

Ledoit, Olivier; Wolf, Michael

Working Paper

Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks

Working Paper, No. 137

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Ledoit, Olivier; Wolf, Michael (2014) : Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks, Working Paper, No. 137, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-90273>

This Version is available at:

<https://hdl.handle.net/10419/111195>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich** ^{UZH}

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

January 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

January 2014

Abstract

Markowitz (1952) portfolio selection requires estimates of (i) the vector of expected returns and (ii) the covariance matrix of returns. Many proposals to address the first question exist already. This paper addresses the second question. We promote a new 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 (that is, the *Goldilocks* principle). In a stylized setting, the nonlinear shrinkage estimator is asymptotically optimal for portfolio selection. In addition to theoretical analysis, we establish superior real-life performance of our new estimator using backtest exercises.

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

Recent proposals by Ledoit and Wolf (2003, 2004a,b), Kan and Zhou (2007), Brandt et al. (2009), DeMiguel et al. (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.² Our proposed solution can be divided into three steps.

The first step is to formulate a loss function that captures the out-of-sample variance of a mean-variance efficient portfolio. We use a stylized version of the Markowitz (1952) problem that abstracts from leverage constraints and short-sales constraints in order to focus on the core of the matter: how to minimize portfolio variance under the constraint that expected return meet a desired target. But the practical applications are more general than the stylized setting used for mathematical analysis. For example, the empirical results presented in Section 5 incorporate leverage constraints and still get favorable results.

The second step is to characterize the behavior of this loss function asymptotically as the

¹Sometimes 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.

²The 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 easy nor too tight but ‘just right’.

number of assets, N , and the sample size, T , go to infinity together, with their ratio, called the *concentration*, converging to a finite, nonzero limit. This framework is known as *large-dimensional asymptotics*. It is appropriate when the dimension of the investable universe is not negligible compared to the sample size. Letting the number of assets go to infinity is not new in finance, as it is an integral part of the Arbitrage Pricing Theory (APT) developed by Ross (1976). Characterizing the limit of the portfolio selection loss function requires a minimum of knowledge about the vector of expected returns. Since there are so many different signals explaining the cross-section, we adopt the most neutral stance possible in order to ensure a well-rounded performance.

The third step is to find a covariance matrix estimator that asymptotically minimizes the loss function within a certain class of dimension $O(N)$. In keeping with much of the multivariate statistics literature, we use the class of rotation-equivariant estimators introduced by Stein (1975, 1986). Rotation equivariance means that rotating the original set of variables makes the estimator rotate the same way. This is appropriate in the general case where we have no *a priori* information about the orientation of the eigenvectors of the population covariance matrix. In practice, this entails keeping the same eigenvectors as the sample covariance matrix, while modifying (also called *shrinking*) the sample eigenvalues nonlinearly in a completely free manner. As there are N eigenvalues, this class is of dimension N . Our contribution is the optimal such rotation-equivariant covariance matrix estimator for portfolio selection in large dimensions.

This new 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 and estimated consistently 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. Beyond finance, the portfolio selection method introduced in this paper represents a contribution to the widely-used technique of beamforming in signal processing (Du et al., 2010) and to optimal fingerprinting in global warming research (Ribes et al., 2009), as the underlying mathematics are exactly identical.

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 contains all figures, tables, and mathematical proofs, as well as an extension of the method to the case where the number of assets is larger than sample size.

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.

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

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

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

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

In practice, the covariance matrix Σ is not known and needs to be estimated from historical data. Let \tilde{S} denote a generic (invertible) estimator of the covariance matrix. The plug-in estimator of the optimal portfolio weights is

$$\tilde{w} := \frac{\sqrt{m'm}}{m'\tilde{S}^{-1}m} \times \tilde{S}^{-1}m. \quad (2.5)$$

All investing takes place *out of sample* by necessity. Since the population covariance matrix Σ is unknown and the covariance matrix estimator \tilde{S} 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 \tilde{S} and m :

$$\mathcal{L}(\tilde{S}, \Sigma, m) := \tilde{w}'\Sigma\tilde{w} = m'm \times \frac{m'\tilde{S}^{-1}\Sigma\tilde{S}^{-1}m}{(m'\tilde{S}^{-1}m)^2}. \quad (2.6)$$

3 Large-Dimensional Asymptotic Limit of the Loss Function

The basic description of the large-dimensional asymptotic framework is standard and is similar to the one provided by Ledoit and Wolf (2013a). Some of the assumptions made in Section 2 are restated more formally below for the sake of completeness.

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 $c \in (0, 1)$ called the concentration. Furthermore, there exists a compact interval included in $(0, 1)$ 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. The assumption $c < 1$ is for ease of exposition only. 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. This extension is presented in Appendix D.

Assumption 3.2 (Population Covariance Matrix). *The population covariance matrix Σ_T is a nonrandom symmetric positive-definite matrix of dimension N . Let $\tau_T := (\tau_{T,1}, \dots, \tau_{T,N})'$ denote a system of eigenvalues of Σ_T sorted in increasing order. The empirical distribution function (e.d.f.) of the 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). $\text{Supp}(H)$, the support of H , is the union of a finite number of closed intervals, bounded away from zero and infinity. Furthermore, 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.

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 Σ_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, Monte Carlo simulations reported in Table 2 of Ledoit and Wolf (2013b) suggest that it is not needed in practice and can be replaced by a bound on the second moment.

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 $S_T = U_T\Lambda_TU_T'$, where Λ_T is a diagonal matrix, and U_T is an orthogonal matrix: $U_TU_T' = U_T'U_T = \mathbb{I}_T$, 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 continuously differentiable limiting sample spectral distribution F such that

$$\forall x \in \mathbb{R} \quad F_T(x) \xrightarrow{\text{a.s.}} F(x). \quad (3.1)$$

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: $\text{Supp}(F) = \bigcup_{k=1}^{\kappa} [a_k, b_k]$, where $0 < a_1 < b_1 < \dots < a_{\kappa} < b_{\kappa} < \infty$.

Assumption 3.4 (Class of Estimators). We consider positive-definite covariance matrix estimators of the type $\tilde{S}_T := U_T\tilde{D}_TU_T'$, where \tilde{D}_T is a diagonal matrix: $\tilde{D}_T := \text{Diag}(\tilde{d}_T(\lambda_{T,1}), \dots, \tilde{d}_T(\lambda_{T,N}))$, and \tilde{d}_T is a real univariate function which can depend on S_T . We assume that there exists a nonrandom real univariate function \tilde{d} defined on $\text{Supp}(F)$ and continuously differentiable such that $\tilde{d}_T(x) \xrightarrow{\text{a.s.}} \tilde{d}(x)$, 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 \tilde{K} such that almost surely, over the set $x \in \bigcup_{k=1}^{\kappa} [a_k - \eta, b_k + \eta]$, $\tilde{d}_T(x)$ is uniformly bounded by \tilde{K} from above and by $1/\tilde{K}$ from below, 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.

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

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

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

We call \tilde{d}_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 \tilde{S}_T are fully characterized by its limiting shrinkage function \tilde{d} .

Estimators in the class defined by Assumption 3.4 are evaluated according to the limit as T and N go to infinity together of the loss function defined in equation (2.6). For this limit to exist, some assumption on the return predictive signal is required.³ 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 \tilde{S}_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 proposition, 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, $i := \sqrt{-1}$.

Proposition 3.1. Under Assumptions 3.1–3.5,

$$m'_T m_T \times \frac{m'_T \tilde{S}_T^{-1} \Sigma_T \tilde{S}_T^{-1} m_T}{\left(m'_T \tilde{S}_T^{-1} m_T\right)^2} \xrightarrow{\text{a.s.}} \frac{\int \frac{dF(x)}{x |s(x)|^2 \tilde{d}(x)^2}}{\left[\int \frac{dF(x)}{\tilde{d}(x)} \right]^2}, \quad (3.3)$$

³The return predictive signal can be interpreted as an estimator of the vector of expected returns, which is not available in practice.

The challenge is to construct an estimator of the covariance matrix whose limiting shrinkage function satisfies equation (4.2) with $\alpha = 1$. The problem is that the function $s(x)$ is not directly observable. As can be seen from equation (3.4), it depends on the limiting distribution of population eigenvalues H . Thus, a consistent estimator of the eigenvalues of the population covariance matrix Σ_T is needed. Fortunately, a recent breakthrough in the literature on large-dimensional asymptotics has delivered just such an estimator. It is worth emphasizing an important point: Under large-dimensional asymptotics, that is, when the number of stocks is not small relative to sample size, it is impossible to estimate the population covariance matrix consistently because the estimation problem is $O(N^2)$. But it is possible to estimate the eigenvalues of the population covariance matrix consistently because the estimation problem is $O(N)$ only. This result is established formally by Ledoit and Wolf (2012, 2013b). Of the two papers, we prefer to borrow the technique of the latter, because it also accomodates the singular case $N > T$, which arises often in finance (cf. Appendix D). The method of Ledoit and Wolf (2013b) for constructing a consistent estimator of population eigenvalues is reproduced in Appendix C for convenience. The exact sense in which this estimator is consistent is stated in the following proposition.

Proposition 4.3 (Ledoit and Wolf, 2013b; Theorem 2.2). *Suppose that Assumptions 3.1–3.3 are satisfied. Let $\hat{\tau}_T := (\hat{\tau}_{T,1}, \dots, \hat{\tau}_{T,N})'$ denote the estimator of population eigenvalues of Ledoit and Wolf (2013b), as re-stated in Appendix C, with eigenvalues sorted in increasing order. As per Assumption 3.2, $\tau_T := (\tau_{T,1}, \dots, \tau_{T,N})'$ denotes a system of eigenvalues of the population covariance matrix Σ_T sorted in increasing order. Then*

$$\frac{1}{N} \sum_{i=1}^N (\hat{\tau}_{T,i} - \tau_{T,i})^2 \xrightarrow{\text{a.s.}} 0 .$$

An asymptotically optimal estimator of the covariance matrix is obtained simply by inserting the population eigenvalues estimator of Proposition 4.3 into equation (3.4).

Theorem 4.1. *Suppose that Assumptions 3.1–3.5 are satisfied, and let $\hat{\tau}_T := (\hat{\tau}_{T,1}, \dots, \hat{\tau}_{T,N})'$ denote the population eigenvalues estimator of Ledoit and Wolf (2013b) as restated in Appendix C. Define the covariance matrix estimator*

$$\hat{S}_T := U_T \hat{D}_T U_T' \quad \text{where} \quad \hat{D}_T := \text{Diag}(\hat{d}_T(\lambda_{T,1}), \dots, \hat{d}_T(\lambda_{T,N})) , \quad (4.4)$$

$$\forall i = 1, \dots, N \quad \hat{d}_T(\lambda_{T,i}) := \frac{1}{\lambda_{T,i} |\hat{s}(\lambda_{T,i})|^2} , \quad (4.5)$$

and, for all $x \in (0, \infty)$, $\hat{s}(x)$ is defined as the unique solution $\hat{s} \in \mathbb{R} \cup \mathbb{C}^+$ to the equation

$$\hat{s} = - \left[x - \frac{1}{T} \sum_{i=1}^N \frac{\hat{\tau}_{T,i}}{1 + \hat{\tau}_{T,i} \hat{s}} \right]^{-1} . \quad (4.6)$$

Then \hat{S}_T 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.

Theorem 4.1 formulates a covariance matrix estimator that is optimal for portfolio selection when the investment universe is large, within a much broader class of estimators than previously considered. In particular, its shrinkage function (4.5) is of a highly nonlinear nature, unlike the linear shrinkage function (3.2) of the estimator of Ledoit and Wolf (2004b). The next section provides empirical evidence that this more flexible, nonlinear estimation strategy indeed generates portfolios with superior out-of-sample properties.

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”.⁴ 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.⁵ 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.

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

⁴ “Monthly” updating is common practice to avoid an unreasonable amount of turnover, and thus transaction costs.

⁵ 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.

details can be found in Appendix E. Of course, we always use the ‘raw’, non-winsorized data in computing the out-of-sample portfolio returns.

Remark 5.1 (Real Returns versus Log Returns). For forming the portfolios, we use real returns. This is because for real returns, unlike for log returns, the return of a portfolio is given by the ‘portfolio’ of the returns (that is, by the corresponding linear combination of the individual stock returns).

On the other hand, for the evaluation of the realized out-of-sample returns, we use log returns. This is because the log returns more closely reflect the out-of-sample performance of a portfolio. For example, the average of the log returns corresponds one-to-one to the net asset value of a portfolio (with initial investment \$1, say) over the investment period, while is this not true for the average of the real returns. ■

5.2 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

$$\min_w w' \Sigma w \quad (5.1)$$

$$\text{subject to } w' \mathbb{1} = 1, \quad (5.2)$$

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

$$w = \frac{\Sigma^{-1} \mathbb{1}}{\mathbb{1}' \Sigma^{-1} \mathbb{1}}. \quad (5.3)$$

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

$$\tilde{w} := \frac{\tilde{S}^{-1} \mathbb{1}}{\mathbb{1}' \tilde{S}^{-1} \mathbb{1}}. \quad (5.4)$$

Alternative strategies, motivated by estimating the optimal w of (5.3) ‘directly’, as opposed to ‘indirectly’ via the estimation of Σ , have been proposed recently by Frahm and Memmel (2010).

Estimating the GMV portfolio is a ‘clean’ problem in terms of evaluating the quality of a covariance matrix estimator, since it abstracts from having to estimate the vector of expected returns at the same time. In addition, researchers have established that estimated GMV portfolios have desirable out-of-sample properties not only in terms of risk but also in terms of reward-to-risk (that is, in terms of the information 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 five portfolios are included in the study.

- **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 \tilde{S} is given by the $N \times N$ identity matrix. This strategy avoids any parameter estimation whatsoever.
- **Sample**: The portfolio (5.4) where \tilde{S} 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 $N > T$.
- **Lin**: The portfolio (5.4) where \tilde{S} is given by the linear shrinkage estimator of Ledoit and Wolf (2004b).
- **NonLin**: The portfolio (5.4) where \tilde{S} is given by the estimator \hat{S} of Theorem 4.1, with the modification of Appendix D when $N > T$.

Remark 5.2 (Alternative $O(1)$ Proposals). As mentioned in the introduction, there have been many recent proposals based on $O(1)$ degrees of freedom. We could not possibly include them all in the empirical study and restrict attention to the linear shrinkage estimator of Ledoit and Wolf (2004b). There are two reasons for this particular choice. First, since it also belongs to class of rotation-equivariant estimators \tilde{S} considered, it is the most natural $O(1)$ competitor to our $O(N)$ nonlinear shrinkage estimator; see Remark 3.1. Second, no other $O(1)$ GMV portfolio has been shown to consistently outperform Lin. For example, DeMiguel et al. (2009a) find that their NC2V norm-constrained portfolio beats Lin in only one out five real-life data sets. ■

Remark 5.3 (Alternative Linear Shrinkage Estimators). Alternatively, the GMV portfolio estimator Lin could also be based on the linear shrinkage estimators \tilde{S} proposed by Ledoit and Wolf (2003, 2004a). But these two estimators do not belong to the class of rotation-equivariant estimators considered and, therefore, are less natural competitors to our nonlinear shrinkage estimator.

Furthermore, several previous studies have compared the three linear shrinkage estimators of Ledoit and Wolf (2003, 2004a,b) without being able to find a clear winner among them; for example, see Ledoit and Wolf (2004a) and DeMiguel et al. (2009a). ■

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

- **AV**: We compute the average of the 10,080 out-of-sample log returns and then multiply by 250 to annualize.
- **SD**: We compute the standard deviation of the 10,080 out-of-sample log returns and then multiply by $\sqrt{250}$ to annualize.

- **IR:** We compute the (annualized) information ratio as the ratio AV/SD .⁶

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 information 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 information ratio, IR , 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 five portfolios, there are ten 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).⁷

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 1.5 percentage points. On the other hand, the standard deviations of Lin and $Nonlin$ decrease by 3.4 and 4.2 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.

The same is not true for the portfolios $Sample$ and FM , though, whose standard deviations decrease by only 0.6 and 1.1 percentage points, respectively, as N increases from $N = 30$ to $N = 100$. Both portfolios break down for $N = 250$, when the sample covariance matrix is nearly singular.

- $1/N$ is consistently outperformed by Lin and $NonLin$ in terms of both the standard deviation and the information ratio. (There is the single case $N = 100$ where the information ratio of $1/N$ is barely higher than the information ratio of Lin .)
- $Sample$ and FM outperform $1/N$ for $N = 30, 50, 100$ in terms of the standard deviation but only for $N = 30, 50$ in terms of the information ratio.
- Although FM improves over $Sample$ in terms of both the standard deviation and the information ratio, it is itself consistently outperformed by Lin and $NonLin$.

⁶This version of the information ratio, which simply uses zero as the benchmark, is widely used in the mutual fund industry.

⁷Since 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.

- NonLin consistently outperforms Lin both in terms of the standard deviation and the information ratio. (There is the single case $N = 30$ where the information ratio of Lin is barely higher than the information ratio of NonLin.) The outperformance in terms of the standard deviation is significant at the 0.01 level for $N \geq 50$. The outperformance is also economically meaningful for $N = 100, 250, 500$, ranging from 0.3 to 0.9 percentage points.

In addition, all realized out-of-sample log returns are graphically displayed via boxplots in Figure 1. One can see that Lin and NonLin generally deliver the tightest distributions.

Summing up, in the global minimum variance portfolio problem, NonLin dominates the remaining four portfolios in terms of the standard deviation, and generally also in terms of the information ratio.

5.3 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 .

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

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

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

$$w' \mathbf{1} = 1, \quad (5.7)$$

where b is a selected target expected return. The problem has the analytical solution

$$w = c_1 \Sigma^{-1} \mathbf{1} + c_2 \Sigma^{-1} m, \quad (5.8)$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2}, \quad (5.9)$$

$$\text{with } A := \mathbf{1}' \Sigma^{-1} \mathbf{1}, \quad B := \mathbf{1}' \Sigma^{-1} m, \quad \text{and} \quad C := m' \Sigma^{-1} m. \quad (5.10)$$

The natural strategy in practice is to replace the unknown Σ by an estimator \tilde{S} in formulas (5.8)–(5.10), yielding a feasible portfolio

$$\tilde{w} := c_1 \tilde{S}^{-1} \mathbf{1} + c_2 \tilde{S}^{-1} m, \quad (5.11)$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2}, \quad (5.12)$$

$$\text{with } A := \mathbf{1}' \tilde{S}^{-1} \mathbf{1}, \quad B := \mathbf{1}' \tilde{S}^{-1} m, \quad \text{and} \quad C := m' \tilde{S}^{-1} m. \quad (5.13)$$

The following five 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 the remaining four portfolios 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).

- **Sample:** The portfolio (5.11)–(5.13) where \tilde{S} 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.
- **BSV:** The portfolio (5.11)–(5.13) where \tilde{S} is given by the identity matrix of dimension $N \times N$. This portfolio corresponds to the proposal of Brandt et al. (2009).
- **Lin:** The portfolio (5.11)–(5.13) where \tilde{S} is given by the linear shrinkage estimator of Ledoit and Wolf (2004b).
- **NonLin:** The portfolio (5.11)–(5.13) where \tilde{S} is given by the estimator \hat{S} of Theorem 4.1, with the modification of Appendix D when $N > T$.

Our stance is that in the context of a ‘full’ Markowitz portfolio, the most important performance measure is the out-of-sample information ratio, IR. In the ‘ideal’ investment problem (5.8)–(5.10), minimizing the variance (for a fixed target expected return b) is equivalent to maximizing the information 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 information ratio than another portfolio at a level that is statistically significant. Since we consider five portfolios, there are ten 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 information ratios is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2008, Section 3.1).⁸

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

- The information ratio of the ‘ideal’ Markowitz portfolio (5.8)–(5.10) is expected to increase in N , as the potential for diversification increases. So the same should be true for any good estimator of the Markowitz portfolio. Of the five portfolios considered, only Lin and NonLin have this desirable property.
- We again observe that Sample breaks down for $N = 250$, when the sample covariance matrix is close to singular.

⁸Since 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.

- The order, from worst to best, of the remaining four portfolios is EW-TQ, BSV, Lin, and NonLin for all values of N .
- The outperformance of NonLin over Lin is significant at the 0.05 level for $N = 50$ and significant at the 0.01 level for $N = 100, 250, 500$. The outperformance is also economically meaningful for $N = 100, 250, 500$, ranging from 0.07 to 0.19.

In addition, all realized out-of-sample log returns are graphically displayed via boxplots in Figure 2. One can see that Lin and NonLin generally deliver the tightest distributions.

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

Remark 5.4. Let us point out that we get higher information ratios here compared to the GMV portfolios across the board for $N \geq 50$. Therefore, using a return predictive signal can really pay off, if done properly. ■

5.4 Robustness Checks

5.4.1 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 3–5. It can be seen that NonLin is better than Lin in terms of the standard deviation in 14 out of the 15 cases, and generally with statistical significance. (The single case in which Lin is better than NonLin, namely the case $N = 20$ in the second subperiod, is significant at the 0.1 level only.) We also note that NonLin is better than Lin in terms of the information ratio in 13 out of the 15 cases.

The results for the Markowitz portfolio with momentum signal are presented in Tables 6–8. It can be seen that NonLin is better than Lin in terms of the information ratio in all 15 cases; though, with one exception, statistical significance only obtains in the first subperiod.

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

5.4.2 Rolling-Window Analysis

As a further comparison, we carry out a rolling-window analysis. Starting in 12/10/1982, we use the most recent 2,500 out-of-sample log-returns to compute (annualized) standard deviations of the GMV portfolios and (annualized) information ratios for the Markowitz portfolios with momentum signal. We repeat this exercise every subsequent trading day, ending in 12/31/2011. For each scenario, this results in 7,581 “ten-year” rolling-window numbers, which are then displayed in time series plots. To avoid a visual overload we only include Lin and NonLin in the figures.

Figure 3 presents the time series plots for the standard deviations of the GMV portfolios. It can be seen that NonLin generally lies below Lin and that the differences tend to increase in N .

Figure 4 presents the time series plots for the information ratios of the Markowitz portfolios with momentum signal. It can be seen that NonLin generally lies above Lin and that the differences tend to increase in N .

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. 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.

A Figures and Tables

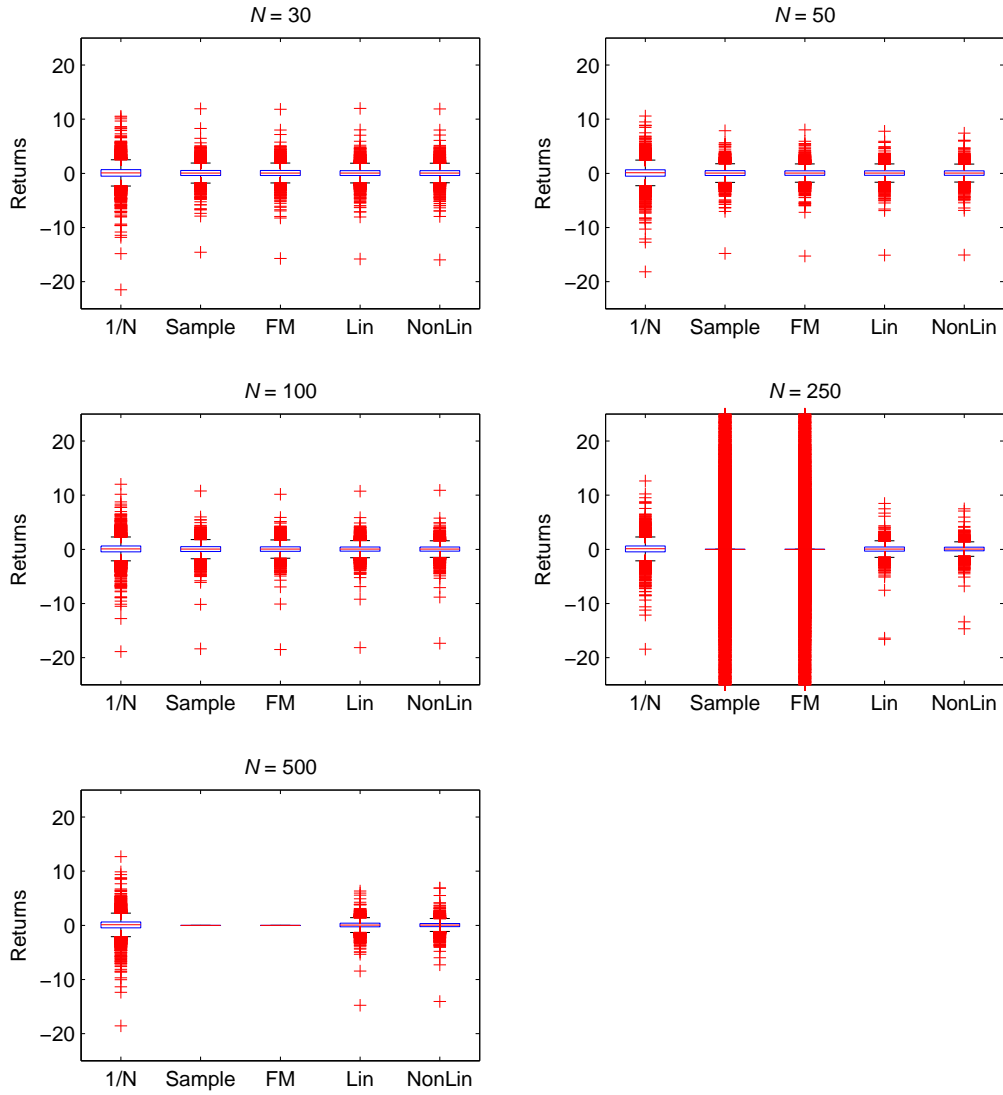


Figure 1: Boxplots of the 10,080 out-of-sample log returns (in percent) of the various GMV portfolios. The investment period is 01/19/1973–12/31/2011. For $N = 250$, Sample and FM produce some very large returns (in absolute value) and their boxplots extend beyond the limits of the vertical axis.

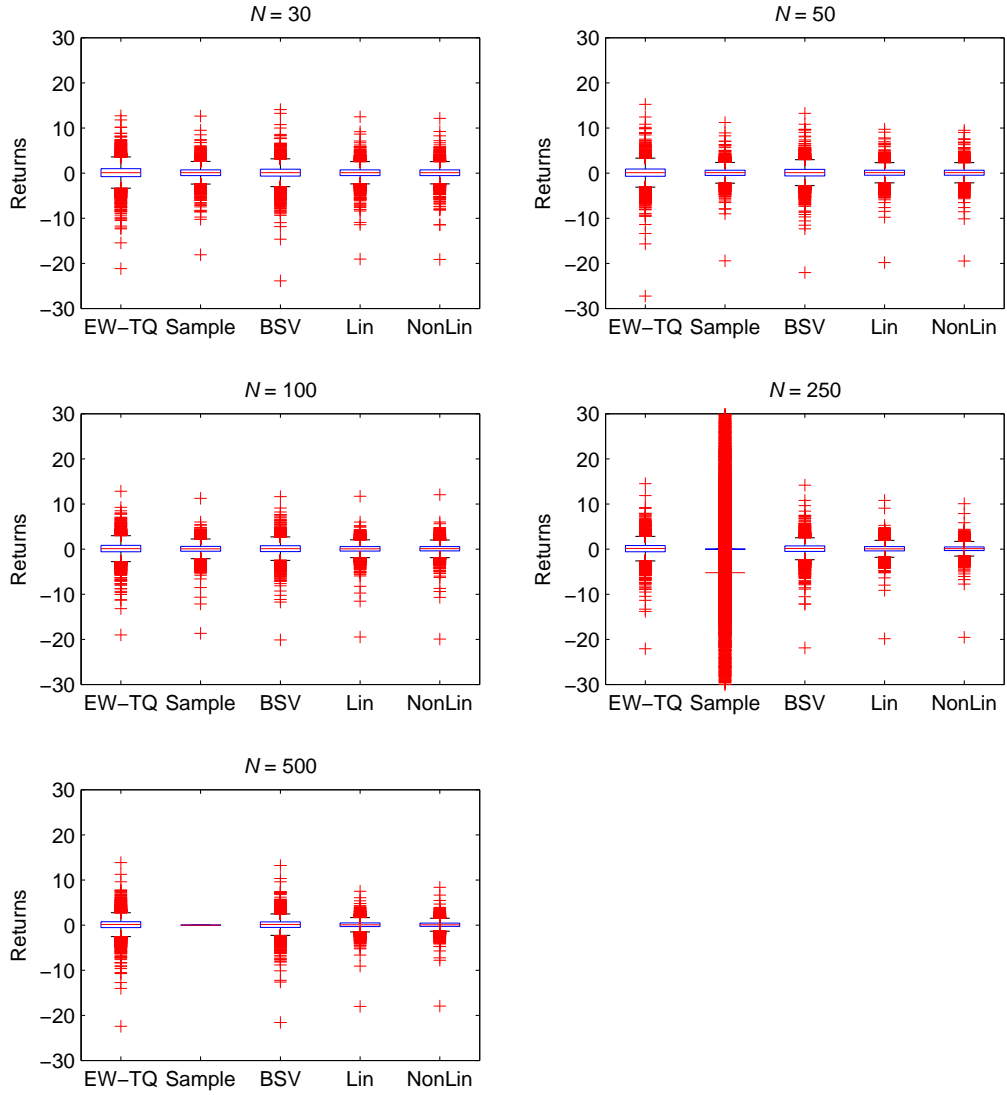


Figure 2: Boxplots of the 10,080 out-of-sample log returns (in percent) of the various Markowitz portfolios with momentum signal. The investment period is 01/19/1973–12/31/2011. For $N = 250$, Sample produces some very large returns (in absolute value) and its boxplot extends beyond the limits of the vertical axis.

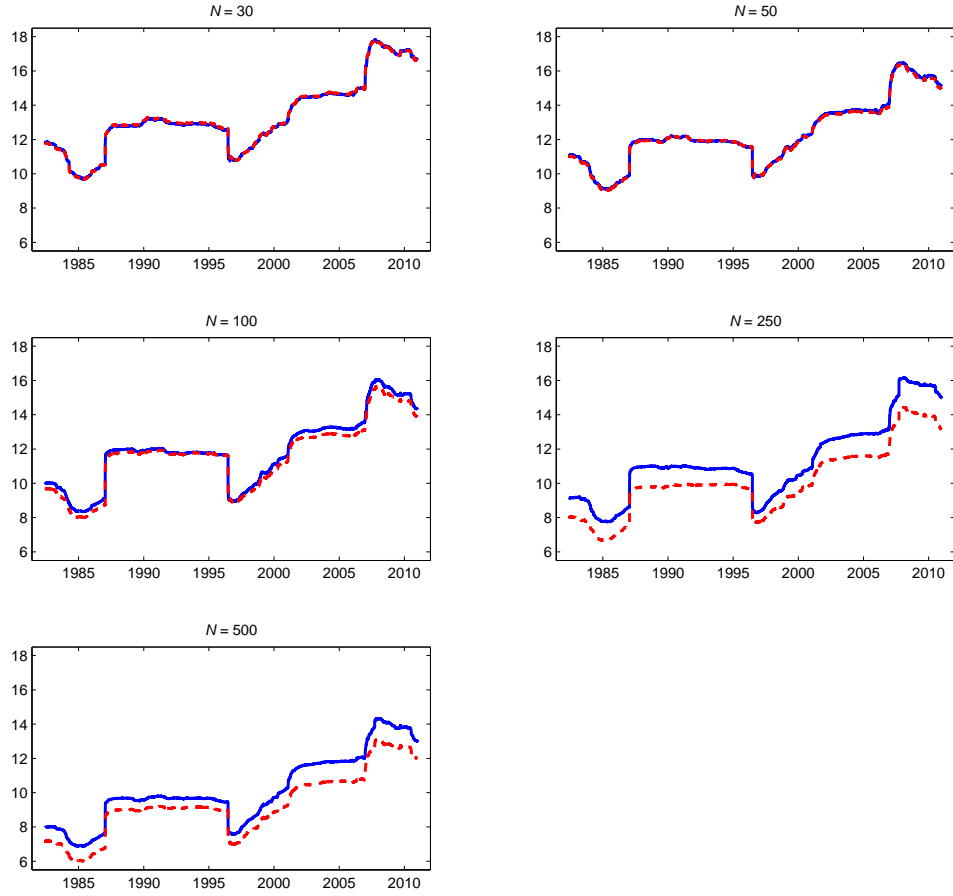


Figure 3: Annualized rolling-window standard deviations of the most recent 2,500 out-of-sample log returns (in percent) for the GMV portfolios based on linear shrinkage (solid lines) and nonlinear shrinkage (dashed lines). The investment period is 01/19/1973–12/31/2011 and the interval of the reported time series is 12/10/1982–12/31/2011.

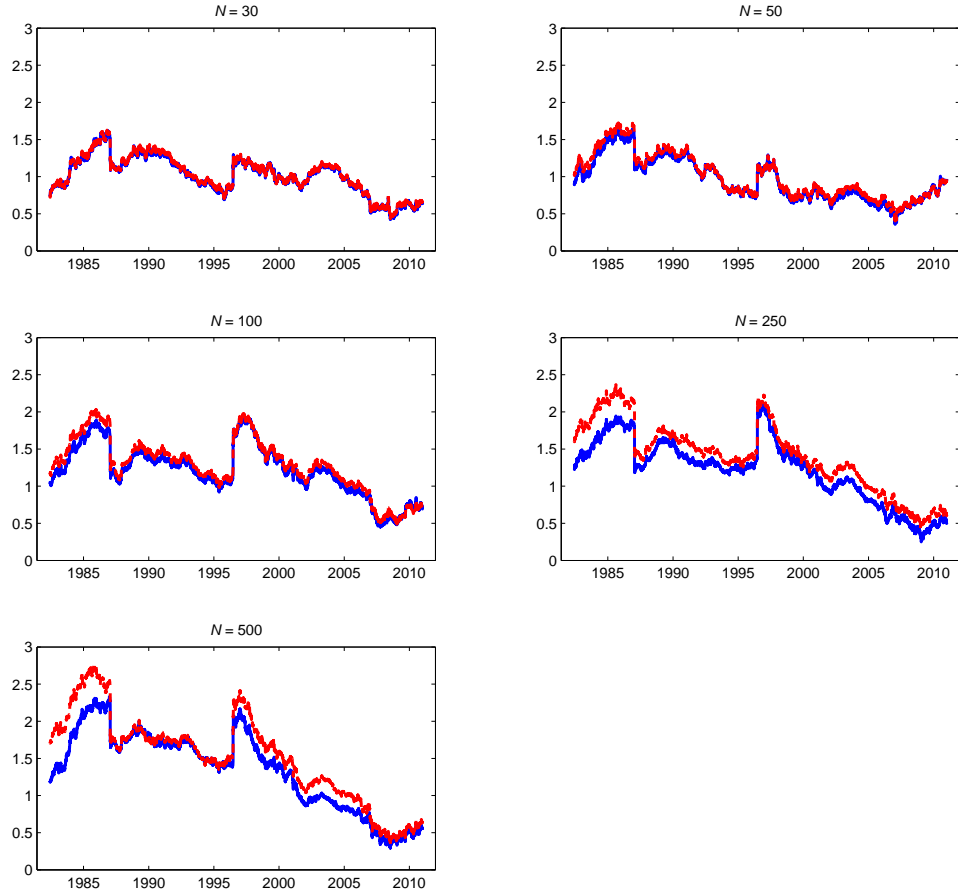


Figure 4: Annualized rolling-window information ratios of the most recent 2,500 out-of-sample log returns (in percent) for the Markowitz portfolios with momentum signal based on linear shrinkage (solid lines) and nonlinear shrinkage (dashed lines). The investment period is 01/19/1973–12/31/2011 and the interval of the reported time series is 12/10/1982–12/31/2011.

Period: 01/19/1973–12/31/2011					
	1/N	Sample	FM	Lin	NonLin
$N = 30$					
AV	15.19	10.61	11.01	11.26	11.22
SD	20.59	14.15	14.08	13.97	13.96
IR	0.74	0.75	0.78	0.81	0.80
$N = 50$					
AV	14.97	10.12	10.50	10.80	10.93
SD	19.74	13.35	13.13	12.89	12.81 ***
IR	0.76	0.76	0.80	0.84	0.85
$N = 100$					
AV	14.55	7.93	8.71	9.14	9.71
SD	19.36	13.53	13.03	12.29	11.99 ***
IR	0.75	0.57	0.67	0.74	0.81
$N = 250$					
AV	14.68	−86.81	−86.81	9.17	10.55
SD	19.23	1,014.13	1,1014.13	11.94	10.65 ***
IR	0.76	−0.09	−0.09	0.77	0.99
$N = 500$					
AV	14.58	NA	NA	10.17	10.19
SD	19.13	NA	NA	10.64	9.74 ***
IR	0.76	NA	NA	0.95	1.05

Table 1: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 10,080 daily out-of-sample log returns 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	Sample	BSV	Lin	NonLin
$N = 30$					
AV	15.40	14.61	16.82	14.82	14.93
SD	25.94	19.26	23.11	18.93	18.92
IR	0.59	0.76	0.73	0.78	0.79
$N = 50$					
AV	19.26	14.31	18.06	15.19	15.71
SD	24.83	17.14	22.26	16.66	16.59
IR	0.78	0.83	0.81	0.91	0.95**
$N = 100$					
AV	19.53	14.18	18.63	15.36	15.98
SD	23.13	15.98	20.96	14.69	14.39
IR	0.86	0.87	0.91	1.02	1.09**
$N = 250$					
AV	18.56	−9.21	17.47	13.73	15.02
SD	22.12	743.03	20.15	14.24	12.96
IR	0.84	−0.01	0.87	0.96	1.15***
$N = 500$					
AV	18.79	NA	17.33	13.11	14.28
SD	21.79	NA	19.83	12.39	11.45
IR	0.86	NA	0.87	1.06	1.25***

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 IR stands for information ratio. All measures are based on 10,080 daily out-of-sample log returns from 01/19/1973 until 12/31/2011. In the rows labeled IR, 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 IR 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					
	1/N	Sample	FM	Lin	NonLin
$N = 30$					
AV	17.43	14.11	14.57	14.70	14.67
SD	15.43	11.44	11.32	11.27	11.21 ***
IR	1.13	1.23	1.29	1.30	1.31
$N = 50$					
AV	16.77	12.22	12.50	13.08	13.81
SD	14.97	10.83	10.65	10.55	10.47 **
IR	1.12	1.13	1.17	1.24	1.31
$N = 100$					
AV	15.32	10.05	10.50	11.20	12.20
SD	14.50	10.48	10.09	9.51	9.17 ***
IR	1.06	0.96	1.04	1.18	1.33
$N = 250$					
AV	15.60	−334.94	−334.94	11.47	12.80
SD	14.23	849.48	849.48	8.77	7.72 ***
IR	1.10	−0.39	−0.39	1.31	1.66
$N = 500$					
AV	15.40	NA	NA	12.45	13.46
SD	14.16	NA	NA	7.72	6.94 ***
IR	1.09	NA	NA	1.61	1.94

Table 3: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 3,360 daily out-of-sample log returns from 01/19/1973 until 05/08/1986. In the rows labeled SD, the lowest 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 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	Lin	NonLin
$N = 30$					
AV	18.83	11.28	11.96	12.11	12.29
SD	16.91	13.01	13.01	12.91*	12.97
IR	1.11	0.87	0.92	0.94	0.95
$N = 50$					
AV	16.94	12.34	12.89	12.02	12.00
SD	15.94	12.43	12.24	12.02	12.00
IR	1.06	0.99	1.05	1.07	1.02
$N = 100$					
AV	17.92	11.64	12.37	12.15	12.50
SD	15.51	12.77	12.26	11.66	11.55*
IR	1.15	0.92	1.01	1.04	1.08
$N = 250$					
AV	18.01	510.04	510.04	12.50	12.93
SD	15.39	1,163.12	1,163.12	10.89	9.88***
IR	1.17	0.44	0.44	1.15	1.31
$N = 500$					
AV	18.09	NA	NA	14.53	13.30
SD	15.30	NA	NA	9.70	9.13***
IR	1.18	NA	NA	1.50	1.46

Table 4: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 3,360 daily out-of-sample log returns 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	Lin	NonLin
$N = 30$					
AV	9.30	6.44	6.51	6.96	6.70
SD	27.35	17.34	17.23	17.08	17.06
IR	0.34	0.37	0.38	0.41	0.39
$N = 50$					
AV	11.21	5.81	6.13	6.34	6.69
SD	26.29	16.21	15.92	15.57	15.44**
IR	0.43	0.36	0.38	0.41	0.43
$N = 100$					
AV	10.40	2.11	3.27	4.06	4.43
SD	25.96	16.62	16.04	15.05	14.60***
IR	0.40	0.13	0.20	0.27	0.30
$N = 250$					
AV	10.42	−435.38	−435.38	3.56	5.92
SD	25.90	1,003.23	1,003.23	15.23	13.52***
IR	0.40	−0.43	−0.43	0.23	0.43
$N = 500$					
AV	10.25	NA	NA	3.49	3.80
SD	25.75	NA	NA	13.62	12.38***
IR	0.40	NA	NA	0.26	0.31

Table 5: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 3,360 daily out-of-sample log returns 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	Sample	BSV	Lin	NonLin
$N = 30$					
AV	14.28	14.67	16.17	14.65	14.71
SD	20.29	14.86	18.14	14.63	14.60
IR	0.70	0.99	0.89	1.00	1.01
$N = 50$					
AV	20.81	14.58	19.21	15.49	16.85
SD	19.51	13.68	17.60	13.33	13.25
IR	1.07	1.07	1.09	1.16	1.27***
$N = 100$					
AV	20.05	16.03	21.05	16.95	17.66
SD	17.79	13.00	16.21	11.86	11.47
IR	1.13	1.23	1.30	1.43	1.54*
$N = 250$					
AV	19.78	15.43	19.75	15.51	17.61
SD	16.99	577.29	15.60	10.32	9.14
IR	1.16	0.03	1.27	1.50	1.93***
$N = 500$					
AV	20.24	NA	20.20	14.80	17.23
SD	16.63	NA	15.36	9.01	8.11
IR	1.21	NA	1.32	1.64	2.12***

Table 6: 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 IR stands for information ratio. All measures are based on 3,360 daily out-of-sample returns from 01/19/1973 until 05/08/1986. In the rows labeled IR, the largest 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 IR 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	Sample	BSV	Lin	NonLin
$N = 30$					
AV	24.38	20.04	22.73	20.42	20.85
SD	25.18	18.40	22.76	18.31	18.34
IR	0.97	1.09	1.00	1.12	1.14
$N = 50$					
AV	23.74	12.78	21.18	13.70	13.85
SD	23.34	17.06	20.72	16.66	16.60
IR	1.02	0.75	1.02	0.82	0.84
$N = 100$					
AV	25.12	18.08	22.58	18.87	19.55
SD	20.60	15.97	18.63	14.84	14.77
IR	1.22	1.13	1.21	1.27	1.32
$N = 250$					
AV	23.91	41.14	20.43	17.10	17.46
SD	20.02	669.21	18.26	13.66	12.89
IR	1.19	0.06	1.12	1.25	1.35
$N = 500$					
AV	23.97	NA	20.35	16.98	17.15
SD	19.73	NA	17.91	11.63	11.22
IR	1.21	NA	1.13	1.46	1.53

Table 7: 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 IR stands for information ratio. All measures are based on 3,360 daily out-of-sample returns from 05/09/1986 until 08/25/1999. In the rows labeled IR, the largest 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 IR 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	Sample	BSV	Lin	NonLin
$N = 30$					
AV	6.71	12.64	7.02	12.74	12.91
SD	32.14	22.17	28.27	21.87	21.84
IR	0.21	0.57	0.25	0.58	0.59
$N = 50$					
AV	13.25	15.57	13.78	16.38	16.43
SD	30.40	20.07	27.34	19.44	19.35
IR	0.44	0.78	0.50	0.84	0.85
$N = 100$					
AV	13.41	8.42	12.25	10.26	10.73
SD	28.11	19.43	25.38	17.76	17.30
IR	0.48	0.43	0.48	0.58	0.62
$N = 250$					
AV	12.00	−84.23	12.24	8.58	9.98
SD	27.90	935.07	25.33	17.75	15.94
IR	0.43	−0.09	0.48	0.48	0.63*
$N = 500$					
AV	12.17	NA	11.45	7.54	8.45
SD	27.53	NA	24.97	15.61	14.19
IR	0.44	NA	0.46	0.48	0.59

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 IR stands for information ratio. All measures are based on 3,360 daily out-of-sample returns from 08/26/1999 until 12/31/2011. In the rows labeled IR, the largest 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 IR 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.

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 $\check{s}_F(x) := \lim_{z \in \mathbb{C}^+ \rightarrow x} s_F(z)$ which is continuous over $x \in \mathbb{R}$. Furthermore, 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 $n^{-1}Y_n Y'_n = n^{-1}X_n \Sigma_n X'_n$ as \underline{F} ; note that the eigenvalues of $n^{-1}Y'_n Y_n$ and $n^{-1}Y_n Y'_n$ only differ by $|n - p|$ zero eigenvalues. It then holds:

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

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

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

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

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 now): $\forall x \in \mathbb{R} \setminus \{0\}$, $\check{s}_{\underline{F}}(x) := \lim_{z \in \mathbb{C}^+ \rightarrow x} s_{\underline{F}}(z)$ exists. Furthermore, the function $\check{s}_{\underline{F}}$ is continuous over $\mathbb{R} \setminus \{0\}$.

It can easily be verified that the function $s(x)$ defined in equation (3.4) is in fact none other than $\check{m}_{\underline{F}}(x)$. Equation (4.2.2) of Bai and Silverstein (2010), for example, gives an expression analogous to equation (3.4). Based on the right-hand side of equation (3.3), we introduce the function

$$\forall x \in \text{Supp}(F) \quad d^*(x) := \frac{1}{x |s(x)|^2} = \frac{1}{x |\check{s}_{\underline{F}}(x)|^2} = \frac{x}{|1 - c - c x \check{s}_{\underline{F}}(x)|^2} . \quad (\text{B.8})$$

B.2 Proof of Proposition 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  ch   (2011) are satisfied. This implies that there exists a constant K_1 independent of T , \tilde{S}_T and m_T such that

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

Note that $\|\tilde{S}_T^{-1}\| \leq \tilde{K}/\underline{h}$ a.s. for large enough T by Assumption 3.4. Therefore,

$$\frac{1}{N} m'_T \tilde{S}_T^{-1} m_T - \frac{1}{N} \text{Tr}(\tilde{S}_T^{-1}) \xrightarrow{\text{a.s.}} 0 .$$

In addition, we have

$$\frac{1}{N} \text{Tr}(\tilde{S}_T^{-1}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\tilde{d}_T(\lambda_{T,i})} = \int \frac{1}{\tilde{d}_T(x)} dF_T(x) \xrightarrow{\text{a.s.}} \int \frac{1}{\tilde{d}(x)} dF(x) .$$

Therefore,

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

A similar line of reasoning leads to

$$\frac{1}{N} m'_T \tilde{S}_T^{-1} \Sigma_T \tilde{S}_T^{-1} m_T - \frac{1}{N} \text{Tr}(\tilde{S}_T^{-1} \Sigma_T \tilde{S}_T^{-1}) \xrightarrow{\text{a.s.}} 0 .$$

Notice that

$$\frac{1}{N} \text{Tr}(\tilde{S}_T^{-1} \Sigma_T \tilde{S}_T^{-1}) = \frac{1}{N} \text{Tr}(U'_T \Sigma_T U_T \tilde{D}_T^{-2}) = \frac{1}{N} \sum_{i=1}^N \frac{u'_{T,i} \Sigma_T u_{T,i}}{\tilde{d}_T(\lambda_{T,i})^2} .$$

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

$$\frac{1}{N} \sum_{i=1}^N \frac{u'_{T,i} \Sigma_T u_{T,i}}{\tilde{d}_T(\lambda_{T,i})^2} \xrightarrow{\text{a.s.}} \int \frac{x}{|1 - c - c x \check{s}_{\underline{F}}(x)|^2 \times \tilde{d}(x)^2} dF(x) .$$

Thus,

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

with the function d^* defined by equation (B.8). Putting equations (B.9) and (B.10) together yields

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

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

B.3 Proof of Proposition 4.1

Recall that the function d^* was defined in equation (B.8). Differentiating the right-hand side of equation (3.3) with respect to $\tilde{d}(x)$ for $x \in \text{Supp}(F)$ yields the first-order condition

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

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

B.4 Proof of Proposition 4.2

Theorem 4 of Ledoit and P  ch   (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

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

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

B.5 Proof of Theorem 4.1

Define the shrinkage function

$$\forall x \in (0, \infty) \quad \hat{d}_T(x) := \frac{1}{x |\hat{s}(x)|^2} .$$

Theorem 2.2 of Ledoit and Wolf (2013b) and Proposition 4.3 of Ledoit and Wolf (2012) imply that $\forall x \in \text{Supp}(F) \quad \widehat{d}_T(x) \xrightarrow{\text{a.s.}} d^*(x)$, and that this convergence is uniform over $x \in \text{Supp}(F)$, apart from arbitrarily small boundary regions of the support. 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 . Theorem 4.1 then follows from Proposition 4.1. ■

C Consistent Estimator of Population Eigenvalues

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 here again 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

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

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

where

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

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

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

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

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

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:

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

D The Singular Case $N > T$

Almost all of the analysis developed above goes through in the singular case $N > T$. The only difference is that now $N - T$ sample eigenvalues are equal to zero. In order to handle them, Section 3.2.2 of Ledoit and Wolf (2013b) introduces the quantity defined as the unique solution $s =: \hat{s}_0$ in $(0, \infty)$ to the equation

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

where $\hat{\tau}_T := (\hat{\tau}_{T,1}, \dots, \hat{\tau}_{T,N})'$ is defined as in Proposition 4.3.

The optimal covariance matrix estimator for portfolio selection is built as in equation (4.4), but with the changed definition

$$\hat{d}_T(\lambda_{T,i}) := \begin{cases} \frac{1}{\left(\frac{N}{T} - 1\right) \hat{s}_0}, & \text{if } \lambda_{T,i} = 0 \\ \frac{1}{\lambda_{T,i} |\hat{s}(\lambda_{T,i})|^2}, & \text{if } \lambda_{T,i} > 0 \end{cases} \quad \text{for } i = 1, \dots, N. \quad (\text{D.2})$$

E 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

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

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

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

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

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

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

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

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

We finally compute upper and lower bounds defined by

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

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 median 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.)

References

- Bai, Z. D. and Silverstein, J. W. (2010). *Spectral Analysis of Large-Dimensional Random Matrices*. Springer, New York, second edition.
- 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.
- 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.
- 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.
- Ilmanen, A. (2011). *Expected Returns: An Investor’s Guide to Harvesting Market Rewards*. Wiley Finance, New York.
- 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      , 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      , 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., Aza  s, 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.

- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.
- 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.
- 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.