) of the loss function defined in equation (2.5). For this limit to exist, some assumption on the return predictive signal is required.<sup>4</sup> The assumption that we make below is coherent with the rotation-equivariant framework that we have built so far.

**Assumption 3.5** (Return Predictive Signal).  *$m_T$  is distributed independently of  $S_T$ , and its distribution is rotation invariant.*

Rotation invariance means that the normalized return predictive signal  $m_T / \sqrt{m'_T m_T}$  is uniformly distributed on the unit sphere. This 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 3.4. The limit of the loss function defined in Section 2 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 3.1.** *Under Assumptions 3.1–3.5,*

$$m'_T m_T \times \frac{m'_T \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} m_T}{\left(m'_T \widehat{\Sigma}_T^{-1} m_T\right)^2} \xrightarrow{\text{a.s.}} \frac{\sum_{k=1}^{\kappa} \int_{a_k}^{b_k} \frac{dF(x)}{x |s(x)|^2 \widehat{\delta}(x)^2} + \mathbb{1}_{\{c>1\}} \frac{1}{c s(0) \widehat{\delta}(0)^2}}{\left[ \int \frac{dF(x)}{\widehat{\delta}(x)} \right]^2} , \quad (3.3)$$

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

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

---

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

All proofs are in Appendix B. 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 B.1 for more specific mathematical details). By contrast, equation (3.3) is one of the major mathematical innovations of this paper.

## 4 Optimal Covariance Matrix Estimator for Portfolio Selection

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

### 4.1 Oracle Estimator

Equation (3.3) 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 4.1.** *Define the oracle shrinkage function*

$$\forall x \in \text{Supp}(F) \quad d^*(x) := \begin{cases} \frac{1}{x |s(x)|^2} & \text{if } x > 0, \\ \frac{1}{(c-1)s(0)} & \text{if } c > 1 \text{ and } x = 0, \end{cases} \quad (4.1)$$

where  $s(x)$  is the complex-valued Stieltjes transform defined by equation (MP). If Assumptions 3.1–3.5 are satisfied then the following statements hold true:

(a) *The oracle estimator of the covariance matrix*

$$S_T^* := U_T D_T^* U_T' \quad \text{where} \quad D_T^* := \text{Diag}(d^*(\lambda_{T,1}), \dots, d^*(\lambda_{T,N})) \quad (4.2)$$

*minimizes in the class of rotation-equivariant estimators defined in Assumption 3.4 the almost sure limit of the portfolio selection loss function introduced in Section 2*

$$\mathcal{L}_T(\widehat{\Sigma}_T, \Sigma_T, m_T) := m_T' m_T \frac{m_T' \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} m_T}{(m_T' \widehat{\Sigma}_T^{-1} m_T)^2}, \quad (4.3)$$

*as  $T$  and  $N$  go to infinity together in the manner of Assumption 3.1.*

(b) *Conversely, any covariance matrix estimator  $\widehat{\Sigma}_T$  that minimize 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  $\alpha$ .*

$S_T^*$  is an oracle estimator because it depends on  $c$  and  $s(x)$ , which are unobservable. Nonetheless, deriving a *bona fide* counterpart to  $S_T^*$  will be easy because estimators for  $c$  and  $s(x)$  are readily available, as demonstrated in Section 4.2 below, therefore we shall keep our attention on  $S_T^*$  for the time being.

As is apparent from part (b) of Theorem 4.1, minimizing the loss function characterizes the optimal shrinkage formula only up to an arbitrary positive scaling factor  $\alpha$ . This is inherent to the problem of portfolio selection: equation (2.3) shows that two covariance matrices that only differ by the scaling factor  $\alpha$  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 4.1.** *Let Assumptions 3.1–3.5 hold, and let the limiting shrinkage function of the estimator  $\widehat{\Sigma}_T$  be of the form  $\widehat{\delta} = \alpha d^*$  for some positive constant  $\alpha$  as per item (b) of Theorem 4.1. Then,*

$$\frac{1}{N} \text{Tr}(\widehat{\Sigma}_T) - \frac{1}{N} \text{Tr}(\Sigma_T) \xrightarrow{\text{a.s.}} 0 \quad (4.4)$$

*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  $S_T^*$ .

It can be noticed that 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 3.5. In subsequent Monte Carlo simulations (Sections 5.3 and 5.4) 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, as the optimal shrinkage formula  $d^*$  is not tied up to any one choice of  $m_T$ .

One of the basic features of the optimal nonlinear transformation  $d^*$  is that it preserves the grand mean of the eigenvalues, as evidenced by Proposition 4.1. 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 P ech e (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  $u'_{T,i} \Sigma_T u_{T,i}$ ; and the latter shows that  $(u'_{T,i} \Sigma_T u_{T,i})_{i=1,\dots,N}$  are less dispersed than  $(\lambda_{T,i})_{i=1,\dots,N}$ . 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 4.1 can be gathered by comparing equation (4.1) with the two shrinkage functions obtained earlier by Ledoit and Wolf (2012). These authors used a generic loss function renowned for its tractability, based on the *Frobenius* norm. The Frobenius norm of a quadratic matrix  $A$  is defined as  $\|A\|_F := \sqrt{\text{Tr}(AA')}$ , so it is essentially quadratic in nature. Ledoit and Wolf (2012) used a Frobenius-norm-based loss function in two

different ways, once with the covariance matrix and then again with its inverse:

$$\mathcal{L}_T^1(\widehat{\Sigma}_T, \Sigma_T) := \frac{1}{N} \left\| \widehat{\Sigma}_T - \Sigma_T \right\|_F^2 \quad \text{and} \quad (4.5)$$

$$\mathcal{L}_T^{-1}(\widehat{\Sigma}_T, \Sigma_T) := \frac{1}{N} \left\| \widehat{\Sigma}_T^{-1} - \Sigma_T^{-1} \right\|_F^2, \quad (4.6)$$

leading to two different optimal shrinkage formulas.

The first unexpected result is that equation (4.1) 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 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, among the two shrinkage functions of Ledoit and Wolf (2012),  $d^*$  is equal to the ‘wrong’ one. Indeed, equation (2.3) makes it clear that Markowitz (1952) portfolio selection involves not the covariance matrix itself but its *inverse*. Thus we might have expected that  $d^*$  is equal to the shrinkage function 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:  $d^*$  is optimal with respect to  $\mathcal{L}_T^1$  instead. This could not have been anticipated without the analytical developments achieved in Theorems 3.1 and 4.1. In particular, there have been several papers recently concerned with the ‘direct’ estimation of the inverse covariance matrix<sup>5</sup> using a loss function of the type (4.6), with Markowitz (1952) portfolio selection listed as a major motivation; for example, see Frahm and Memmel (2010), Wang et al. (2012), and Bodnar et al. (2014). But our result shows that, unexpectedly, this approach is suboptimal in the context of portfolio selection.

In the end, the best way to gain comfort with this mathematical result may be to seek an economic interpretation of it. Starting from equation (2.2), we can express the vector of optimal portfolio weights as

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

where  $a_T$  is a suitably chosen scalar coefficient, and the last approximation comes from Theorem 1.4 of Ledoit and P ech e (2011) as mentioned earlier. Thus, the mean-variance efficient portfolio can be decomposed into a linear combination of sample eigenvector portfolios, with the  $i$ th sample eigenvector portfolio assigned a weight approximately proportional to  $u_{T,i}' m_T / (u_{T,i}' \Sigma_T u_{T,i})$ . This weighting scheme is intuitively appealing because it represents the out-of-sample reward-to-risk ratio of the  $i$ th sample eigenvector portfolio. Thus we can be confident that the proposed nonlinear shrinkage formula makes sense economically.

<sup>5</sup>Also called the *precision matrix* in the statistics literature.

## 4.2 *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, 2013b). 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 pushed to Appendix C. 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 4.1.** *Suppose that Assumptions 3.1–3.5 are satisfied, and let  $\widehat{s}(x)$  denote the estimator of the Stieltjes transform  $s(x)$  introduced by Ledoit and Wolf (2013b) and reproduced in Appendix C. Then the covariance matrix estimator*

$$\widehat{S}_T := U_T \widehat{D}_T U_T' \quad \text{where} \quad \widehat{D}_T := \text{Diag}(\widehat{d}_T(\lambda_{T,1}), \dots, \widehat{d}_T(\lambda_{T,N})) \quad (4.7)$$

$$\text{and } \forall i = 1, \dots, N \quad \widehat{d}_T(\lambda_{T,i}) := \begin{cases} \frac{1}{\lambda_{T,i} |\widehat{s}(\lambda_{T,i})|^2} & \text{if } \lambda_{T,i} > 0, \\ \frac{1}{(\frac{N}{T} - 1) \widehat{s}(0)} & \text{if } N > T \text{ and } \lambda_{T,i} = 0 \end{cases} \quad (4.8)$$

*minimizes in the class of rotation-equivariant estimators defined in Assumption 3.4 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 3.1.*

This corollary is given without proof as it is an immediate consequence of Theorem 4.1 above, via Ledoit and Wolf (2012, Proposition 4.3; 2013b, Theorem 2.2).

## 5 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, as in Section 2, 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.

### 5.1 Data and General Portfolio Formation Rules

We download daily data from the Center for Research in Security Prices (CRSP) starting in 01/01/1972 and ending in 12/31/2011. For simplicity, we adopt the common convention that 21 consecutive trading days constitute one ‘month’. The out-of-sample period ranges

from 01/19/1973 until 12/31/2011, resulting in a total of 480 ‘months’ (or 10,080 days). All portfolios are updated ‘monthly’.<sup>6</sup> 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 a complete return ‘history’ over the next 21 days.<sup>7</sup> 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.

## 5.2 Restriction to Rotation-Equivariant Covariance Matrix Estimators

In the estimation of the various Markowitz portfolios described below, attention is restricted to rotation-equivariant covariance matrix estimators. The reason for doing so is because the theoretical analysis of our paper is based on the rotation-equivariant framework, and also because it is the only way to compare apples to apples. Such a restriction rules out, among others, the diagonal of the sample covariance matrix as proposed by Schäfer and Strimmer (2005), factor models as proposed by Fan et al. (2008), or linear shrinkage to structured targets as proposed by Ledoit and Wolf (2003, 2004a).

To break rotation equivariance would require *a priori* beliefs about the orientation of the population eigenvectors — even though they are unobservable. It is generally true that *if* we have prior beliefs about some of the parameters that need to be estimated, and *if* we know how strong our beliefs are, and *if* we turn out to be right, then we are better off incorporating these beliefs into the construction of estimator. But that is a lot of *ifs*, as researchers who have been tempted by the Bayesian side well know.

There are potentially as many beliefs as there are researchers in the field. Some may think that the Fama and French (1992) three-factor model is the right one, others the Carhart (1997) four-factor model, the recent five-factor model of Fama and French (2014), the twelve BARRA<sup>TM</sup> style factors, or even 68 GICS<sup>®</sup> industry factors. The lack of consensus makes rotation equivariance the only scientifically neutral choice, and the one whose relevance is most universal.

Shrinkage and prior beliefs are not substitutes but complements. Prior beliefs typically concern a small number of dimensions. Something needs to be done on the remainder, which constitutes the bulk of the dimensions, where no prior belief can reasonably be intuited. For example, if we trade the Russell 3000 index, even with an industry factor model, there are still

---

<sup>6</sup>‘Monthly’ updating is common practice to avoid an unreasonable amount of turnover, and thus transaction costs.

<sup>7</sup>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.

2,932 undifferentiated dimensions. This is where shrinkage must be brought in. It should be possible to combine a low-dimensional factor model favored by the researcher with shrinkage in all the other dimensions, perhaps by some pre-conditioning method. This development, however, must be deferred to future research. It is logically necessary to solve the pure problem of optimal shrinkage in the rotation-equivariant case first, as we do in this paper, before we can even think of the best way to combine this solution with prior beliefs in a low-dimensional factor model.

Fortunately, there have been several recent proposals for the estimation of Markowitz portfolios that all fall in the rotation-equivariant framework, and we shall compare our linear shrinkage estimator to them in the remainder of the chapter.

### 5.3 Global Minimum Variance Portfolio

We first focus on the well-known problem of estimating the global minimum variance (GMV) portfolio, in the absence of short-sales constraints. The problem is formulated as

$$\min_w w' \Sigma w \tag{5.1}$$

$$\text{subject to } w' \mathbf{1} = 1, \tag{5.2}$$

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

$$w = \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}' \Sigma^{-1} \mathbf{1}}. \tag{5.3}$$

The natural strategy in practice is to replace the unknown  $\Sigma$  by an estimator  $\hat{\Sigma}$  in formula (5.3), yielding a feasible portfolio

$$\hat{w} := \frac{\hat{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}}. \tag{5.4}$$

Alternative strategies, motivated by estimating the optimal  $w$  of (5.3) ‘directly’, as opposed to ‘indirectly’ via the estimation of  $\Sigma$ , 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 (that is, 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 seven portfolios are included in the study.

- $\mathbf{1}/N$ : The equal-weighted portfolio promoted by DeMiguel et al. (2009b), among others. This portfolio can be viewed as a special case of (5.4) where  $\hat{\Sigma}$  is given by the  $N \times N$  identity matrix. This strategy avoids any parameter estimation whatsoever.



- **Sample:** The portfolio (5.4) 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  $\Sigma$  and set the gross-exposure constraint parameter equal to  $c = 2$ .
- **Lin:** The portfolio (5.4) where the matrix  $\widehat{\Sigma}$  is given by the linear shrinkage estimator of Ledoit and Wolf (2004b).
- **NonLin:** The portfolio (5.4) where  $\widehat{\Sigma}$  is given by the estimator  $\widehat{S}$  of Corollary 4.1.
- **NL-Inv:** The portfolio (5.4) where  $\widehat{\Sigma}^{-1}$  is given by the ‘direct’ nonlinear shrinkage estimator of  $\Sigma^{-1}$  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 (2013a).

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 (5.3). 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. A high out-of-sample average return, AV, and a high out-of-sample Sharpe ratio, SR, are naturally also desirable, but should be considered of secondary importance from the point of view of evaluating the quality of a covariance matrix estimator.

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 seven portfolios, there are 21 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).<sup>8</sup>

The results are presented in Table 1 and can be summarized as follows.

- The standard deviation of the true GMV portfolio (5.3) 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 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’ portfolios.
- The performance of FZY and Lin is comparable.
- NonLin has the uniformly best performance 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  $N = 50, 100, 250, 500$ . The outperformance is also economically meaningful for  $N = 100, 250, 500$ , ranging from 0.3 to 0.6 percentage points. Furthermore, NonLin always outperforms Lin in terms of the Sharp ratio also.
- NonLin also outperforms NL-Inv, though the differences are always small.

Summing up, in the global minimum variance portfolio problem, NonLin dominates the remaining six portfolios in terms of the standard deviation and, in addition, dominates Lin in terms of the Sharpe ratio.

## 5.4 Markowitz Portfolio with Momentum Signal

We now turn attention to a ‘full’ Markowitz portfolio with a signal.

As discussed at the beginning of Section 2, by now a large number of variables have been documented that can be used to construct a signal in practice. For simplicity and reproducibility, we use the well-known momentum factor (or simply momentum for short) of Jegadeesh and Titman (1993). For a given period investment period  $h$  and a given stock, the momentum is the the geometric average of the previous 12 ‘monthly’ returns on the stock but excluding the most recent ‘month’. Collecting the individual momentums of all the  $N$  stocks contained in the portfolio universe yields the return predictive signal  $m$ .

---

<sup>8</sup>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.

In the absence of short-sales constraints, the investment problem is formulated as

$$\min_w w' \Sigma w \quad (5.5)$$

$$\text{subject to } w'm = b \quad \text{and} \quad w' \mathbf{1} = 1, \quad (5.6)$$

where  $b$  is a selected target expected return. The analytical solution of the problem is given in Sections 3.8 and 3.9 of the textbook by Huang and Litzenberger (1988). The natural strategy in practice is to replace the unknown  $\Sigma$  by an estimator  $\widehat{\Sigma}$ , yielding a feasible portfolio

$$\widehat{w} := \frac{Cb - A}{BC - A^2} \widehat{\Sigma}^{-1} m + \frac{B - Ab}{BC - A^2} \widehat{\Sigma}^{-1} \mathbf{1}, \quad (5.7)$$

$$\text{where } A := m' \widehat{\Sigma}^{-1} \mathbf{1}, \quad B := m' \widehat{\Sigma}^{-1} m, \quad \text{and} \quad C := \mathbf{1}' \widehat{\Sigma}^{-1} \mathbf{1}. \quad (5.8)$$

The following eight portfolios are included in the study.

- **EW-TQ:** The equal-weighted portfolio of the top-quintile stocks according to momentum  $m$ . This strategy does not make use of the momentum signal beyond sorting of the stocks in quintiles.

The value of the target expected return  $b$  for portfolios listed below is then given by the arithmetic average of the momentums of the stocks included in this portfolio (that is, the expected return of EW-TQ according to the signal  $m$ ).

- **BSV:** The portfolio (5.7)–(5.8) where  $\widehat{\Sigma}$  is given by the identity matrix of dimension  $N \times N$ . This portfolio corresponds to the proposal of Brandt et al. (2009).
- **Sample:** The portfolio (5.7)–(5.8) 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.
- **KZ:** The three-fund portfolio described by equation (68) of Kan and Zhou (2007); note that this portfolio is not available when  $N \geq T - 4$ .
- **TZ:** The three-fund portfolio KZ combined with the equal-weighted portfolio as proposed in Section 2.3 of Tu and Zhou (2011); note that this portfolio is not available when  $N \geq T - 4$ .
- **Lin:** The portfolio (5.7)–(5.8) where  $\widehat{\Sigma}$  is given by the linear shrinkage estimator of Ledoit and Wolf (2004b).
- **NonLin:** The portfolio (5.7)–(5.8) where  $\widehat{\Sigma}$  is given by the estimator  $\widehat{S}$  of Corollary 4.1.
- **NL-Inv:** The portfolio (5.7)–(5.8) where  $\widehat{\Sigma}^{-1}$  is given by the ‘direct’ nonlinear shrinkage estimator of  $\Sigma^{-1}$  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 (2013a).

Our stance is that in the context of a ‘full’ Markowitz portfolio, the most important performance measure is the out-of-sample Sharpe ratio, SR. In the ‘ideal’ investment problem (5.5)–(5.5), minimizing the variance (for a fixed target expected return  $b$ ) is equivalent to maximizing the Sharpe ratio (for a fixed target expected return  $b$ ). In practice, because of

estimation error in the signal, the various strategies do not have the same expected return and, thus, focusing on the out-of-sample standard deviation is inappropriate.

We also consider the question whether one portfolio delivers a higher out-of-sample Sharpe ratio than another portfolio at a level that is statistically significant. Since we consider eight portfolios, there are 28 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 Sharpe ratios is obtained by the prewhitened  $HAC_{PW}$  method described in Ledoit and Wolf (2008, Section 3.1).<sup>9</sup>

The results are presented in Table 2 and can be summarized as follows.

- We again observe that Sample breaks down for  $N = 250$ , when the sample covariance matrix is close to singular.
- KZ and TZ have the lowest Sharpe ratios throughout and some of the numbers are even negative.
- The overall order, from worst to best, of the remaining five portfolios is EW-TQ, BSV, Lin, NL-Inv, and NonLin.
- NonLin has the uniformly best performance and the outperformance over Lin is statistically significant at the 0.05 level for  $N = 250, 500$ . The outperformance is also economically meaningful for  $N = 100, 250, 500$ , ranging from 0.04 to 0.15.

Summing up, in a ‘full’ Markowitz problem with momentum signal, NonLin dominates the remaining seven portfolios in terms of the Sharpe ratio.

**Remark 5.1.** It should be pointed out that for all shrinkage estimators of the covariance matrix (that is, Lin, NonLin, and NL-Inv), the Sharpe ratios here are higher compared to the GMV portfolios for all  $N$ . Therefore, using a return predictive signal can really pay off, if done properly. ■

## 5.5 Analysis of Weights

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

- **Min:** Minimum weight.
- **Max:** Maximum weight.
- **SD:** Standard deviation of weights.

---

<sup>9</sup>Since the out-of-sample size is very large at 10,080, there is no need to use the computationally more expensive bootstrap method described in Ledoit and Wolf (2008, Section 3.2), which is preferred for small sample sizes.

- **MAD-EW:** Mean absolute deviation from equal-weighted portfolio computed as

$$\frac{1}{N} \sum_{i=1}^N \left| \hat{w}_i - \frac{1}{N} \right|.$$

For each characteristic, we then report the average outcome over the 480 portfolio formations (that is, over the 480 ‘months’).

The results for the global minimum variance portfolio are displayed in Table 3. Not surprisingly, the most dispersed weights 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  $N = 100, 250, 500$ .

The results for the Markowitz portfolio with momentum signal are displayed in Table 4. Not surprisingly, the most dispersed weights are found for Sample, followed by three shrinkage methods, EW-TQ, and BSV. The least dispersed weights are always found for KZ and TZ, which is owed to the fact that these two portfolios are not fully invested in the  $N$  stocks but also invest (generally to a large extent) in the risk-free rate. NonLin and NL-Inv are comparably dispersed to Lin for  $N = 30, 50$  but less dispersed than Lin for  $N = 100, 250, 500$ .

## 5.6 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.

### 5.6.1 Subperiod Analysis

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

The results for the global minimum variance portfolio are presented in Tables 5–7. It can be seen that 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  $N \geq 50$ .

The results for the Markowitz portfolio with momentum signal are presented in Tables 8–10. It can be seen that NonLin is better than Lin in terms of the Sharpe ratio in 14 out of the 15 cases; though statistical significance only obtains in the first subperiod for  $N = 250, 500$ .

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

### 5.6.2 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.

The results for global minimum variance portfolio are presented in Table 11. It can be seen that they are similar to the results in Table 1. In particular, NonLin has the uniformly best performance in terms of the standard deviation and the outperformance over Lin is statistically significant for  $N = 100, 250, 500$ .

The results for the Markowitz portfolio with momentum signal are presented in Table 12. It can be seen that they are similar to the results in Table 2. In particular, NonLin has the uniformly best performance in terms of the Sharpe ratio, though the outperformance over Lin is not statistically significant.

### 5.6.3 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 D. Of course, we always use the ‘raw’, non-winsorized data in computing the out-of-sample portfolio returns.

The results for the global minimum variance portfolio are presented in Table 13. 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 somewhat worse. Again, NonLin has the uniformly best performance.

The results for the Markowitz portfolio with momentum signal are presented in Table 14. While BSV has the best performance for  $N = 30$ , NonLin has the best performance for  $N = 50, 100, 250, 500$ . In particular, NonLin always outperforms Lin, though no longer with statistical significance.

### 5.6.4 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.

The results for the global minimum variance portfolio are presented in Table 15. 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 in terms of the standard deviation, although the differences to Lin, NonLin, and NL-Shrink are always small. Furthermore, the results for Sample are always better compared allowing short sales whereas the results for the Lin, NonLin, and NL-Shrink are always worse. 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.

The results for Markowitz portfolio with momentum signal are presented in Table 16. Note that Sample is now available for all  $N$ , whereas KZ and TZ are not available at all. In contrast

to the previous results for the global minimum variance portfolio, improved estimation of the covariance matrix still pays off, even if to a lesser extent compared to allowing short sales. In particular Lin, NonLin, and NL-Inv improve upon Sample in terms of the Sharpe ratio for all  $N$ . While BSV has the best performance for  $N = 30$ , NonLin has the best performance for  $N = 50, 100, 250, 500$ . In particular, NonLin always outperforms Lin, though no longer with statistical significance.

### 5.6.5 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 starting at 01/1945 until 12/2011. We use the  $T = 120$  most recent months of previous data to estimate a covariance matrix. Consequently, the out-of-sample investment period ranges from 01/1955 until 12/2011, yielding 684 out-of-sample returns. The remaining details are as before.

The results for the global minimum variance portfolio are presented in Table 17. 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  $N = 50, 100, 250, 500$ .

The results for the Markowitz portfolio with momentum signal are presented in Table 18. It can be seen that the relative performance of the various portfolios is similar to that in Table 2. 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; the outperformance is statistically significant for all  $N$ .

## 5.7 Illustration of Nonlinear vs. 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 Assumption 3.4. Therefore, they can only differ in their eigenvalues, but not in their eigenvectors.

The specific data chosen is roughly in the middle of the out-of-sample investment period<sup>10</sup> for an investment universe of size  $N = 500$ . Figure 1 displays the shrunk eigenvalues (that is, the eigenvalues of linear and nonlinear shrinkage) as 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.

<sup>10</sup>Specifically, use ‘month’ number 250 out of the 480 ‘months’ in the out-of-sample investment period.

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 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.

## 5.8 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 (a) the global minimum variance portfolio and (b) a full Markowitz portfolio with momentum signal. We have compared nonlinear shrinkage to a number of other strategies to estimate Markowitz portfolios, most of them proposed in the last decade in leading finance and econometrics journals.

Our main analysis is based on daily data with an out-of-sample investment period ranging from 1973 throughout 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).

Overall, nonlinear shrinkage is the clear winner in both applications.

In the context of estimating the global minimum variance portfolio, where the performance criterion is the standard deviation of the realized out-of-sample returns in excess of the risk-free rate, 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. (2009b). 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 minimum variance portfolio.

The statements of the previous paragraph only apply to ‘unrestricted’ estimation of the global minimum variance portfolio when short sales (that is, 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 (that is, when portfolio weights are constrained to be non-negative). In this case, the sample covariance, linear shrinkage, and nonlinear shrinkage have comparable performance (with the sample covariance matrix actually being best by a very slim margin), while the equal-weighted portfolio continues to be worst.

In the context of estimating a full Markowitz portfolio with momentum signal, where the performance criterion is the Sharpe ratio of the realized out-of-sample returns, linear shrinkage comes in second place, followed the proposal of Brandt et al. (2009) which, in this context, corresponds to ‘estimating’ the covariance matrix by the identity matrix. Fourth place is taken by the approach of simply investing equally in the top quintile of stocks after ranking them according to momentum. The two last places are taken by the three-fund rule of Kan and Zhou (2007) and the combination of this rule with the equally-weighted portfolio as proposed by Tu and Zhou (2011).

The statements of the previous paragraph are for ‘unrestricted’ estimation of the Markowitz portfolio with momentum signal when short sales (that is, negative portfolio weights) are allowed. They generally continue to hold, in terms of the relative ranking, when short sales are not allowed (that is, when portfolio weights are constrained to be non-negative), though the absolute differences diminish. The only difference is that the sample covariance matrix now outperforms the identity covariance matrix and equally investing in the top-quintile portfolio when the number of assets is comparable to or even larger than the estimation window used to estimate a covariance matrix. Nevertheless, both linear and nonlinear shrinkage continue to always outperform the sample covariance matrix.

**Remark 5.2** (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 2. Though nobody could expect this *a priori*, the mathematical solution turns out to be one-to-one the same as in 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, 2013b). 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 covariance matrix under a generic Frobenius-norm-based loss function is studied by Ledoit and Wolf (2012, 2013a). 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 ■

**Remark 5.3** (Optimal Versus Naïve Diversification). DeMiguel et al. (2009b) claim that it is very difficult to outperform the ‘naïve’ equally-weighted portfolio with ‘sophisticated’ Markowitz portfolios due to the estimation error in the inputs required by Markowitz portfolios; their claim is concerning outperformance in terms of the Sharpe ratio and certainty equivalents. In contrast, we find that shrinkage estimation of the covariance matrix combined with the momentum signal results in consistently higher Sharpe ratios compared to the equally-weighted portfolio.<sup>11</sup> In particular, the outperformance also holds in recent times when the momentum signal was not as strong anymore as in the more distant past; see Tables 7 and 10. This finding is encouraging to ‘sophisticated’ investment managers: If they can come up with a good signal and combine it with nonlinear shrinkage estimation of the covariance matrix, outperforming the equally-weighted portfolio is anything but a hopeless task. ■

## 6 Conclusion

Despite its relative simplicity, Markowitz (1952) portfolio selection remains a cornerstone of finance, both for researchers and fund managers. When applied in practice, it requires two inputs: (i) an estimate of the vector of expected returns and (ii) 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 (for which end there exists a large literature already).

Compared to 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’ (that is, the *Goldilocks* principle) is  $O(N)$  free parameters.

Our theoretical analysis is based on a stylized version of the Markowitz (1952) 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

---

<sup>11</sup>Of the many scenarios considered, there is a single one where the equally-weighted portfolio has a higher Sharpe ratio than linear shrinkage combined with the momentum signal, namely with monthly data for  $N = 30$ ; see Tables 17 and 18. On the other hand, the equally-weighted portfolio has always a lower Sharpe ratio than nonlinear shrinkage combined with the momentum signal.

purpose of portfolio selection. We term 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 (a) for the global minimum variance portfolio problem and (b) for a ‘full’ Markowitz problem with a signal, we have found that nonlinear shrinkage outperforms previously suggested estimators and, in particular, dominates linear shrinkage.

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

## References

- Bai, Z. D. and Silverstein, J. W. (2010). *Spectral Analysis of Large-Dimensional Random Matrices*. Springer, New York, second edition.
- Bell, D. R., Ledoit, O., and Wolf, M. (2014). A new portfolio formation approach to mispricing of marketing performance indicators: An application to customer satisfaction. *Customer Needs and Solutions*, 1:263–276.
- Bodnar, T., Gupta, A. K., and Parolya, N. (2014). Optimal linear shrinkage estimator for large-dimensional precision matrix. Preprint arXiv:1308.0931v3.
- Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, 22(9):3411–3447.
- Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418.
- Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.
- Chincarini, L. B. and Kim, D. (2006). *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management*. McGraw-Hill, New York.
- DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the  $1/N$  portfolio strategy? *Review of Financial Studies*, 22:1915–1953.
- DeMiguel, V., Martin-Utrera, A., and Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37:3018–3034.

- Du, L., Li, J., and Stoica, P. (2010). Fully automatic computation of diagonal loading levels for robust adaptive beamforming. *Aerospace and Electronic Systems, IEEE Transactions on*, 46(1):449–458.
- Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.
- Fama, E. F. and French, K. R. (2014). A five-factor asset pricing model. Fama-miller working paper, Chicago Booth. Available at <http://ssrn.com/abstract=2287202>.
- Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.
- Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.
- Frahm, G. and Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, 159(2):289–302.
- Green, J., Hand, J. R. M., and Zhang, X. F. (2013). The superview of return predictive signals. *Review of Accounting Studies*, 18:692–730.
- Harvey, C. R., Liu, Y., and Zhu, H. (2013). ... and the cross-section of expected returns. Working paper, Duke University. <http://dx.doi.org/10.2139/ssrn.2249314>.
- Haugen, R. A. and Baker, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management*, 17(3):35–40.
- Huang, C. and Litzenberger, R. (1988). *Foundations for Financial Economics*. North-Holland, New York.
- Ilmanen, A. (2011). *Expected Returns: An Investor's Guide to Harvesting Market Rewards*. Wiley Finance, New York.
- IPCC (2001). Climate change 2001: the scientific basis. In Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der LINDEN, P. J., Dai, X., Maskell, K., and Johnson, C., editors, *Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 881pp.
- IPCC (2007). Climate change 2007: the scientific basis. In Solomon, S., Qin, D., Manning, M., Marquis, M., Averyt, K., Tignor, M. M., Miller, H. L., and Chen, Z., editors, *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 996pp.

- Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 54(4):1651–1684.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.
- Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.
- Ledoit, O. and P ech e, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 150(1–2):233–264.
- Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.
- Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4):110–119.
- Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.
- Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15:850–859.
- Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott Magazine*, September:86–89.
- Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.
- Ledoit, O. and Wolf, M. (2013a). Optimal estimation of a large-dimensional covariance matrix under Stein’s loss. Working Paper ECON 122, Department of Economics, University of Zurich.
- Ledoit, O. and Wolf, M. (2013b). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. Working Paper ECON 105, Department of Economics, University of Zurich.
- Mar cenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.
- McLean, R. D. and Pontiff, J. (2013). Does academic research destroy stock return predictability? In *AFFI/EUROFIDAI, Paris December 2012 Finance Meetings Paper*.
- Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.

- Nielsen, F. and Aylursubramanian, R. (2008). Far from the madding crowd — Volatility efficient indices. Research insights, MSCI Barra.
- Ribes, A., Azais, J.-M., and Planton, S. (2009). Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance matrix estimate. *Climate Dynamics*, 33(5):707–722.
- Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1). Article 32.
- Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.
- Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Mathematical Sciences*, 34(1):1373–1403.
- Stieltjes, T. J. (1894). Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse 1<sup>re</sup> Série*, 8(4):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(1):27–42.
- Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.
- Wang, C., Pan, G., and Cao, L. (2012). A shrinkage estimation of large dimensional precision matrices using random matrix theory. Preprint arXiv:1211.2400.

## A Tables and Figures

Period: 01/19/1973–12/31/2011							
	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	11.14	8.64	8.65	8.68	8.52	8.71	8.72
SD	20.05	14.21	14.11	14.11	14.16	<b>14.08*</b>	14.08
SR	0.56	0.61	0.61	0.62	0.60	0.62	0.62
$N = 50$							
AV	9.54	4.65	4.99	4.70	5.10	5.21	5.22
SD	19.78	13.15	13.01	12.83	12.75	<b>12.68***</b>	12.68
SR	0.48	0.35	0.38	0.37	0.40	0.41	0.41
$N = 100$							
AV	10.53	4.74	5.31	4.96	4.99	5.10	5.12
SD	19.34	13.11	12.71	11.75	11.79	<b>11.52***</b>	11.55
SR	0.54	0.36	0.42	0.42	0.42	0.44	0.44
$N = 250$							
AV	9.57	275.02	275.02	6.73	5.81	6.26	6.43
SD	18.95	3,542.90	3,542.90	10.69	10.91	<b>10.34***</b>	10.49
SR	0.50	0.08	0.08	0.63	0.53	0.61	0.61
$N = 500$							
AV	9.78	NA	NA	5.90	5.03	5.34	5.41
SD	18.95	NA	NA	10.21	10.20	<b>9.65***</b>	9.75
SR	0.52	NA	NA	0.58	0.49	0.55	0.55

Table 1: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/19/1973–12/31/2011

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	12.47	13.37	11.22	4.14	2.38	11.29	11.87	11.87
SD	25.77	23.15	18.86	16.31	19.64	18.58	18.57	18.57
SR	0.48	0.58	0.60	0.25	0.12	0.61	<b>0.64</b>	0.64
$N = 50$								
AV	16.26	14.90	10.32	0.13	0.50	11.70	12.12	12.04
SD	24.60	22.18	16.86	2.66	2.25	16.30	16.23	16.24
SR	0.66	0.67	0.61	0.05	0.22	0.72	<b>0.75</b>	0.74
$N = 100$								
AV	15.74	14.98	10.93	-4.70	1.23	11.97	12.31	12.31
SD	22.44	20.32	16.00	25.34	1.85	14.68	14.30	14.30
SR	0.70	0.74	0.68	-0.19	0.66	0.82	<b>0.86</b>	0.86
$N = 250$								
AV	14.17	13.03	275.01	NA	NA	10.04	11.12	11.38
SD	21.77	19.86	3,542.92	NA	NA	12.82	12.14	12.32
SR	0.65	0.66	0.08	NA	NA	0.78	<b>0.92**</b>	0.92
$N = 500$								
AV	14.67	13.58	NA	NA	NA	8.92	9.94	9.87
SD	21.54	19.63	NA	NA	NA	11.82	11.09	11.26
SR	0.68	0.69	NA	NA	NA	0.75	<b>0.90**</b>	0.88

Table 2: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.



Period: 01/19/1973–12/31/2011

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 50$							
Min	0.0333	-0.0729	-0.0653	-0.0690	-0.0522	-0.0584	-0.0582
Max	0.0333	0.2720	0.2562	0.2698	0.1902	0.2210	0.2228
SD	0.0000	0.0737	0.0686	0.0726	0.0562	0.0614	0.0616
MAD-EW	0.0000	0.0515	0.0479	0.0507	0.0423	0.0444	0.0444
$N = 50$							
Min	0.0200	-0.0762	-0.0680	-0.0695	-0.0531	-0.0537	-0.0536
Max	0.0200	0.2219	0.2055	0.2211	0.1461	0.1556	0.1589
SD	0.0000	0.0552	0.0506	0.0533	0.0411	0.0415	0.0418
MAD-EW	0.0000	0.0386	0.0353	0.0367	0.0309	0.0306	0.0306
$N = 100$							
Min	0.0100	-0.0837	-0.0733	-0.0598	-0.0499	-0.0423	-0.0424
Max	0.0100	0.1776	0.1595	0.1684	0.0989	0.0846	0.0890
SD	0.0000	0.0407	0.0362	0.0333	0.0270	0.0234	0.0237
MAD-EW	0.0000	0.0288	0.0256	0.0218	0.0205	0.0180	0.0181
$N = 250$							
Min	0.0040	-7.2464	-7.2464	-0.0438	-0.0362	-0.0260	-0.0263
Max	0.0040	6.7094	6.7094	0.1225	0.0530	0.0357	0.0362
SD	0.0000	1.9296	1.9296	0.0172	0.0150	0.0108	0.0108
MAD-EW	0.0000	1.4159	1.4159	0.0098	0.0118	0.0085	0.0086
$N = 500$							
Min	0.0020	NA	NA	-0.0359	-0.0232	-0.0167	-0.0164
Max	0.0020	NA	NA	0.0998	0.0293	0.0199	0.0203
SD	0.0000	NA	NA	0.0106	0.0083	0.0059	0.0059
MAD-EW	0.0000	NA	NA	0.0052	0.0066	0.0046	0.0047

Table 3: Average characteristics of the weight vectors of various estimators of the GMV portfolio. Min stands for minimum weight; Max stands for maximum weight; SD stands for standard deviation of the weights; and MAD-EW stands for mean absolute deviation from the equal-weighted portfolio (that is, from  $1/N$ ). All measures reported are the averages of the corresponding characteristic over the 480 portfolio formations.

Period: 01/19/1973–12/31/2011

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 50$								
Min	0.0000	-0.0687	-0.1371	-0.0098	-0.0092	-0.1095	-0.1164	-0.1165
Max	0.1667	0.1287	0.3173	0.0136	0.0139	0.2344	0.2612	0.2629
SD	0.0678	0.0452	0.0988	0.0051	0.0051	0.0805	0.0853	0.0854
MAD-EW	0.0533	0.0343	0.0722	0.0326	0.0321	0.0618	0.0642	0.0642
$N = 50$								
Min	0.0000	-0.0466	-0.1198	-0.0118	-0.0110	-0.0895	-0.0883	-0.0886
Max	0.1000	0.0823	0.2588	0.0143	0.0144	0.1733	0.1838	0.1873
SD	0.0404	0.0271	0.0716	0.0051	0.0049	0.0554	0.0558	0.0561
MAD-EW	0.0320	0.0206	0.0519	0.0196	0.0191	0.0428	0.0425	0.0426
$N = 100$								
Min	0.0000	-0.0257	-0.1105	-0.0130	-0.0120	-0.0682	-0.0591	-0.0594
Max	0.0500	0.0446	0.2027	0.0144	0.0141	0.1148	0.0986	0.1032
SD	0.0201	0.0135	0.0501	0.0045	0.0043	0.0343	0.0301	0.0305
MAD-EW	0.0160	0.0102	0.0362	0.0100	0.0096	0.0265	0.0236	0.0237
$N = 250$								
Min	0.0000	-0.0110	-7.2457	NA	NA	-0.0452	-0.0334	-0.0338
Max	0.0200	0.0187	6.7088	NA	NA	0.0620	0.0426	0.0433
SD	0.0080	0.0054	1.9294	NA	NA	0.0182	0.0133	0.0134
MAD-EW	0.0064	0.0041	1.4157	NA	NA	0.0144	0.0105	0.0106
$N = 500$								
Min	0.0000	-0.0056	NA	NA	NA	-0.0285	-0.0209	-0.0207
Max	0.0100	0.0095	NA	NA	NA	0.0342	0.0239	0.0243
SD	0.0040	0.0027	NA	NA	NA	0.0099	0.0071	0.0071
MAD-EW	0.0032	0.0020	NA	NA	NA	0.0079	0.0056	0.0056

Table 4: Average characteristics of the weight vectors of various estimators of the Markowitz portfolio with momentum signal. Min stands for minimum weight; Max stands for maximum weight; SD stands for standard deviation of the weights; and MAD-EW stands for mean absolute deviation from the equal-weighted portfolio (that is, from  $1/N$ ). All measures reported are the averages of the corresponding characteristic over the 480 portfolio formations.

Period: 01/19/1973–05/08/1986							
	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	7.64	4.97	5.00	4.97	5.49	5.40	5.39
SD	15.43	11.63	11.54	11.63	11.42	11.39	11.39
SR	0.50	0.43	0.43	0.43	0.48	<b>0.47</b>	0.47
$N = 50$							
AV	7.20	3.63	3.93	3.57	4.58	4.68	4.68
SD	14.83	10.91	10.74	10.88	10.54	<b>10.47**</b>	10.47
SR	0.49	0.33	0.37	0.33	0.43	0.45	0.45
$N = 100$							
AV	8.62	6.06	6.09	6.54	6.82	6.92	6.90
SD	14.50	10.39	10.03	9.61	9.43	<b>9.18***</b>	9.18
SR	0.59	0.58	0.61	0.68	0.72	0.75	0.75
$N = 250$							
AV	6.16	-527.27	-527.27	4.78	5.08	6.09	5.96
SD	14.18	2,009.60	2,009.60	8.41	8.51	<b>7.81***</b>	7.86
SR	0.43	-0.26	-0.26	0.57	0.60	0.78	0.76
$N = 500$							
AV	6.91	NA	NA	3.35	4.20	5.45	5.41
SD	14.14	NA	NA	7.82	7.50	<b>7.03***</b>	7.14
SR	0.49	NA	NA	0.43	0.56	0.77	0.76

Table 5: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 05/08/1986. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 05/09/1986–08/25/1999

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	12.56	12.06	12.01	12.04	12.23	12.08	12.10
SD	16.24	12.91	12.81	12.81	12.78	<b>12.77</b>	12.77
SR	0.77	0.93	0.94	0.94	0.96	0.95	0.95
$N = 50$							
AV	12.94	7.22	7.69	7.58	7.20	7.62	7.60
SD	15.82	12.04	11.94	11.90	11.86	<b>11.85</b>	11.85
SR	0.82	0.60	0.64	0.64	0.61	0.64	0.64
$N = 100$							
AV	12.45	6.68	7.23	6.92	6.84	6.56	6.61
SD	15.37	11.42	11.10	10.63	10.39	<b>10.26***</b>	10.28
SR	0.81	0.58	0.65	0.65	0.66	0.64	0.64
$N = 250$							
AV	12.04	652.67	652.67	7.92	7.41	6.72	6.70
SD	15.09	2,126.98	2,126.98	10.07	9.78	<b>9.55***</b>	9.64
SR	0.80	0.31	0.31	0.79	0.76	0.70	0.70
$N = 500$							
AV	12.12	NA	NA	9.45	7.96	6.93	7.29
SD	15.01	NA	NA	9.57	9.21	<b>8.93***</b>	9.06
SR	0.81	NA	NA	0.99	0.86	0.78	0.80

Table 6: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate from 05/09/1986 until 08/25/1999. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 08/26/1999–12/31/2011

	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	13.22	8.90	8.96	9.03	7.84	8.65	8.66
SD	26.55	17.43	17.32	17.27	17.54	<b>17.38</b>	17.38
SR	0.50	0.51	0.52	0.52	0.45	0.50	0.50
$N = 50$							
AV	8.49	3.10	3.34	2.95	3.52	3.34	3.38
SD	26.53	15.97	15.80	15.30	15.36	<b>15.25**</b>	15.25
SR	0.32	0.19	0.21	0.19	0.23	0.22	0.22
$N = 100$							
AV	10.52	1.47	2.61	1.43	1.32	1.82	1.86
SD	25.99	16.65	16.16	14.46	14.84	<b>14.46***</b>	14.49
SR	0.40	0.09	0.16	0.10	0.09	0.13	0.13
$N = 250$							
AV	10.50	699.67	699.67	7.50	4.95	5.97	6.63
SD	25.47	5,394.22	5,394.22	13.07	13.76	<b>12.99***</b>	13.24
SR	0.41	0.13	0.13	0.57	0.36	0.46	0.50
$N = 500$							
AV	10.30	NA	NA	4.91	2.93	3.65	3.53
SD	25.54	NA	NA	12.64	13.07	<b>12.24***</b>	12.34
SR	0.40	NA	NA	0.39	0.22	0.30	0.29

Table 7: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate from 08/26/1999 until 12/31/2011. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/19/1973–05/08/1986

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	11.84	12.15	8.05	7.29	7.65	8.93	8.74	8.73
SD	21.37	18.79	15.33	22.12	33.77	15.03	15.05	15.04
SR	0.55	0.65	0.53	0.33	0.23	<b>0.59</b>	0.58	0.58
$N = 50$								
AV	13.62	12.02	6.97	-0.25	0.27	8.28	9.02	8.89
SD	19.50	17.66	13.52	2.61	1.49	13.14	13.05	13.06
SR	<b>0.70</b>	0.68	0.52	-0.10	0.18	0.63	0.69	0.68
$N = 100$								
AV	12.64	13.77	9.51	0.86	1.27	10.16	10.38	10.36
SD	17.77	16.20	12.74	1.69	1.91	11.58	11.32	11.33
SR	0.71	0.85	0.75	0.51	0.67	0.88	<b>0.92</b>	0.91
$N = 250$								
AV	11.90	12.32	-527.28	NA	NA	7.99	10.02	9.96
SD	16.88	15.55	2,009.60	NA	NA	10.21	9.31	9.39
SR	0.70	0.79	-0.26	NA	NA	0.78	<b>1.08***</b>	1.06
$N = 500$								
AV	12.35	12.73	NA	NA	NA	6.86	9.65	9.48
SD	16.59	15.32	NA	NA	NA	8.73	8.12	8.29
SR	0.74	0.83	NA	NA	NA	0.79	<b>1.19***</b>	1.14

Table 8: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 05/08/1986. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 05/09/1986–08/25/1999

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	14.06	11.89	12.27	-1.04	0.56	12.21	12.36	12.36
SD	23.70	21.94	19.13	3.66	1.29	18.68	18.55	18.55
SR	0.59	0.54	0.64	-0.28	0.43	0.65	<b>0.67</b>	0.67
$N = 50$								
AV	19.81	18.06	10.08	0.70	1.47	11.46	11.74	11.73
SD	22.96	20.48	16.65	1.36	2.84	16.19	16.16	16.17
SR	<b>0.86</b>	0.88	0.61	0.52	0.52	0.71	0.73	0.73
$N = 100$								
AV	20.25	18.20	12.63	0.64	1.34	13.20	14.00	14.07
SD	20.42	18.33	15.59	2.91	1.97	14.38	14.30	14.28
SR	0.99	0.99	0.81	0.22	0.68	0.92	0.98	<b>0.99</b>
$N = 250$								
AV	18.12	15.08	652.50	NA	NA	11.74	12.44	12.51
SD	19.57	17.78	2127.18	NA	NA	11.95	11.86	12.01
SR	0.93	0.85	0.31	NA	NA	0.98	<b>1.05</b>	1.04
$N = 500$								
AV	18.95	16.35	NA	NA	NA	12.19	12.67	12.71
SD	19.39	17.62	NA	NA	NA	11.05	10.82	11.03
SR	0.98	0.93	NA	NA	NA	1.10	<b>1.17</b>	1.15

Table 9: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate from 05/09/1986 until 08/25/1999. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 08/26/1999–12/31/2011

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	11.50	16.05	13.34	6.16	-1.08	12.71	14.54	14.52
SD	31.21	27.82	21.58	17.20	3.99	21.47	21.54	21.54
SR	0.37	0.58	0.62	0.36	-0.27	0.59	<b>0.68</b>	0.67
$N = 50$								
AV	15.35	14.63	13.91	-0.06	-0.23	15.35	15.59	15.50
SD	30.15	27.29	19.82	3.54	2.22	19.03	18.95	18.97
SR	0.51	0.54	0.70	-0.02	-0.10	0.81	<b>0.82</b>	0.82
$N = 100$								
AV	14.33	12.97	10.65	-15.61	1.06	12.54	12.52	12.53
SD	27.89	25.30	19.06	43.75	1.65	17.49	16.76	16.78
SR	0.51	0.51	0.56	-0.36	0.65	0.72	<b>0.75</b>	0.75
$N = 250$								
AV	12.51	11.68	699.80	NA	NA	10.40	10.92	11.67
SD	27.45	25.02	5,394.16	NA	NA	15.69	14.67	14.93
SR	0.46	0.47	0.13	NA	NA	0.66	0.74	<b>0.78</b>
$N = 500$								
AV	12.71	11.67	NA	NA	NA	7.70	7.51	7.42
SD	27.22	24.71	NA	NA	NA	14.84	13.63	13.78
SR	0.47	0.47	NA	NA	NA	0.52	<b>0.55</b>	0.54

Table 10: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 3,360 daily out-of-sample returns in excess of the risk-free rate from 08/26/1999 until 12/31/2011. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.



Period: 01/19/1973–12/31/2011							
	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	11.14	8.68	8.73	8.61	8.80	8.78	8.77
SD	20.05	14.15	14.11	14.12	14.11	<b>14.09</b>	14.09
SR	0.56	0.61	0.62	0.61	0.62	0.62	0.62
$N = 50$							
AV	9.54	6.20	6.32	6.06	6.21	6.25	6.26
SD	19.78	12.87	12.82	12.70	12.67	<b>12.67</b>	12.67
SR	0.48	0.48	0.49	0.48	0.49	0.49	0.49
$N = 100$							
AV	10.53	4.74	5.07	5.23	5.17	5.31	5.30
SD	19.34	11.95	11.85	11.51	11.59	<b>11.48</b> ***	11.48
SR	0.54	0.40	0.43	0.45	0.45	0.46	0.46
$N = 250$							
AV	9.57	7.43	7.47	6.69	6.77	6.66	6.79
SD	18.95	11.83	11.55	10.21	10.51	<b>10.09</b> ***	10.12
SR	0.50	0.63	0.65	0.66	0.64	0.66	0.67
$N = 500$							
AV	9.78	1,200.31	1,200.31	6.83	6.66	6.25	6.33
SD	18.95	8,551.27	8,551.26	9.54	10.33	<b>9.43</b> ***	9.87
SR	0.52	0.14	0.14	0.72	0.64	0.66	0.64

Table 11: Annualized performance measures (in percent) for various estimators of the GMV portfolio. The past window to estimate the covariance matrix is taken to be of length  $T = 500$  days instead of  $T = 250$  days. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/19/1973–12/31/2011

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	12.47	13.37	11.32	0.29	0.15	11.51	11.94	11.92
SD	25.77	23.15	18.44	0.93	1.22	18.32	18.46	18.46
SR	0.48	0.58	0.61	0.31	0.13	0.63	<b>0.65</b>	0.65
$N = 50$								
AV	16.26	14.90	11.58	1.91	2.23	12.17	12.54	12.51
SD	24.60	22.18	16.25	7.52	10.80	16.12	16.11	16.11
SR	0.66	0.67	0.71	0.25	0.21	0.76	<b>0.78</b>	0.78
$N = 100$								
AV	15.74	14.98	11.07	0.53	4.11	11.88	11.94	11.92
SD	22.44	20.32	14.54	2.29	22.56	14.15	13.95	13.96
SR	0.70	0.74	0.76	0.23	0.18	0.84	<b>0.86</b>	0.85
$N = 250$								
AV	14.17	13.03	11.25	3.28	0.60	10.72	11.11	11.24
SD	21.77	19.86	13.70	9.44	5.76	12.26	11.83	11.87
SR	0.65	0.66	0.82	0.35	0.10	0.87	<b>0.94</b>	0.94
$N = 500$								
AV	14.67	13.58	1,205.18	NA	NA	10.53	10.47	10.57
SD	21.54	19.63	8,551.62	NA	NA	11.95	10.80	11.22
SR	0.68	0.69	0.14	NA	NA	0.88	<b>0.97</b>	0.94

Table 12: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. The past window to estimate the covariance matrix is taken to be of length  $T = 500$  days instead of  $T = 250$  days. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/19/1973–12/31/2011							
	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	11.14	8.27	8.30	8.31	8.21	8.42	8.44
SD	20.05	14.57	14.42	14.46	14.38	14.38	<b>14.37</b>
SR	0.56	0.57	0.58	0.57	0.57	0.59	0.59
$N = 50$							
AV	9.54	4.60	4.94	4.55	5.00	5.29	5.31
SD	19.78	13.56	13.34	13.09	13.13	12.97***	<b>12.95</b>
SR	0.48	0.34	0.37	0.35	0.38	0.41	0.41
$N = 100$							
AV	10.53	4.84	5.29	5.71	4.89	5.04	5.08
SD	19.34	13.86	13.29	12.20	12.54	11.95***	<b>11.93</b>
SR	0.54	0.35	0.40	0.47	0.39	0.42	0.43
$N = 250$							
AV	9.57	-2,498.27	-2,498.27	6.99	6.72	6.70	6.75
SD	18.95	12,130.15	12,130.15	10.81	11.75	<b>10.58***</b>	10.58
SR	0.50	-0.21	-0.21	0.65	0.57	0.63	0.64
$N = 500$							
AV	9.78	NA	NA	5.86	5.44	5.56	5.68
SD	18.95	NA	NA	10.33	10.83	<b>9.71***</b>	9.78
SR	0.52	NA	NA	0.57	0.50	0.57	0.58

Table 13: Annualized performance measures (in percent) for various estimators of the GMV portfolio. In the estimation of a covariance matrix, the past returns are winsorized as described in Appendix D. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/19/1973–12/31/2011

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	12.47	13.37	11.27	0.82	2.08	11.38	11.93	11.94
SD	25.77	23.15	19.16	3.61	4.68	18.83	18.85	18.84
SR	0.48	0.58	0.59	0.23	0.45	0.60	<b>0.63</b>	0.63
$N = 50$								
AV	16.26	14.90	10.71	1.62	-5.88	11.46	12.04	11.99
SD	24.60	22.18	17.12	8.05	32.69	16.65	16.53	16.52
SR	0.66	0.67	0.63	0.20	-0.18	0.69	<b>0.73</b>	0.73
$N = 100$								
AV	15.74	14.98	10.47	0.52	1.12	11.39	11.75	11.75
SD	22.44	20.32	16.39	18.21	6.16	15.05	14.50	14.50
SR	0.70	0.74	0.64	0.03	0.18	0.76	<b>0.81</b>	0.81
$N = 250$								
AV	14.17	13.03	-2,498.52	NA	NA	10.70	11.11	11.14
SD	21.77	19.86	12,130.23	NA	NA	13.82	12.33	12.37
SR	0.65	0.66	-0.21	NA	NA	0.77	<b>0.90*</b>	0.90
$N = 500$								
AV	14.67	13.58	NA	NA	NA	9.16	9.90	9.91
SD	21.54	19.63	NA	NA	NA	12.59	11.13	11.26
SR	0.68	0.69	NA	NA	NA	0.73	<b>0.89**</b>	0.88

Table 14: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. In the estimation of a covariance matrix, the past returns are winsorized as described in Appendix D. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/19/1973–12/31/2011							
	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	11.14	9.75	NA	NA	9.87	9.92	9.92
SD	20.05	<b>14.21</b>	NA	NA	14.41	14.26	14.26
SR	0.56	0.69	NA	NA	0.69	0.70	0.70
$N = 50$							
AV	9.54	6.97	NA	NA	6.99	7.10	7.08
SD	19.78	<b>12.96</b>	NA	NA	13.06	12.99	12.99
SR	0.48	0.54	NA	NA	0.54	0.55	0.55
$N = 100$							
AV	10.53	6.93	NA	NA	7.22	7.37	7.35
SD	19.34	<b>12.04</b>	NA	NA	12.14	12.17	12.18
SR	0.54	0.58	NA	NA	0.59	0.61	0.60
$N = 250$							
AV	9.57	7.40	NA	NA	7.40	7.51	7.58
SD	18.95	<b>11.06</b>	NA	NA	11.09	11.20	11.29
SR	0.50	0.67	NA	NA	0.67	0.67	0.67
$N = 500$							
AV	9.78	7.21	NA	NA	7.16	7.54	7.56
SD	18.95	<b>10.57</b>	NA	NA	10.59	10.74	10.83
SR	0.52	0.68	NA	NA	0.68	0.70	0.70

Table 15: Annualized performance measures (in percent) for various estimators of the GMV portfolio. A lower bound of zero is imposed on all portfolio weights, so that short sales are not allowed. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/19/1973–12/31/2011

	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	12.47	13.37	12.06	NA	NA	12.12	13.21	13.20
SD	25.77	23.15	23.07	NA	NA	23.05	23.11	23.11
SR	0.48	<b>0.58</b>	0.52	NA	NA	0.53	0.57	0.57
$N = 50$								
AV	16.26	14.90	13.16	NA	NA	13.65	14.54	14.52
SD	24.60	22.18	20.93	NA	NA	20.84	21.04	21.04
SR	0.66	0.67	0.63	NA	NA	0.65	<b>0.69</b>	0.69
$N = 100$								
AV	15.74	14.98	14.50	NA	NA	14.85	15.69	15.67
SD	22.44	20.32	18.63	NA	NA	18.59	18.73	18.71
SR	0.70	0.74	0.78	NA	NA	0.80	<b>0.84</b>	0.84
$N = 250$								
AV	14.17	13.03	12.57	NA	NA	13.04	14.50	14.51
SD	21.77	19.86	16.60	NA	NA	16.57	16.82	16.84
SR	0.65	0.66	0.76	NA	NA	0.79	<b>0.86</b>	0.86
$N = 500$								
AV	14.67	13.58	13.66	NA	NA	13.82	15.27	15.09
SD	21.54	19.63	15.26	NA	NA	15.26	15.55	15.63
SR	0.68	0.69	0.90	NA	NA	0.91	<b>0.98</b>	0.97

Table 16: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. A lower bound of zero is imposed on all portfolio weights, so that short sales are not allowed. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 10,080 daily out-of-sample returns in excess of the risk-free rate from 01/19/1973 until 12/31/2011. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/1955–12/2011							
	1/N	Sample	FM	FZY	Lin	NonLin	NL-Inv
$N = 30$							
AV	7.05	6.43	6.46	6.64	6.82	7.35	7.30
SD	16.69	14.25	13.68	13.90	13.48	13.45	<b>13.42</b>
SR	0.42	0.45	0.47	0.48	0.51	0.55	0.54
$N = 50$							
AV	6.74	3.30	3.95	3.25	4.37	4.47	4.56
SD	16.44	15.12	14.06	13.16	12.90	12.61**	<b>12.56</b>
SR	0.41	0.22	0.28	0.25	0.34	0.35	0.36
$N = 100$							
AV	6.29	9.22	8.77	6.30	6.66	6.32	6.45
SD	16.06	25.98	21.36	12.99	13.67	12.70***	<b>12.81</b>
SR	0.39	0.35	0.41	0.48	0.49	0.50	0.50
$N = 250$							
AV	7.29	NA	NA	4.56	3.24	4.61	4.89
SD	16.05	NA	NA	11.76	11.92	10.77***	<b>10.75</b>
SR	0.45	NA	NA	0.39	0.27	0.43	0.45
$N = 500$							
AV	7.03	NA	NA	4.98	4.05	4.81	4.88
SD	15.98	NA	NA	11.35	10.59	<b>10.23**</b>	10.37
SR	0.44	NA	NA	0.44	0.38	0.47	0.47

Table 17: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 684 monthly out-of-sample returns in excess of the risk-free rate from 01/1955 until 12/2011. In the rows labeled SD, the lowest number appears in bold face. 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: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.

Period: 01/1955–12/2011								
	EW-TQ	BSV	Sample	KZ	TZ	Lin	NonLin	NL-Inv
$N = 30$								
AV	7.87	7.03	6.29	7.50	8.06	6.55	7.05	7.08
SD	19.90	18.08	17.49	40.55	22.41	16.31	16.16	16.15
SR	0.40	0.39	0.36	0.18	0.36	0.40	0.44*	<b>0.44</b>
$N = 50$								
AV	8.94	9.58	5.57	2.48	6.00	6.71	7.77	7.77
SD	19.32	17.88	17.82	40.80	20.95	15.24	14.60	14.63
SR	0.46	0.54	0.31	0.06	0.29	0.44	<b>0.53**</b>	0.53
$N = 100$								
AV	9.32	9.74	4.03	-1.87	3.18	8.16	8.55	8.63
SD	18.47	16.95	28.82	36.16	19.80	14.50	12.99	13.09
SR	0.50	0.57	0.14	-0.05	0.16	0.56	<b>0.66*</b>	0.66
$N = 250$								
AV	10.88	10.62	NA	NA	NA	5.52	8.22	8.50
SD	17.91	16.58	NA	NA	NA	13.98	11.85	11.83
SR	0.61	0.64	NA	NA	NA	0.39	0.69**	<b>0.72</b>
$N = 500$								
AV	10.17	10.21	NA	NA	NA	5.44	7.82	7.53
SD	17.70	16.41	NA	NA	NA	12.90	11.06	11.54
SR	0.57	0.62	NA	NA	NA	0.42	<b>0.71***</b>	0.65

Table 18: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 684 monthly out-of-sample returns in excess of the risk-free rate from 01/1955 until 12/2011. In the rows labeled SR, the largest number appears in bold face. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled Lin and NonLin, significant outperformance of one of the two portfolios over the other in terms of SR is denoted by asterisks: \*\*\* denotes significance at the 0.01 level; \*\* denotes significance at the 0.05 level; and \* denotes significance at the 0.1 level.



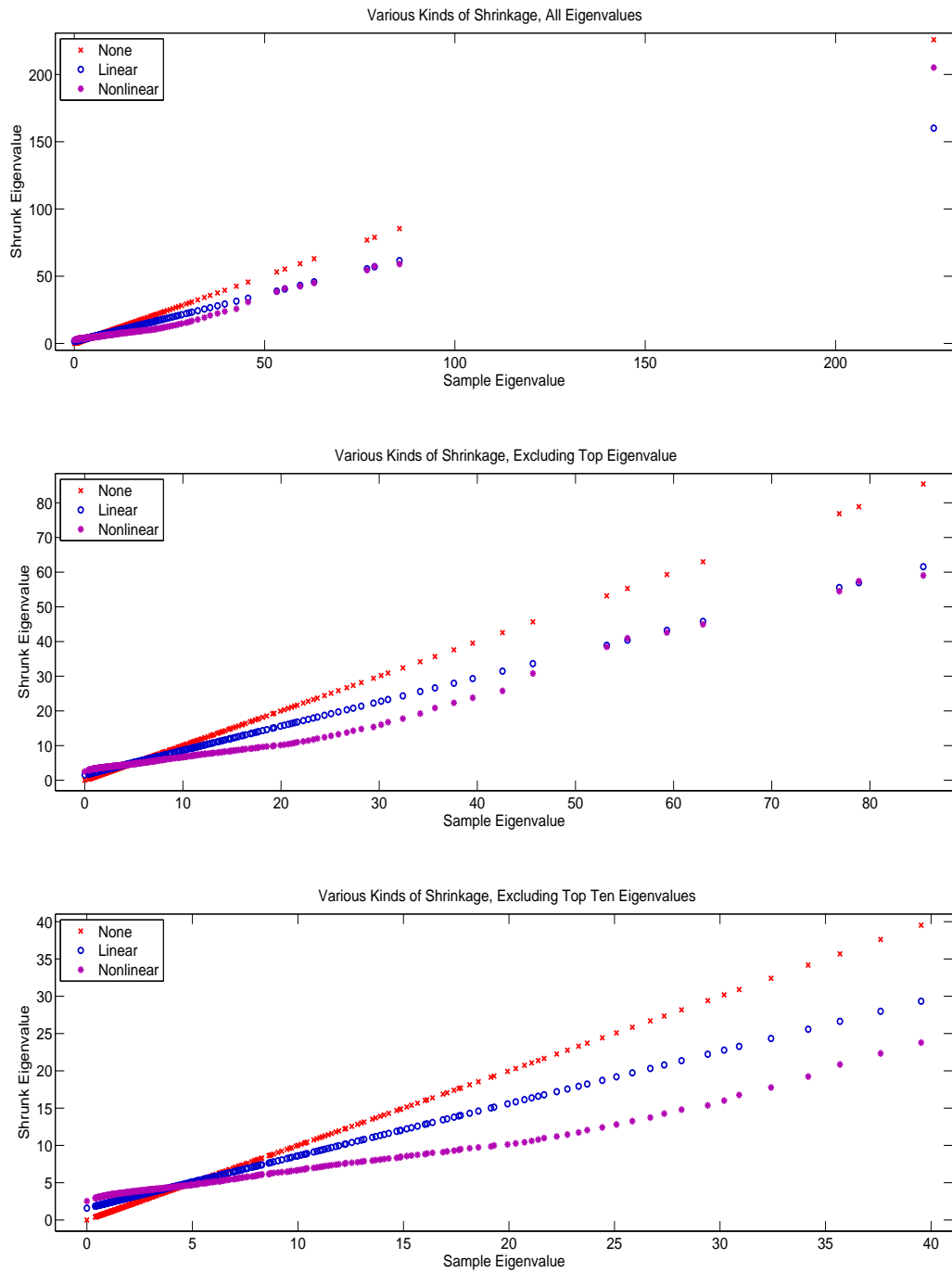


Figure 1: Shrunken eigenvalues as a function of sample eigenvalues for sample covariance matrix, linear shrinkage, and nonlinear shrinkage for an exemplary data set. The size of the investment universe is  $N = 500$ .

## B Mathematical Proofs

### B.1 Preliminaries

We shall use the notations  $\operatorname{Re}(z)$  and  $\operatorname{Im}(z)$  for the real and imaginary parts of a complex number  $z$ , so that

$$\forall z \in \mathbb{C} \quad z = \operatorname{Re}(z) + i \cdot \operatorname{Im}(z) .$$

For any increasing function  $G$  on the real line,  $s_G$  denotes the Stieltjes transform of  $G$ :

$$\forall z \in \mathbb{C}^+ \quad s_G(z) := \int \frac{1}{\lambda - z} dG(\lambda) .$$

The Stieltjes transform admits a well-known inversion formula:

$$G(b) - G(a) = \lim_{\eta \rightarrow 0^+} \frac{1}{\pi} \int_a^b \operatorname{Im} [s_G(\xi + i\eta)] d\xi , \quad (\text{B.1})$$

as long as  $G$  is continuous at both  $a$  and  $b$ . Bai and Silverstein (2010, p.112) give the following version for the equation that relates  $F$  to  $H$  and  $c$ . The quantity  $s =: s_F(z)$  is the unique solution in the set

$$\left\{ s \in \mathbb{C} : -\frac{1-c}{z} + cs \in \mathbb{C}^+ \right\} \quad (\text{B.2})$$

to the equation

$$\forall z \in \mathbb{C}^+ \quad s = \int \frac{1}{\tau [1 - c - czs] - z} dH(\tau) . \quad (\text{B.3})$$

Although the Stieltjes transform of  $F$ ,  $s_F$ , is a function whose domain is the upper half of the complex plane, it admits an extension to the real line  $\forall x \in \mathbb{R} - \{0\}$   $\check{s}_F(x) := \lim_{z \in \mathbb{C}^+ \rightarrow x} s_F(z)$  which is continuous over  $x \in \mathbb{R} - \{0\}$ . When  $c < 1$ ,  $\check{s}_F(0)$  also exists and  $F$  has a continuous derivative  $F' = \pi^{-1} \operatorname{Im} [\check{s}_F]$  on all of  $\mathbb{R}$  with  $F' \equiv 0$  on  $(-\infty, 0]$ . (One should remember that, although the argument of  $\check{s}_F$  is real-valued now, the output of the function is still a complex number.)

Recall that the limiting e.d.f. of the eigenvalues of  $n^{-1}Y_n'Y_n = n^{-1}\Sigma_n^{1/2}X_n'X_n\Sigma_n^{1/2}$  was defined as  $F$ . In addition, define the limiting e.d.f. of the eigenvalues of  $n^{-1}Y_nY_n' = n^{-1}X_n\Sigma_nX_n'$  as  $\underline{F}$ ; note that the eigenvalues of  $n^{-1}Y_n'Y_n$  and  $n^{-1}Y_nY_n'$  only differ by  $|n-p|$  zero eigenvalues. It then holds:

$$\forall x \in \mathbb{R} \quad \underline{F}(x) = (1-c) \mathbb{1}_{[0,\infty)}(x) + cF(x) \quad (\text{B.4})$$

$$\forall x \in \mathbb{R} \quad F(x) = \frac{c-1}{c} \mathbb{1}_{[0,\infty)}(x) + \frac{1}{c} \underline{F}(x) \quad (\text{B.5})$$

$$\forall z \in \mathbb{C}^+ \quad s_{\underline{F}}(z) = \frac{c-1}{z} + cs_F(z) \quad (\text{B.6})$$

$$\forall z \in \mathbb{C}^+ \quad s_F(z) = \frac{1-c}{cz} + \frac{1}{c} s_{\underline{F}}(z) . \quad (\text{B.7})$$

Although the Stieltjes transform of  $\underline{F}$ ,  $s_{\underline{F}}$ , is again a function whose domain is the upper half of the complex plane, it also admits an extension to the real line (except at zero):  $\forall x \in \mathbb{R} \setminus \{0\}$ ,  $\check{s}_{\underline{F}}(x) := \lim_{z \in \mathbb{C}^+ \rightarrow x} s_{\underline{F}}(z)$  exists. Furthermore, the function  $\check{s}_{\underline{F}}$  is continuous over

$\mathbb{R} \setminus \{0\}$ . When  $c > 1$ ,  $\check{s}_F(0)$  also exists and  $\underline{F}$  has a continuous derivative  $\underline{F}' = \pi^{-1} \text{Im} [\check{s}_F]$  on all of  $\mathbb{R}$  with  $\underline{F}' \equiv 0$  on  $(-\infty, 0]$ .

It can easily be verified that the function  $s(x)$  defined in equation (MP) is in fact none other than  $\check{s}_F(x)$ . Equation (4.2.2) of Bai and Silverstein (2010), for example, gives an expression analogous to equation (MP). Based on the right-hand side of equation (3.3), we can rewrite the function  $d^*$  introduced in Theorem 4.1 as:

$$\forall x \in \bigcup_{k=1}^{\kappa} [a_k, b_k] \quad d^*(x) = \frac{1}{x |\check{s}_F(x)|^2} = \frac{x}{|1 - c - cx \check{s}_F(x)|^2}. \quad (\text{B.8})$$

## B.2 Proof of Theorem 3.1

Given that it is only the normalized quantity  $(m'_T m_T)^{-1/2} m_T$  that appears in this proposition, the parametric form of the distribution of the underlying quantity  $m_T$  is irrelevant, as long as  $(m'_T m_T)^{-1/2} m_T$  is uniformly distributed on the unit sphere. Thus, we can assume without loss of generality that  $m_T$  is normally distributed with mean zero and covariance matrix the identity.

In this case, the assumptions of Lemma 1 of Ledoit and P ech e (2011) are satisfied. This implies that there exists a constant  $K_1$  independent of  $T$ ,  $\widehat{\Sigma}_T$  and  $m_T$  such that

$$\mathbb{E} \left[ \left( \frac{1}{N} m'_T \widehat{\Sigma}_T^{-1} m_T - \frac{1}{N} \text{Tr} \left( \widehat{\Sigma}_T^{-1} \right) \right)^6 \right] \leq \frac{K_1 \left\| \widehat{\Sigma}_T^{-1} \right\|}{N^3}.$$

Note that  $\left\| \widehat{\Sigma}_T^{-1} \right\| \leq \widehat{K}/h$  a.s. for large enough  $T$  by Assumption 3.4. Therefore,

$$\frac{1}{N} m'_T \widehat{\Sigma}_T^{-1} m_T - \frac{1}{N} \text{Tr} \left( \widehat{\Sigma}_T^{-1} \right) \xrightarrow{\text{a.s.}} 0.$$

In addition, we have

$$\frac{1}{N} \text{Tr} \left( \widehat{\Sigma}_T^{-1} \right) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\widehat{\delta}_T(\lambda_{T,i})} = \int \frac{1}{\widehat{\delta}_T(x)} dF_T(x) \xrightarrow{\text{a.s.}} \int \frac{1}{\widehat{\delta}(x)} dF(x).$$

Therefore,

$$\frac{1}{N} m'_T \widehat{\Sigma}_T^{-1} m_T \xrightarrow{\text{a.s.}} \int \frac{1}{\widehat{\delta}(x)} dF(x). \quad (\text{B.9})$$

A similar line of reasoning leads to

$$\frac{1}{N} m'_T \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} m_T - \frac{1}{N} \text{Tr} \left( \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} \right) \xrightarrow{\text{a.s.}} 0.$$

Notice that

$$\frac{1}{N} \text{Tr} \left( \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} \right) = \frac{1}{N} \text{Tr} \left( U'_T \Sigma_T U_T \widehat{\Delta}_T^{-2} \right) = \frac{1}{N} \sum_{i=1}^N \frac{u'_{T,i} \Sigma_T u_{T,i}}{\widehat{\delta}_T(\lambda_{T,i})^2}.$$

Using Theorem 4 of Ledoit and P ech e (2011), we obtain that

$$\frac{1}{N} \sum_{i=1}^N \frac{u'_{T,i} \Sigma_T u_{T,i}}{\widehat{\delta}_T(\lambda_{T,i})^2} \xrightarrow{\text{a.s.}} \int \frac{d^*(x)}{\widehat{\delta}(x)^2} dF(x),$$

with the function  $d^*$  defined by equation (4.1). Thus,

$$\frac{1}{N} m'_T \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} m_T \xrightarrow{\text{a.s.}} \int \frac{d^*(x)}{\widehat{\delta}(x)^2} dF(x) . \quad (\text{B.10})$$

Putting equations (B.9) and (B.10) together yields

$$N \frac{m'_T \widehat{\Sigma}_T^{-1} \Sigma_T \widehat{\Sigma}_T^{-1} m_T}{\left(m'_T \widehat{\Sigma}_T^{-1} m_T\right)^2} \xrightarrow{\text{a.s.}} \frac{\int \frac{d^*(x)}{\widehat{\delta}(x)^2} dF(x)}{\left(\int \frac{1}{\widehat{\delta}(x)} dF(x)\right)^2} .$$

Theorem 3.1 then follows from noticing that  $N^{-1} m'_T m_T \xrightarrow{\text{a.s.}} 1$ . ■

### B.3 Proof of Theorem 4.1

Differentiating the right-hand side of equation (3.3) with respect to  $\widehat{\delta}(x)$  for  $x \in \text{Supp}(F)$  yields the first-order condition

$$-2 \frac{d^*(x) F'(x)}{\widehat{\delta}(x)^3} \left[ \int \frac{dF(y)}{\widehat{\delta}(y)} \right]^{-2} + 2 \left[ \int \frac{dF(y)}{\widehat{\delta}(y)^2} \right]^{-3} \frac{F'(x)}{\widehat{\delta}(x)} \left[ \int \frac{d^*(y) dF(y)}{\widehat{\delta}(y)^2} \right] = 0 ,$$

which is verified if and only if  $\widehat{\delta}(x)/d^*(x)$  is a constant independent of  $x$ . The proportionality constant must be strictly positive because the covariance matrix estimator  $\widehat{\Sigma}_T$  is positive definite, as stated in Assumption 3.4. ■

### B.4 Proof of Proposition 4.1

Theorem 4 of Ledoit and Péché (2011) and the paragraphs immediately above it imply that

$$\frac{1}{N} \sum_{i=1}^N u'_{T,i} \Sigma_T u_{T,i} \xrightarrow{\text{a.s.}} \int d^*(x) dF(x) . \quad (\text{B.11})$$

It can be seen that the left-hand side of equation (B.11) is none other than  $N^{-1} \text{Tr}(\Sigma_T)$ . In addition, note that

$$\frac{1}{N} \text{Tr}(\widehat{\Sigma}_T) = \frac{1}{N} \sum_{i=1}^N \widehat{\delta}_T(\lambda_{T,i}) = \int \widehat{\delta}_T(x) dF_T(x) \xrightarrow{\text{a.s.}} \int \widehat{\delta}(x) dF(x) = \alpha \int d^*(x) dF(x) . \quad (\text{B.12})$$

Comparing equations (B.11) and (B.12) yields the desired result. ■

## C Consistent Estimator of the Stieltjes Transform $s(x)$

The estimation method developed by Ledoit and Wolf (2013b) is reproduced below solely for the sake of convenience. Interested readers are invited to consult the original paper for details. Note that Ledoit and Wolf (2013b) denote the number of variables by  $p$  rather than by  $N$  and the sample size by  $n$  rather than by  $T$ .

The key idea is to introduce a nonrandom multivariate function, called the *Quantized Eigenvalues Sampling Transform*, or QuEST for short, which discretizes, or *quantizes*, the relationship between  $F$ ,  $H$ , and  $c$  defined in equations (B.2) and (B.3). For any positive integers  $T$  and  $N$ , the QuEST function, denoted by  $Q_{T,N}$ , is defined as

$$Q_{T,N} : [0, \infty)^N \longrightarrow [0, \infty)^N$$

$$\mathbf{v} := (v_1, \dots, v_N)' \longmapsto Q_{T,N}(\mathbf{v}) := (q_{T,N}^1(\mathbf{v}), \dots, q_{T,N}^N(\mathbf{v}))',$$

where

$$\forall i = 1, \dots, N \quad q_{T,N}^i(\mathbf{v}) := N \int_{(i-1)/N}^{i/N} (F_{T,N}^{\mathbf{v}})^{-1}(u) du, \quad (\text{C.1})$$

$$\forall u \in [0, 1] \quad (F_{T,N}^{\mathbf{v}})^{-1}(u) := \sup\{x \in \mathbb{R} : F_{T,N}^{\mathbf{v}}(x) \leq u\}, \quad (\text{C.2})$$

$$\forall x \in \mathbb{R} \quad F_{T,N}^{\mathbf{v}}(x) := \lim_{\eta \rightarrow 0^+} \frac{1}{\pi} \int_{-\infty}^x \operatorname{Im} [s_{T,N}^{\mathbf{v}}(\xi + i\eta)] d\xi, \quad (\text{C.3})$$

and  $\forall z \in \mathbb{C}^+ \quad s := s_{T,N}^{\mathbf{v}}(z)$  is the unique solution in the set

$$\left\{ s \in \mathbb{C} : -\frac{T-N}{Tz} + \frac{N}{T} s \in \mathbb{C}^+ \right\} \quad (\text{C.4})$$

to the equation

$$s = \frac{1}{N} \sum_{i=1}^N \frac{1}{v_i \left(1 - \frac{N}{T} - \frac{N}{T} z s\right) - z}. \quad (\text{C.5})$$

It can be seen that equation (C.3) quantizes equation (B.1), that equation (C.4) quantizes equation (B.2), and that equation (C.5) quantizes equation (B.3). Thus,  $F_{T,N}^{\mathbf{v}}$  is the limiting distribution (function) of sample eigenvalues corresponding to the population spectral distribution (function)  $N^{-1} \sum_{i=1}^N \mathbb{1}_{[v_i, \infty)}$ . Furthermore, by equation (C.2),  $(F_{T,N}^{\mathbf{v}})^{-1}$  represents the inverse spectral distribution function, also known as the *quantile* function. By equation (C.1),  $q_{T,N}^i(\mathbf{v})$  can be interpreted as a ‘smoothed’ version of the  $(i-0.5)/N$  quantile of  $F_{T,N}^{\mathbf{v}}$ . Ledoit and Wolf (2013b) estimate the eigenvalues of the population covariance matrix simply by inverting the QuEST function numerically:

$$\hat{\tau}_T := \operatorname{argmin}_{\mathbf{v} \in (0, \infty)^N} \frac{1}{N} \sum_{i=1}^N [q_{T,N}^i(\mathbf{v}) - \lambda_{T,i}]^2. \quad (\text{C.6})$$

From this estimator of the population eigenvalues, Ledoit and Wolf (2013b) deduce an estimator of the Stieltjes transform  $s(x)$  as follows: for all  $x \in (0, \infty)$ , and also for  $x = 0$  in the case  $c > 1$ ,  $\hat{s}(x)$  is defined as the unique solution  $\hat{s} \in \mathbb{R} \cup \mathbb{C}^+$  to the equation

$$\hat{s} = - \left[ x - \frac{1}{T} \sum_{i=1}^N \frac{\hat{\tau}_{T,i}}{1 + \hat{\tau}_{T,i} \hat{s}} \right]^{-1}. \quad (\text{C.7})$$

## D Winsorization of Past Returns

Unusually large returns (in absolute value) can have undesirable impacts if such data are used to estimate a covariance matrix. We mitigate this problem by properly truncating very small and very large observations in any cross-sectional data set. Such truncation is commonly referred to as ‘Winsorization’, a method that is widely used by quantitative portfolio managers; for example, see Chincarini and Kim (2006, p.180).

Consider a set of numbers  $a_1, \dots, a_N$ . We first compute a robust measure of location that is not (heavily) affected by potential outliers. To this end we use the trimmed mean of the data with trimming fraction  $\eta \in (0, 0.5)$  on the left and on the right. This number is simply the mean of the middle  $(1 - 2\eta) \cdot 100\%$  of the data. More specifically, denote by

$$a_{(1)} \leq a_{(2)} \leq \dots \leq a_{(N)} \quad (\text{D.1})$$

the ordered data (from smallest to largest) and denote by

$$M := \lfloor \eta \cdot N \rfloor \quad (\text{D.2})$$

the smallest integer less than or equal to  $\eta \cdot N$ . Then the trimmed mean with trimming fraction  $\eta$  is defined as

$$\bar{a}_\eta := \frac{1}{N - 2M} \sum_{i=M+1}^{N-M} a_{(i)} . \quad (\text{D.3})$$

We employ the value of  $\eta = 0.1$  in practice.

We next compute a robust measure of spread. To this end, we use the mean absolute deviation (MAD) given by

$$\text{MAD}(a) := \frac{1}{N} \sum_{i=1}^N |a_i - \text{med}(a)| , \quad (\text{D.4})$$

where  $\text{med}(a)$  denotes the sample median of  $a_1, \dots, a_N$ .

We finally compute upper and lower bounds defined by

$$a_{lo} := \bar{a}_{0.1} - 5 \cdot \text{MAD}(a) \quad \text{and} \quad a_{up} := \bar{a}_{0.1} + 5 \cdot \text{MAD}(a) . \quad (\text{D.5})$$

The motivation here is that for a normally distributed sample, it will hold that  $\bar{a} \approx \bar{a}_{0.1}$  and  $s(a) \approx 1.5 \cdot \text{MAD}(a)$ , where  $\bar{a}$  and  $s(a)$  denote the sample mean and the sample standard deviation of  $a_1, \dots, a_N$ , respectively. As a result, for a ‘well-behaved’ sample, there will usually be no points below  $a_{lo}$  or above  $a_{up}$ . Our truncation rule is then that any data point  $a_i$  below  $a_{lo}$  will be changed to  $a_{lo}$  and any data point  $a_i$  above  $a_{up}$  will be changed to  $a_{up}$ . We apply this truncation rule, one day at a time, to the past stock return data used to estimate a covariance matrix. (Of course, we do not apply this truncation rule to *future* stock return data used to compute portfolio out-of-sample returns.)