Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Portfolio Construction by Mitigating Error Amplification: The Bounded-Noise Portfolio

Long Zhao, Deepayan Chakrabarti, Kumar Muthuraman

informs
http://pubsonline.informs.org/journal/opre/

OPERATIONS RESEARCH
Articles in Advance, pp. 1–19
ISSN 0030-364X (print), ISSN 1526-5463 (online)

CONTEXTUAL AREAS

Portfolio Construction by Mitigating Error Amplification: The Bounded-Noise Portfolio

Long Zhao,[a] Deepayan Chakrabarti,[a] Kumar Muthuraman[a]

[a] McCombs School of Business, University of Texas, Austin, Texas 78712

Contact: long.zhao@mccombs.utexas.edu, https://orcid.org/0000-0002-8211-118X (LZ); deepayan.chakrabarti@mccombs.utexas.edu (DC); kumar.muthuraman@mccombs.utexas.edu, https://orcid.org/0000-0002-2872-1980 (KM)

Received: March 14, 2017
Revised: March 18, 2018; September 9, 2018; December 9, 2018
Accepted: December 17, 2018
Published Online in Articles in Advance: June 26, 2019

Keywords: portfolio choice; estimation error

https://doi.org/10.1287/opre.2019.1858

Abstract. We address the problem of poor portfolio performance when a minimum-variance portfolio is constructed using the sample estimates. Estimation errors are mostly blamed for the poor portfolio performance. However, we argue that even small unbiased estimation errors can lead to significantly bad performance because the optimization step amplifies errors, in a nonsymmetric way. Instead of trying to independently improve the estimation step or fix the optimization step for robustness, we disentangle the well-estimated aspects from the poorly estimated aspects of the covariance matrix. By using a single parameter held constant over all data sets and time periods, our method achieves excellent performance both empirically and in the simulation. We also show how to use information from the sample mean to construct mean-variance portfolios that have higher out-of-sample Sharpe ratios.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2019.1858.

1. Introduction

The celebrated mean-variance portfolio-optimization approach proposed by Markowitz (1952) lays out a clear methodology for constructing asset portfolios that minimize risk, for any performance/reward target. His seminal work helped to initiate an era for the mathematical analysis of financial problems. However, the out-of-sample performance of these mean-variance portfolios in the real world is often unacceptable (Jobson and Korkie 1981; Frost and Savarino 1986, 1988; Jorion 1986; Michaud 1989). This poor performance stems largely from an inability to make precise parameter estimates. Even the simpler variance-minimizing portfolio often has a similarly unacceptable performance (Jagannathan and Ma 2003, DeMiguel et al. 2009a).

The most direct way to solve the minimum-variance (Min-Var) portfolio problem involves two steps. First, we find the best estimates of the covariance matrix using historical data. Next, we use these estimates as inputs to the optimization problem and solve it to obtain a portfolio. Both steps are relatively straightforward. However, even the best covariance estimation in the first step, although unbiased, has errors. When we use this estimate in place of the true covariance matrix, the hope is that the resulting portfolio is still close to the optimal portfolio. However, this is not the case. The large number of covariance estimates relative to the limited size of historical data often is blamed for this poor result. But, as we argue in this paper, the initial error stemming from limited data is amplified by the structure of the optimization procedure itself.

A plethora of research papers suggest ways to address this poor out-of-sample performance. Either these papers try to improve the first estimation step to yield better covariance estimates or modify the second optimization step to produce a better out-of-sample performance (see Section 2). However, DeMiguel et al. (2009a) examine 14 popular methods in terms of their Sharpe ratio, certainty-equivalent return, and turnover and find that none of the methods consistently outperforms the naïve equally weighted portfolio. Better results are obtained on many real-world financial data sets by using *norm-constrained* portfolios (Brodie et al. 2009, DeMiguel et al. 2009b, Fan et al. 2012). Instead of just minimizing portfolio variance, the norm-constrained portfolios seek to minimize a weighted sum of the portfolio variance and a norm of the portfolio weights. Covariance estimation errors often manifest as large weights of some assets, and penalizing portfolio weights limits this problem.

However, the norm-constrained approach presents several problems, stemming primarily from the ad hoc nature of merely modifying the objective to keep the portfolio weights low. First, Green and Hollifield (1992) argue that the optimal portfolio can have sizeable asset weights. Hence, although norm constraints might help, they also might be wrong because they exclude the optimal solution, which involves large portfolio weights. Second, the choice of the norm is arbitrary. Third, the performance of the norm-constrained portfolios depends on the selection of a parameter that captures the importance of keeping the portfolio weights low; that is, the coefficient of the norm. The best parameter value depends on the particular financial data set and the amount of training data, and it even changes over the time horizon of a data set. This makes parameter tuning particularly important.

1.1. Our Main Ideas

In this paper, we argue that the compounding effect of the optimization-driven error amplification on the initial estimation errors is the primary cause of the unacceptable performance. Indeed, this amplification is in a sense nonsymmetric; that is, different kinds of errors are amplified in different ways. Hence, the unbiased initial estimation errors do not translate to the unbiased estimates of the optimal portfolio. Instead of trying to independently improve the estimation step or fix the optimization step for robustness, we disentangle the well-estimated aspects of the covariance matrix from the poorly estimated and handle these appropriately when constructing our portfolio.

Our approach has four steps. First, we examine the estimation of the covariance matrix. It turns out that some eigenvectors of the covariance matrix are easier to estimate than others.[1] This suggests the need to split the set of eigenvectors into two groups: the well estimated and the poorly estimated. However, instead of splitting by using an arbitrary threshold on estimation errors, we use the impact of the estimation errors on the portfolio objective to dictate the split. We call the split groups *signal* and *noise*.

Second, we construct a *signal-only* portfolio from the well-estimated signal eigenvectors. This portfolio by itself performs significantly better than the classical Min-Var portfolio obtained by plugging the sample covariance matrix into the Min-Var optimization problem.

Third, realizing that "poorly estimated" does not imply unimportant, we see how we can benefit from the noise eigenvectors. Although each eigenvector in the noise space is poorly estimated, we argue that, when taken together, the space spanned by them is well estimated. This phenomenon is understandable because this space is orthogonal to the space spanned by the signal eigenvectors. The orthogonality implies that a portfolio from the noise space has the potential to improve performance when combined with the signal-only portfolio. We devise an upper bound on the out-of-sample variance of any portfolio constructed from these noisy eigenvectors and use this upper bound to build a *conservative noise-only* portfolio.

Finally, we combine the signal-only portfolio with the conservative noise-only portfolio to generate a single portfolio that we call the bounded-noise portfolio. Using simulated data and 12 standard data sets with different rebalance frequencies and training lengths, we show that the bounded-noise portfolio does well not only in simulation but also on these real-world data sets. Moreover, unlike norm-constrained portfolios, we use the same value of the scalar threshold parameter that defines the signal/noise split for all data sets. In many ways, this property is critical because it ensures that the out-of-sample performance does not rely on one's ability to fine-tune a very sensitive parameter.

In summary, we provide a mechanism to disentangle signal from noise; construct the signal-only portfolio and the conservative noise-only portfolio; and combine the two to show that the resulting portfolio significantly outperforms popular portfolios in the literature. This entire process requires only one easily interpretable parameter, whose value is invariant to the financial data sets and to the length and time of the training data.

We explore the performance of our methodology in multiple ways: testing on both real-world and simulated data and providing mathematical justifications. The performance in the real world is understandably the ultimate prize. However, testing on simulated data allows for exhaustive tests and keeps the focus on estimation errors alone. Understanding mathematically why a method does well is reassuring and will enable us to understand the method's limitations, enabling further improvement. Together, they help make a strong case that we are not at the mercy of few sample paths or being favored by certain data sets only.

1.2. Other Contributions

Mean-variance portfolios, as opposed to Min-Var portfolios, also need to estimate the expected returns to construct portfolios because of the expected return target constraint. They are often considered more challenging to construct, especially because estimating the expected return is harder than estimating the covariance matrix (Merton 1980) and more essential (Black and Litterman 1992, Chopra and Ziemba 1993). Hence, prior literature (Jagannathan and Ma 2003, Brodie et al. 2009, DeMiguel et al. 2009b, Fan et al. 2012) has mostly focused on the Min-Var problem to bypass this issue. However, expected returns are important drivers of the Sharpe ratio. In Sections 5 and 7, we demonstrate how our method can be extended to use information of sample means to construct a mean-variance portfolio with a significantly better out-of-sample Sharpe ratio than the competing methods.

We also provide a detailed discussion on the relation between our method and the norm-constrained methods. Our analysis shows that the best-performing norm-constrained portfolios might correspond to wrong constraints, which could render the optimal portfolio infeasible. However, these wrong constraints indirectly lead to portfolios with similar properties as our bounded-noise portfolio.

1.3. Outline

The rest of the paper is organized as follows. Section 2 discusses related literature. In Section 3, we discuss how certain estimation errors are amplified via the optimization solver, which results in poor portfolio performance. In Section 4, by mitigating the error amplification, we construct the bounded-noise portfolio. In Section 5, we demonstrate how to extend the bounded-noise idea to maximizing the Sharpe ratio. In Section 6, we detail relations to several existing portfolio-optimization methods. In Section 7, we provide exhaustive comparisons of our portfolios with eight other different portfolio-construction methods. Concluding remarks are made in Section 8.

2. Literature Review

The Min-Var portfolio optimization is to find a portfolio $\boldsymbol{w}$ that minimizes variance $\boldsymbol{w}'\Sigma\boldsymbol{w}$ subject to the budget constraint $\boldsymbol{w}'\mathbf{1} = 1$. Solving this optimization problem with the estimated covariance matrix $\hat{\Sigma}$ in place of the unknown true covariance Σ gives us the estimated Min-Var portfolio in place of the true Min-Var portfolio. The poor out-of-sample performance of the estimated Min-Var portfolio is well-known (Jobson and Korkie 1981; Frost and Savarino 1986, 1988; Jorion 1986; Michaud 1989). Michaud (1989) was the first to describe the original portfolio-optimization framework as error maximization. He argues that the solver overweighs those securities that have large estimated returns, negative correlations, and small variance, which are most likely to have estimation errors. Even the naïve equally weighted portfolio that spreads the budget equally among all assets performs better (DeMiguel et al. 2009a). We can group the papers trying to improve performance into three categories. The first category tries to develop methods that provide better covariance estimates than the sample covariance matrix. In the second category, the estimated Min-Var portfolio is combined with the equally weighted portfolio to maximize a utility measure other than variance. The third category includes modification of the optimization problem itself with the hope of improving performance.

Improving Covariance Estimation. A plethora of research exists on the estimation of the covariance matrix in the context of portfolio optimization.[2] One common approach is to shrink the sample covariance. Ledoit and Wolf (2003) shrink the sample covariance matrix toward the single-index covariance matrix. One can also shrink the eigenvalues of the sample covariance matrix linearly (Ledoit and Wolf 2004) or nonlinearly (Ledoit and Wolf 2012, 2017). The former is equivalent to shrinking the sample covariance matrix toward identity matrix. The shrinkage level is chosen such that it is asymptotically optimal under the Frobenius norm. The shrinkage methods have been shown to dominate the multifactor models on the real-world data (Ledoit and Wolf 2003). A second approach is to use robust statistics to counteract sudden movements in the stock price. DeMiguel and Nogales (2009) provide a careful evaluation on both simulated and real-world data sets and show that the robust statistics can indeed improve performance. A third approach is to use the information from the option price documented in DeMiguel et al. (2013b). They indicate that using option-implied volatility can reduce the out-of-sample standard deviation by more than 10% for various modified Min-Var portfolios on two real-world data sets.

Combining with the Equally Weighted Portfolio. The second category is inspired by the good performance of the equally weighted portfolio (Jobson and Korkie 1980, DeMiguel et al. 2009a, Duchin and Levy 2009). With five reasonable assumptions, Frahm and Memmel (2010) prove that the portfolio constructed by carefully combining the estimated Min-Var portfolio with any reference portfolio dominates the former. They use a loss function that is closely related to out-of-sample variance. In the extensive simulation test and a small real-world data set evaluation, they take the equally weighted portfolio as the reference portfolio and demonstrate the benefit of the combination. By minimizing the expected utility loss, Tu and Zhou (2011) estimate the combination level of each of four different portfolios and the equally weighted portfolio. Using an exhaustive assessment of both the simulated and the real-world data sets, they show that the new portfolios perform better than the equally weighted portfolio. DeMiguel et al. (2013a) use different criteria and calibration methods to decide the combination level and show that the combined portfolios can achieve good performance across several real-world data sets.

Modifying the Optimization. In the third category, the portfolio optimization is modified by penalizing portfolios with some predefined characteristics (or, equivalently, by adding extra constraints based on these characteristics). The most common modification is to avoid aggressive short positions. An extreme case is the no-shorting portfolio, which avoids shorting altogether. This approach is analyzed in Jagannathan and Ma (2003), who argue that the "wrong" no-shorting constraint helps because it reduces the effects of the estimation error. They give evidence for better performance using both simulated and real-world data.

A weaker version of the no-shorting constraint involves penalizing a norm of the portfolio weights,

$$\min_{w} w'\Sigma w + \eta \|w\|_p^p \qquad \text{subject to } w'\mathbf{1} = 1. \qquad (1)$$

Two common norms are the $\mathbb{L}_1$ norm (Welsch and Zhou 2007, Brodie et al. 2009, Fan et al. 2012) and the $\mathbb{L}_2$ norm (Lauprête 2001, DeMiguel et al. 2009b). Among these studies, Fan et al. (2012) is the only one that uses both simulated and real-world data to show better performance and that also provides a mathematical justification. Lauprête (2001) takes the view that norm-constrained portfolios are regularizations that counteract the deviations from the normality of the distribution of returns. Empirical evidence is provided via simulations, but only one real-world data set is used. DeMiguel et al. (2009b) provide more comprehensive empirical results. They show that the norm-constrained portfolios dominate the equally weighted portfolio and the estimated Min-Var portfolio, in terms of both the out-of-sample variance and the Sharpe ratio. They also show the relation between norm-constrained portfolios and Bayesian priors on the sample covariance matrix. Gotoh and Takeda (2011) find that the norm constraints are equivalent to the robust constraints associated with the return vector, and Olivares-Nadal and DeMiguel (2018) point out that the norm constraints can be interpreted as the transaction costs. These relations indicate that the same basic idea underpins many seemingly disparate models.

Our approach is complementary to each of the three categories. Estimation errors might be reduced by the first set of methods, but it cannot be eliminated, and we show that this error is amplified by the solver of the portfolio optimization. The second category is based on the good performance of the equally weighted portfolio. We provide theoretical reasons for its good performance and also show its relation with our bounded-noise portfolio. In the third category, the norm of the portfolio weights is penalized, but the penalty factor and the norm must be chosen for each data set. We show how the right penalty can be chosen by using a single constant parameter in our method. This constant parameter value is robust and applies not only to every period of a data set, but also across data sets.

There is another stream of related literature that does not fit into the three categories above. Laloux et al. (1999, 2000) and Plerou et al. (2002) use results from random matrix theory to help estimate better correlation matrices. They also use the signal/noise terminology, but define these differently. They begin by assuming that the correlation matrix is a random matrix generated by independent asset returns. Then, any deviation of the eigenvalue distributions from that dictated by random matrix theory is considered information or a signal. By this definition, for example, the lowest eigenvalues and corresponding eigenvectors are likely to be considered a part of the signal in their case (Plerou et al. 2002), but noise in ours. Laloux et al. (2000) modify their correlation matrix (not covariance) by replacing all the noise eigenvalues with their average and show that it performs better than the sample correlation matrix.

3. Estimation Error and Its Amplification

The basis of our approach stems from the fact that some eigenvalues and corresponding eigenvectors of the true covariance matrix are better estimated than others. In this section, we describe the estimation errors and how they get amplified in the optimization step. While doing this, we also present our signal and noise space and show how portfolios in these spaces can be combined.

3.1. Estimation Error

In Proposition 3.1, we show that the relative errors (percentage deviations from the true values) in estimating the large eigenvalues of the true covariance matrix are small, whereas the relative errors in estimating the small eigenvalues are large. We use Σ to represent the true covariance matrix and $\hat{\Sigma}$ to represent the sample covariance matrix. $\|\cdot\|_{op}$ denotes the operator norm. The sample size is n, and the number of assets is p. Proofs are deferred to the online appendix.

Proposition 3.1 (Eigenvalue Concentration). *Let λ_i and $\hat{\lambda}_i$ represent the i^{th} largest eigenvalues of Σ and $\hat{\Sigma}$, respectively. Then, we have*

$$\frac{|\lambda_i - \hat{\lambda}_i|}{\lambda_i} \le \frac{\|\Sigma - \hat{\Sigma}\|_{op}}{\lambda_i}.$$

Estimation errors for the eigenvectors are a bit more complicated to characterize. Lemma 3.2 shows that the estimation error not only depends on $\|\Sigma - \hat{\Sigma}\|_{op}$, but also how separated the eigenvalues are.

Lemma 3.2 [Concentration of Eigenvectors (Yu et al. 2015)]. *Let $\Sigma, \hat{\Sigma} \in \mathbb{R}^{p\times p}$ be symmetric, with eigenvalues $\lambda_1 \ge \ldots \ge \lambda_p$ and $\hat{\lambda}_1 \ge \ldots \ge \hat{\lambda}_p$, respectively. Fix $1 \le r \le s \le p$, and assume that $\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}) > 0$, where $\lambda_0 = \infty$ and $\lambda_{p+1} = -\infty$. Let $d = s - r + 1$. Let $V = (v_r, v_{r+1}, \ldots, v_s) \in \mathbb{R}^{p\times d}$ and $\hat{V} = (\hat{v}_r, \hat{v}_{r+1}, \ldots, \hat{v}_s) \in \mathbb{R}^{p\times d}$ have orthogonal columns satisfying $\Sigma v_j = \lambda_j v_j$ and $\hat{\Sigma}\hat{v}_j = \hat{\lambda}_j \hat{v}_j$; then there exists an orthogonal matrix $\hat{O} \in \mathbb{R}^{d\times d}$, such that*

$$\|\hat{V}\hat{O} - V\|_F \le \frac{2^{3/2} d^{1/2} \|\hat{\Sigma} - \Sigma\|_{op}}{\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1})}.$$

Vershynin (2011) provides a description of $\|\Sigma - \hat{\Sigma}\|_{op}$ in terms of n and p: under mild conditions, a high-probability upper bound of $\|\Sigma - \hat{\Sigma}\|_{op}$ is roughly of order $(p/n)^{\frac{1}{2}-\frac{2}{q}}$, where the qth moment of the data are bounded. Thus, for a given number of assets p, the difference decays when more observations are available, as expected.

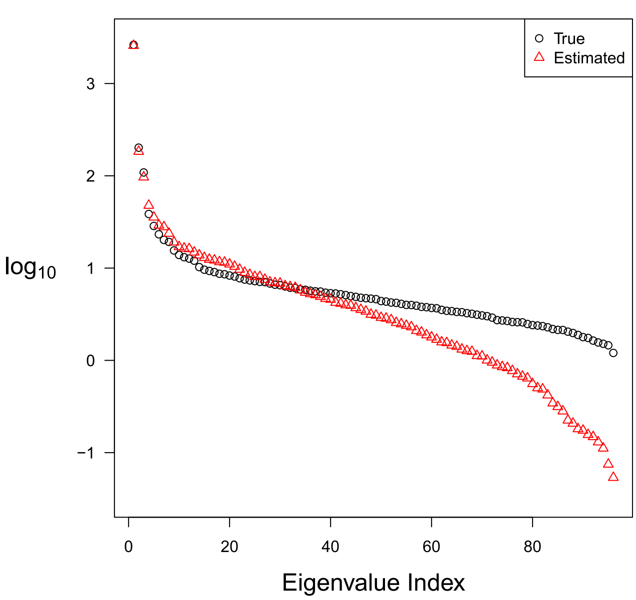
Previous work on financial data sets shows that a few factors can explain a significant portion of the variance of asset returns (Fama and French 2015). This finding suggests that Σ has only a few large eigenvalues (whose corresponding eigenvectors mirror the relevant factors), whereas most of the eigenvalues are small (so their eigenvectors have just a small contribution to the variance of asset returns).

This intuition is supported by our observations from a historical covariance matrix constructed from the monthly returns of the Fama–French value-weighted data set with 96 instruments, aggregated over 625 months. Figure 1 shows the eigenvalues of this "true" covariance matrix, as well as those of a sample covariance matrix simulated from the covariance matrix (both of which are ordered from largest to smallest eigenvalue). Observe that the largest eigenvalues are well separated, but the smallest ones are densely packed (note that we scale the y-axis logarithmically). Note also that the relative difference between the estimated and the true eigenvalues is small for the largest eigenvalues, implying that they are relatively well estimated. In addition to these simulation results and the arguments from the finance literature, we see widespread evidence of similar phenomena in the eigenvalue spectra of many real-world networks (Mihail and Papadimitriou 2002, Chakrabarti and Faloutsos 2006).

3.2. Error Amplification in Portfolio Optimization

The previous discussion shows that the largest eigenvalues and related eigenvectors in the sample covariance

Figure 1. (Color online) Distribution of True and Estimated Eigenvalues



$\hat{\Sigma}$ are relatively good estimates of the corresponding eigenvalues and eigenvectors of the true covariance matrix Σ. The smaller eigenvalues and the corresponding eigenvectors are poor estimates. Hence, we separate the true eigenvectors $(v_1, \ldots, v_p)$ into two sets: from index 1 to k, and from $k+1$ to p. When the split index k is chosen appropriately, we expect the first set to be better estimated than the second set. We will show that the first set of estimated eigenvalues and eigenvectors are also more reliable for portfolio construction, whereas the remaining ones are not. This will guide our algorithm for estimating k from training data, as detailed later in Section 4.1.

For now, given a k, denote the space spanned by $v_1, \ldots, v_k$ as $\mathcal{S}$ and the space spanned by the other eigenvectors as $\mathcal{N}$. To understand how these two parts influence portfolio optimization, we first provide a new characterization of the true Min-Var portfolio.

Lemma 3.3 (Portfolio Decomposition). *For any separation* $(\mathcal{S}, \mathcal{N})$, *the optimal portfolio* $\boldsymbol{w}^*$ *can be expressed as*

$$\boldsymbol{w}^* = \alpha \boldsymbol{w}_S^* + (1-\alpha)\boldsymbol{w}_N^*, \tag{2}$$

$$\alpha = \frac{1/RV(\boldsymbol{w}_S^*)}{1/RV(\boldsymbol{w}_S^*) + 1/RV(\boldsymbol{w}_N^*)}. \tag{3}$$

Here, $\boldsymbol{w}_S^*$ *and* $\boldsymbol{w}_N^*$ *are defined as the solution to the following optimization problems,*

$$\boldsymbol{w}_S^* = \arg\min_{\boldsymbol{w}} \ \boldsymbol{w}'\Sigma\boldsymbol{w}, \quad \text{subject to} \ \boldsymbol{w}'\mathbf{1} = 1, \ \boldsymbol{w} \in \mathcal{S}, \qquad \Bigg| \qquad \boldsymbol{w}_N^* = \arg\min_{\boldsymbol{w}} \ \boldsymbol{w}'\Sigma\boldsymbol{w}, \quad \text{subject to} \ \boldsymbol{w}'\mathbf{1} = 1, \ \boldsymbol{w} \in \mathcal{N}.$$

That is, $\boldsymbol{w}_S^*$ *is the solution to the Min-Var problem given the restriction of being a linear combination of the first* k *eigenvectors (the vectors that span* $\mathcal{S}$*) and* $\boldsymbol{w}_N^*$ *the solution with the restriction of being a linear combination of the other eigenvectors. In the above,* $RV(\boldsymbol{w})$ *is the out-of-sample variance (henceforth, the realized variance)*[3] *of* $\boldsymbol{w}$*, namely,*

$$RV(\boldsymbol{w}) = \boldsymbol{w}'\Sigma\boldsymbol{w}.$$

Thus, the true Min-Var portfolio can be seen as a convex combination of two portfolios: one restricted to space $\mathcal{S}$ and the other confined to space $\mathcal{N}$. The weight of each portfolio is proportional to the inverse of its realized variance.

Now consider the estimated Min-Var portfolio $\hat{\boldsymbol{w}}^*$. It can be expressed in the same form as in Lemma 3.3, but with the true parameters replaced by their estimated counterparts. In particular, the eigenspace $\mathcal{S}$ is replaced by $\hat{\mathcal{S}} = \text{span}(\hat{v}_1, \ldots, \hat{v}_k)$; $\mathcal{N}$ is replaced by $\hat{\mathcal{N}} = \text{span}(\hat{v}_{k+1}, \ldots, \hat{v}_p)$; the portfolios $\boldsymbol{w}_S^*$ and $\boldsymbol{w}_N^*$ are replaced by $\hat{\boldsymbol{w}}_S^*$ and $\hat{\boldsymbol{w}}_N^*$. We use $\hat{\boldsymbol{w}}_S^*$ instead of $\hat{\boldsymbol{w}}_{\hat{S}}^*$ solely to simplify notation. Also, crucially, the realized variance $RV(w) = w'\Sigma w$ is replaced by the *estimated* variance $EV(w) = w'\hat{\Sigma}w$. Thus, the relative weight of $\hat{w}_S^*$ to $\hat{w}_N^*$ in the overall portfolio $\hat{w}^*$ [Equation (3)] is now driven by the estimated variance instead of the realized variance.

To further illustrate the differences between the realized variance and the estimated variance, we perform simulations on the Fama–French value-weighted data set comprising 96 stocks. In the simulation, we assume that the true covariance matrix Σ and the true expected return μ are the sample covariance matrix and the sample mean using all monthly data from July 1963 to July 2015 (625 observations). We also assume that the returns follow a multivariate normal distribution with mean μ and covariance Σ, and we draw 120 observations (10-year monthly data) from this distribution.

We calculate the realized and the estimated variances for various split indices k. We repeat this experiment 100 times and calculate related averages. Figure 2 shows the ratio of realized variance to estimated variance for $\hat{w}_S^*$ and $\hat{w}_N^*$. The realized variance of $\hat{w}_S^*$ is similar to its estimated variance when k is small [Figure 2(a)]. However, for $\hat{w}_N^*$, the realized variance is much larger than its estimated variance [Figure 2(b)]. Indeed, it is at least 20 times larger for any k. This underestimation means that $\hat{w}_N^*$, which uses the poorly estimated parameters, gets overweighted significantly when $\hat{w}_S^*$ and $\hat{w}_N^*$ are combined to construct $\hat{w}^*$.

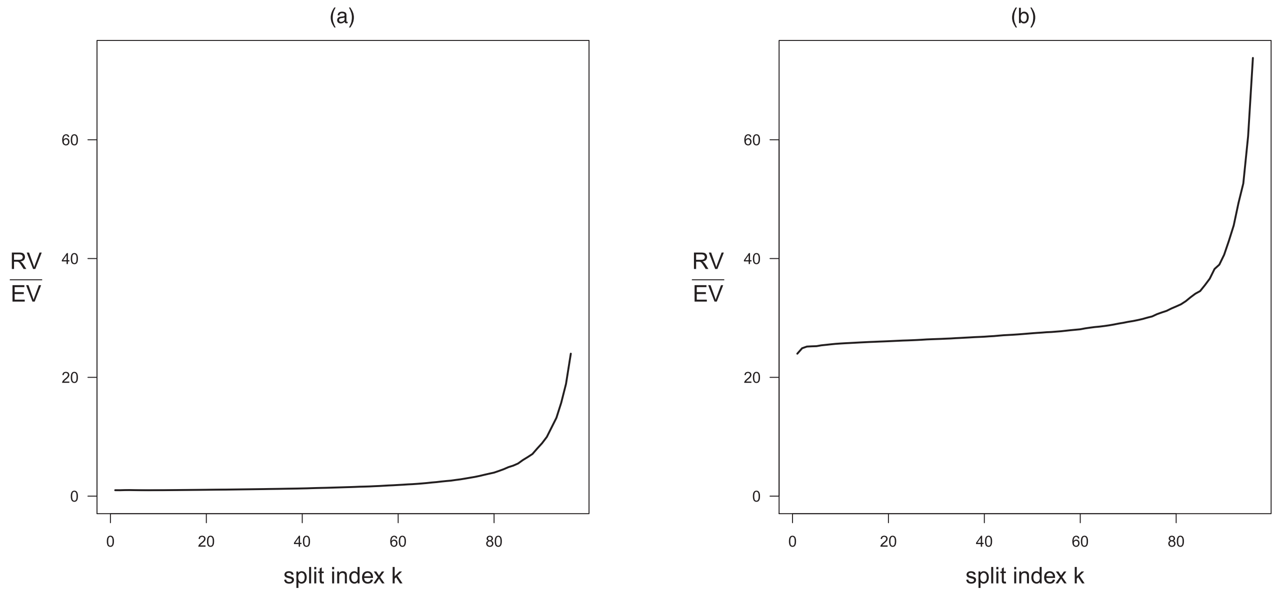
4. The Bounded-Noise Portfolio

The previous discussion shows the utility of separating a "signal" space $\hat{\mathcal{S}}$ (and the *signal-only portfolio* $\hat{w}_S^*$) from a "noise" space $\hat{\mathcal{N}}$ (and the *aggressive noise-only portfolio* $\hat{w}_N^*$) using a signal/noise split index k on the eigenvectors of the covariance matrix $\hat{\Sigma}$. In this section, we begin by formally defining the signal/noise split. Rather than splitting by the estimation errors, we show in Section 4.1 how we can use the effect of the estimation errors on the optimization objective to dictate the split. Given this split, we then construct the signal-only portfolio by minimizing its estimated variance in Section 4.2. To take advantage of the information contained in the noise space, in Section 4.3, we use the idea of minimizing the upper bound of the realized variance to construct the conservative noise-only portfolio. This bound also provides a way to combine the conservative noise-only portfolio cautiously with the signal-only portfolio. We describe the combination procedure in Section 4.4. We call the combined portfolio the bounded-noise portfolio (the BN portfolio). We discuss the entire algorithm including the procedure to estimate the parameters needed in the algorithm, in Section 4.5.

4.1. Splitting into Signal and Noise

Our intuition for a signal is that $(\hat{\lambda}_i, \hat{\mathbf{v}}_i) \approx (\lambda_i, \mathbf{v}_i)$. However, this intuition can be refined based on the specifics of the portfolio-optimization problem. The simulation in Section 3.2 shows the underestimation of the realized variance by the estimated variance.

Figure 2. The Ratio between RV and EV



Notes. Panel (a): The top-k eigenvector portfolio $\hat{w}_S^*$. Panel (b): The bottom $(p-k)$-eigenvector portfolio $\hat{w}_N^*$.

This underestimation leads to the aggressive noise-only portfolio being overweighted in the estimated Min-Var portfolio. Hence, we characterize the signal space as all eigenpairs $(\hat{\lambda}_i, \hat{\mathbf{v}}_i)$ that are such that

$$\text{amplification ratio } \phi_i \triangleq \frac{RV(\hat{\mathbf{v}}_i)}{EV(\hat{\mathbf{v}}_i)} = \frac{RV(\hat{\mathbf{v}}_i)}{\hat{\lambda}_i} \leq 1 + \gamma, \tag{4}$$

where the parameter $\gamma > 0$ allows for some flexibility. We set $\gamma = 0.25$ for all experiments and provide sensitivity analysis on γ in Section 7.5. Because the realized variance is unknown, ϕ_i needs to be estimated, and we provide the procedure in Section 4.5. The first advantage of Equation (4) is that accurate estimation of eigenvalues and eigenvectors (i.e., $(\hat{\lambda}_i, \hat{\mathbf{v}}_i) \approx (\lambda_i, \mathbf{v}_i)$) is sufficient to ensure $\phi_i \leq 1 + \gamma$, but is not necessary.[4] Another advantage is that it does not impose separate conditions on eigenvalues and eigenvectors; instead, it captures, via a single formula, the way in which these quantities affect portfolio optimization.

Definition 1 (Signal and Noise). Let the eigenvalues of the estimated covariance matrix $\hat{\Sigma}$ be set in decreasing order: $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$. Let the corresponding eigenvectors be denoted by $\hat{\mathbf{v}}_i$. Let the eigenvalues λ_i and the corresponding eigenvectors $\mathbf{v}_i$ of the true covariance matrix Σ also be ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$. For a given γ, the signal/noise split point k^* is defined as follows:

$$k^* = \max\{k | \phi_i \leq 1 + \gamma, \ \forall i \leq k\}.$$

The space spanned by $\{\hat{\mathbf{v}}_i \mid i \leq k^*\}$ is defined as the **signal space**, whereas the space spanned by $\{\hat{\mathbf{v}}_i \mid i > k^*\}$ is defined as the **noise space**. It is possible for one of these spaces to be empty. The corresponding sets of eigenvalue and eigenvector pairs, namely, $\{(\hat{\lambda}_i, \hat{\mathbf{v}}_i) \mid i \leq k^*\}$ *and* $\{(\hat{\lambda}_i, \hat{\mathbf{v}}_i) \mid i > k^*\}$, are referred to as **signal** and **noise**, respectively.

In Definition 1, we assume the true covariance matrix is strictly positive definite. With such an assumption, given p, as $n \to \infty$, all $(\hat{\lambda}_i, \hat{\mathbf{v}}_i)$ pairs are considered to be signal:

$$\begin{aligned} |\phi_i - 1| &= \left| \frac{\hat{\mathbf{v}}_i' \Sigma \hat{\mathbf{v}}_i}{\hat{\lambda}_i} - 1 \right| = \frac{|\hat{\mathbf{v}}_i'(\Sigma - \hat{\Sigma})\hat{\mathbf{v}}_i|}{\hat{\lambda}_i} \\ &\leq \frac{\|\Sigma - \hat{\Sigma}\|_{op}}{\max(0, \lambda_p - \|\Sigma - \hat{\Sigma}\|_{op})} \to 0, \end{aligned}$$

where the last inequality follows from Proposition 3.1 and the definition of the operator norm.

4.2. The Signal-Only Portfolio

By construction, the signal space consists of sample eigenvectors whose estimated variance is a reliable indicator of their realized variance. Thus, the signal-only portfolio, $\hat{w}_S^*$, constructed from these sample eigenvectors, should also be reliable. Mathematically speaking, this portfolio is equivalent to a principal component analysis-based portfolio that ignores a certain number of the low eigenvalues of $\hat{\Sigma}$ and corresponding eigenvectors.

4.3. The Conservative Noise-Only Portfolio

Eigenvectors in the noise space are poorly estimated. Hence, although the estimated variance of a portfolio from the noise space might be low, its realized variance might be much higher. Our idea is simple: Because estimates of variance are too unreliable in the noise space, we instead develop an upper bound for the realized variance of any portfolio in the noise space. Then, we choose the portfolio that minimizes this upper bound.

Proposition 4.1 (Bounding Realized Variance of any Portfolio from the Noise Space). *Let the noise space eigenvectors of* $\hat{\Sigma}$ *be* $\hat{v}_{k+1}, \ldots, \hat{v}_p$, *the space spanned by them be* $\hat{\mathcal{N}}$, *and the matrix whose columns are* $(\hat{v}_{k+1}, \ldots, \hat{v}_p)$ *be* $\hat{\mathrm{N}}$. *For any* $w \in \hat{\mathcal{N}}$,

$$RV(w) \leq EV(w) + m\|w\|_2^2, \tag{5}$$

where m *is the largest eigenvalue of the matrix* $\hat{\mathrm{N}}'(\Sigma - \hat{\Sigma})\hat{\mathrm{N}}$.

We call m the *noise bound*. Because the true covariance matrix, Σ, is unknown, we need to estimate the noise bound, and the procedure is provided in Section 4.5. The realized variance of any portfolio w from the noise space can be upper-bounded by a function $BRV(w)$, which is defined as $BRV(w) = EV(w) + m\|w\|_2^2$. Here, BRV stands for the bounded realized variance. It is natural now to choose a portfolio from the noise space that minimizes this upper bound:

$$\begin{aligned} \min_{w} \quad & BRV(w), \\ \text{subject to} \quad & w'\mathbf{1} = 1, \\ & w \in \hat{\mathcal{N}}. \end{aligned} \tag{6}$$

This portfolio is not necessarily close to the optimal noise portfolio w_N^*. However, it is conservative because it is the bound that is minimized. Thus, we call this portfolio the *conservative noise-only portfolio* and denote it as $\hat{w}_N^{BN}$.

4.4. Combining the Two Portfolios

Finally, we must combine the signal-only portfolio, $\hat{w}_S^*$, with the conservative noise-only portfolio, $\hat{w}_N^{BN}$, into a single portfolio. Equation (3) shows that the combination weights each portfolio by the inverse of its realized variance. For the signal-only portfolio, the estimated variance is a good proxy for the realized variance. However, the same is not true for the conservative noise-only portfolio. Hence, instead of using its erroneous estimated variance, we use the upper bound.[5] Thus, the BN portfolio is given by

$$\begin{aligned}\hat{\boldsymbol{w}}^{BN} &= \alpha^{BN}\hat{\boldsymbol{w}}_S^* + (1-\alpha^{BN})\hat{\boldsymbol{w}}_N^{BN},\\ \alpha^{BN} &= \frac{1/EV(\hat{\boldsymbol{w}}_S^*)}{1/EV(\hat{\boldsymbol{w}}_S^*) + 1/BRV(\hat{\boldsymbol{w}}_N^{BN})}.\end{aligned} \tag{7}$$

Given the split, k^*, and the noise bound, m, we can obtain the analytical form of both the signal and the conservative noise portfolio, which can be plugged into Equation (7) to express $\hat{\boldsymbol{w}}^{BN}$ as

$$\hat{\boldsymbol{w}}^{BN} = \frac{\sum_{i=1}^{k^*}\frac{\hat{\boldsymbol{v}}_i'\mathbf{1}}{\hat{\lambda}_i}\hat{\boldsymbol{v}}_i + \sum_{i=k^*+1}^{p}\frac{\hat{\boldsymbol{v}}_i'\mathbf{1}}{\hat{\lambda}_i+m}\hat{\boldsymbol{v}}_i}{\sum_{i=1}^{k^*}\frac{(\hat{\boldsymbol{v}}_i'\mathbf{1})^2}{\hat{\lambda}_i} + \sum_{i=k^*+1}^{p}\frac{(\hat{\boldsymbol{v}}_i'\mathbf{1})^2}{\hat{\lambda}_i+m}}. \tag{8}$$

In other words, the BN portfolio adds the noise bound, m, to the eigenvalues whose corresponding eigenvectors belong to the noise space while adding 0 to the other eigenvalues. Thus, Equation (8) is equivalent to saying that the BN portfolio, $\hat{\boldsymbol{w}}^{BN}$, is the solution to the following optimization problem:

$$\begin{aligned}\min_{\boldsymbol{w}} \quad & \boldsymbol{w}'(\hat{\Sigma} + M)\boldsymbol{w},\\ \text{subject to} \quad & \boldsymbol{w}'\mathbf{1} = 1,\\ \text{where} \quad & M = m\hat{N}\hat{N}'.\end{aligned} \tag{9}$$

4.5. Estimating k^* and m

Both k^* and m are functions of Σ and $\hat{\Sigma}$. Because Σ is unknown, they need to be estimated. Instead of assuming a particular distribution of returns (say, Gaussian), we estimate them using the bootstrap method. In particular, we draw bootstrap samples from the observed returns and construct the bootstrap covariance matrix $\hat{\Sigma}_B$. Then, we estimate k^* and m by using $(\hat{\Sigma}, \hat{\Sigma}_B)$ in place of $(\Sigma, \hat{\Sigma})$ in Definition 1 and Proposition 4.1. Plugging in these estimates in Equation (8) gives us the BN portfolio weights. The algorithm is summarized below.

1. Estimation of the split, k^*, and the noise bound, m.

(a) Draw $L = 1{,}000$ bootstrap samples from the observed sample returns. Construct the corresponding bootstrap covariance matrices $\hat{\Sigma}_{Bj}, j = 1, 2, \ldots, L$.

(b) Calculate the bootstrap analogs $\tilde{\phi}_{ji}$ of ϕ_i for each $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, L$:

$$\tilde{\phi}_{ji} = \frac{\tilde{\mathbf{v}}_{ji}'\hat{\Sigma}\tilde{\mathbf{v}}_{ji}}{\tilde{\lambda}_{ji}},$$

where $\tilde{\lambda}_{ji}$ and $\tilde{\mathbf{v}}_{ji}$ are the i^{th} eigenvalue and eigenvector of $\hat{\Sigma}_{Bj}$, respectively.

(c) Estimate the split, k^*, as follows:

$$\begin{aligned}\hat{k} = \max\Big\{k \mid \text{median}\{\tilde{\phi}_{ji} \mid j \in [1, L]\}) \\ \le 1 + \gamma = 1 + 0.25 = 1.25, \forall i \le k\Big\}.\end{aligned}$$

(d) Estimate the noise bound, m, using the following estimator

$$\hat{m} = \text{median}\Big\{\lambda_{max}\Big(\hat{N}_{Bj}'(\hat{\Sigma} - \hat{\Sigma}_{Bj})\hat{N}_{Bj}\Big)\Big| j = 1, 2, \ldots, L\Big\},$$

where λ_{max} denotes the largest eigenvalue of a matrix, and $\hat{N}_{Bj}$ is the matrix of eigenvectors of $\hat{\Sigma}_{Bj}$ in the noise space: $\hat{N}_{Bj} = (\tilde{\mathbf{v}}_{j\hat{k}+1}, \ldots, \tilde{\mathbf{v}}_{jp})$.

2. Replace the split, k^*, and the noise bound, m, with their estimation $\hat{k}$ and $\hat{m}$ in Equation (8) to get the BN portfolio.

Note that the median is used instead of the mean in steps (c) and (d) to ensure robustness of the estimates. Figure 3 contrasts the classical approach with the bounded-noise procedure.

5. Bounded-Noise Portfolios for Mean-Variance Optimization

Our discussions up to this point and in much of the related literature (Jagannathan and Ma 2003, Brodie et al. 2009, DeMiguel et al. 2009b, Fan et al. 2012) focus on the Min-Var portfolio to avoid the problem of estimating the expected returns. However, this focus restricts our ability to optimize for other measures such as the Sharpe ratio. In this section, we show how to adapt the bounded-noise idea to the problem of maximizing the Sharpe ratio.

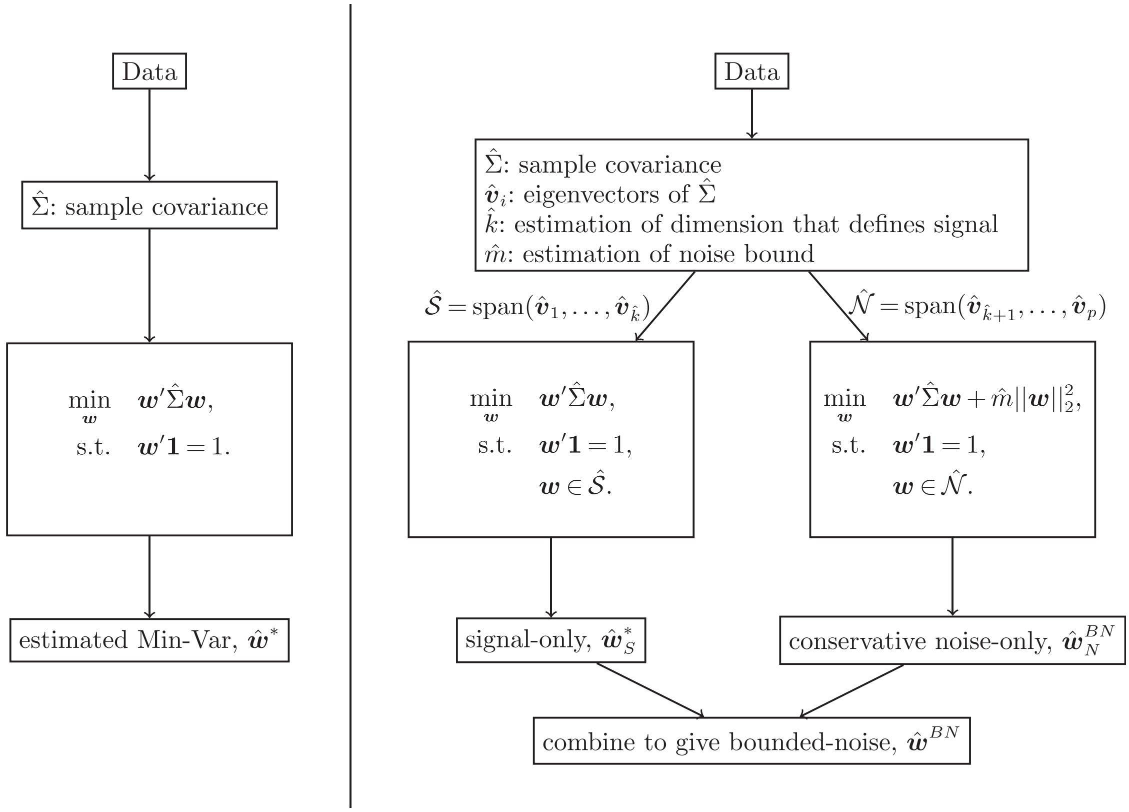
One difficulty in achieving a high Sharpe ratio is that stretching for higher estimated expected returns often requires aggressive positions, which can cause unexpected increases in the realized variance if the errors in the covariance estimation are not adequately accounted for. That is to say, any gains in expected returns can be swamped by the increases in the realized variance, leading to a Sharpe ratio even lower than that of the estimated Min-Var portfolio. However, our upper bound on the realized variance allows us to overcome this issue.

We propose the following formulation for the mean-variance portfolio problem:

$$\begin{aligned}\max_{\boldsymbol{w}} \quad & \hat{\boldsymbol{\mu}}'\boldsymbol{w}\\ \text{subject to} \quad & \boldsymbol{w}'(\hat{\Sigma} + \hat{M})\boldsymbol{w} \le c\sigma_{\min}^2\\ & \boldsymbol{w}'\mathbf{1} = 1,\\ \text{where} \quad & \hat{M} = \hat{m}\hat{N}\hat{N}'.\end{aligned} \tag{10}$$

Here, $\hat{\boldsymbol{\mu}}$ is the vector of mean return estimated from the samples, $c \ge 1$ is a constant, and $\sigma_{\min}^2$ is the optimal objective value of the BN optimization problem [Equation (9)] with $\hat{M}$ replacing M. If $c = 1$, we recover

Figure 3. Diagram of the Estimated Min-Var Portfolio Compared with the Bounded-Noise Portfolio



the BN portfolio. When $c > 1$, Equation (10) maximizes the estimated expected return by allowing a higher realized variance than the BN portfolio. Crucially, the inequality in Equation (10) is based not on the estimated variance but on our upper bound for the realized variance. Because the Sharpe ratio has the realized standard deviation $RSD(w) = \sqrt{RV(w)}$ as its denominator, this inequality ensures that the denominator cannot become too large and overshadow the gains in the expected returns.

We choose the value of c via cross-validation over all previous periods. In the experiments, we set $c = 1$ for the first two years (i.e., we use the BN portfolio). Then, each time we generate a new portfolio, we choose $c \in [1, 1.5]$ such that the previous overall out-of-sample Sharpe ratio is maximized.[6] For example, at the end of the 10th year, we calculate the Sharpe ratio of the previous 10 years' monthly returns for various $c \in [1, 1.5]$. Then, we use the c that gives the highest Sharpe ratio to construct the portfolio for the next period. We call this the BNvar portfolio.

6. Connections to Existing Portfolio-Optimization Methods

Empirical studies have shown that the norm-constrained portfolios work very well in practice (DeMiguel et al. 2009b). The preferred reasoning for its good performance is that the norm penalties on portfolio weights prevent large weights, which are often the result of estimation errors. Apart from the obvious issue of disallowing large portfolio weights even when the optimal portfolio might have them (Green and Hollifield 1992), this approach only fixes one particular effect of estimation error.

We first show that the norm-constrained portfolios might impose the "wrong" constraints. Coupled with the idea of signal/noise split, we show why the norm-constrained portfolios work. We also discuss the relation between the BN portfolio and the equally weighted portfolio. Finally, we provide an interpretation of the BN portfolio as a new way to combine the estimated Min-Var portfolio and the equally weighted portfolio.

6.1. Imposing the Wrong Constraints to Combat Estimation Error

A penalty on the p-norm of portfolio weights, $\|w\|_p$, is equivalent to a constraint of the form $\|w\|_p \le \delta$ for some $\delta > 0$. Such a constraint can be justified if it renders infeasible a large set of poorly performing portfolios that might otherwise be selected because of estimation errors. However, the constraint must not be so restrictive that even the optimal portfolio w^* becomes infeasible.

Figure 4 shows how the realized standard deviation (RSD) varies with different constraint levels, δ, for the $\mathbb{L}_1$ and $\mathbb{L}_2$ norm-constrained portfolios under the simulations using the Fama–French value-weighted data set with 96 assets. In both cases, as expected, the RSD is too high at the extremes, because the constraints become either too strict or too weak. However, the optimum RSD is achieved for a constraint level at which the optimal is infeasible; indeed, the optimum δ is about half of the norm of the optimal portfolio $\|\boldsymbol{w}^*\|_p$. This agrees with Green and Hollifield (1992), who show that the optimal portfolio could have large weights. Thus, the norm-constrained methods can achieve a low RSD only by imposing the wrong constraints, and they cannot be justified simply as a means of capping the estimation error effects.

6.2. Relations to the Norm-Constrained Portfolios

It is easy to see that the $\mathbb{L}_2$ norm-constrained portfolio is a special case of the bounded-noise portfolio. Recall that we can interpret the BN portfolio as the solution to the Min-Var problem using a modified covariance matrix $\hat{\Sigma} + \hat{M}$, where $\hat{M} = \hat{m}\hat{N}\hat{N}'$ [see Equation (9)]. If the noise space contains all eigenvectors, we have $\hat{M} = \hat{m}I$. This yields the norm-constrained portfolio with the $\mathbb{L}_2$-norm penalty and regularization parameter $\hat{m}$.

To get further insight, we must understand how norm-constrained portfolios interact with the signal and noise spaces. The next lemma shows how, for a given k, any portfolio can be split into two unique "projection" portfolios on the signal and noise spaces, and a specific mixing proportion.

Lemma 6.1 (Projection Portfolios). *Denote the eigenvectors of $\hat{\Sigma}$ by $\hat{v}_1, \ldots, \hat{v}_p$. For any integer k between 1 and p, let $\hat{\mathcal{S}} = span(\hat{v}_1, \ldots, \hat{v}_k)$ and $\hat{\mathcal{N}} = span(\hat{v}_{k+1}, \ldots, \hat{v}_p)$. Also introduce matrix $\hat{S} = (\hat{v}_1, \ldots, \hat{v}_k)$, and matrix $\hat{N} = (\hat{v}_{k+1}, \ldots, \hat{v}_p)$. For any weight $\boldsymbol{w}$ that satisfies $\boldsymbol{w}'\mathbf{1} = 1$, there is a unique decomposition,*

$$\boldsymbol{w} = \theta \boldsymbol{w}_S + (1-\theta)\boldsymbol{w}_N, \tag{11}$$

such that $\boldsymbol{w}_S \in \hat{\mathcal{S}}$, $\boldsymbol{w}_S'\mathbf{1} = 1$, and $\boldsymbol{w}_N \in \hat{\mathcal{N}}$, $\boldsymbol{w}_N'\mathbf{1} = 1$. These "projection portfolios" $\boldsymbol{w}_S$ and $\boldsymbol{w}_N$, and the inferred mixing proportion θ, are given by

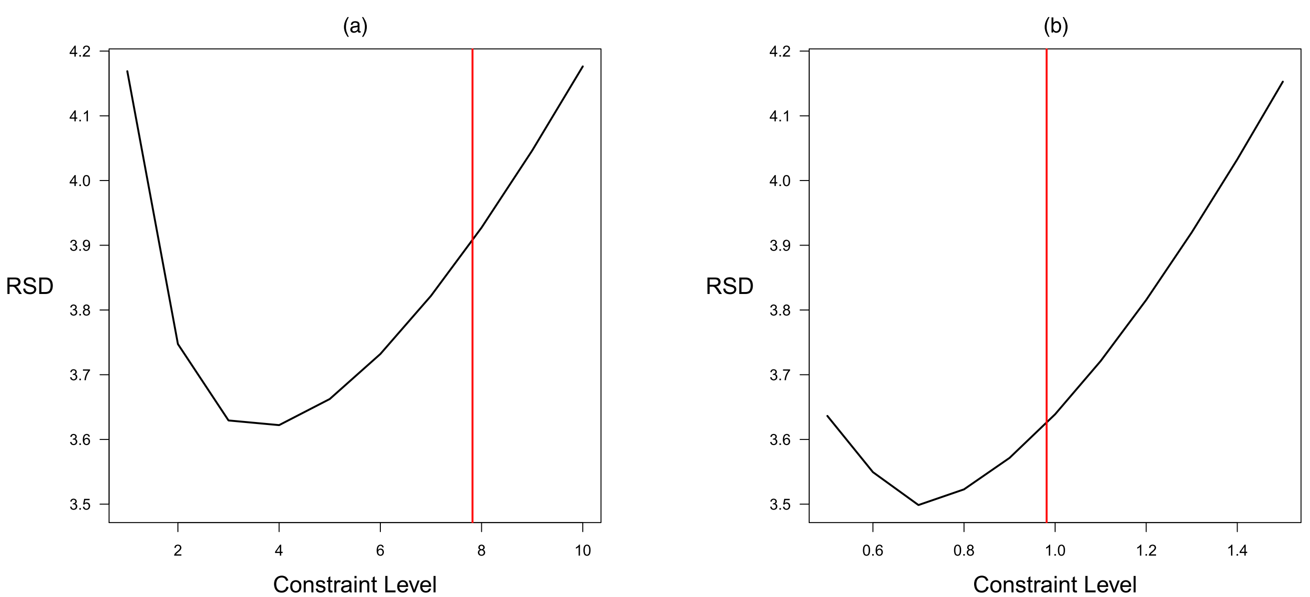
$$\theta = \boldsymbol{w}'\hat{S}\hat{S}'\mathbf{1}, \quad \boldsymbol{w}_S = \frac{\hat{S}\hat{S}'\boldsymbol{w}}{\boldsymbol{w}'\hat{S}\hat{S}'\mathbf{1}}, \quad \boldsymbol{w}_N = \frac{\hat{N}\hat{N}'\boldsymbol{w}}{\boldsymbol{w}'\hat{N}\hat{N}'\mathbf{1}}. \tag{12}$$

Also, as discussed in Section 3.2, the mixing proportion for the estimated Min-Var portfolio is

$$\frac{1/EV(\hat{\boldsymbol{w}}_S^*)}{1/EV(\hat{\boldsymbol{w}}_S^*) + 1/EV(\hat{\boldsymbol{w}}_N^*)}. \tag{13}$$

The strong performance of norm-constrained portfolios could be because they have better projection portfolios than the estimated Min-Var portfolio or because they use a better mixing proportion than relying on the estimated variance [Equation (13)]. We explore this by simulating sample returns from a multivariate normal distribution (with μ and Σ from the Fama–French value-weighted data set) and constructing portfolios from these samples. We then calculate the RSD of the corresponding projection portfolios. All results are averaged over 100 iterations. For brevity, we will call the $\mathbb{L}_1$-norm constrained portfolio the $\mathbb{L}_1$ portfolio with weight vector $\hat{w}^{\mathbb{L}_1}$; the $\mathbb{L}_2$ portfolio with weight

Figure 4. (Color online) Realized Standard Deviation (RSD) with Respect to Different Norm-Constraint Levels



Notes. Panel (a): L_1-norm, the vertical line: $\delta = ||w^*||_1$. Panel (b): L_2-norm, the vertical line: $\delta = ||w^*||_2$.

Table 1. RSD of Projection Portfolios

Portfolio weights	$\hat{w}_S^*$	$\hat{w}_S^{BN}$	$\hat{w}_S^{\mathbb{L}_1}$	$\hat{w}_S^{\mathbb{L}_2}$	$\hat{w}_N^*$	$\hat{w}_N^{BN}$	$\hat{w}_N^{\mathbb{L}_1}$	$\hat{w}_N^{\mathbb{L}_2}$
Mean monthly RSD (%)	3.696	3.696	3.753	3.719	7.687	4.917	6.008	4.928

vector $\hat{w}^{\mathbb{L}}$ is defined accordingly.[7] All projection portfolios are built by using the estimated signal/noise split $\hat{k}$.

Table 1 compares the RSD of the projection portfolios for the $\mathbb{L}_1$ and $\mathbb{L}_2$ portfolios, as well as the BN portfolio. Observe that the signal-space projections of all portfolios have similar RSD (indeed, $\hat{w}_S^* = \hat{w}_S^{BN}$ by construction). Thus, even though the $\mathbb{L}_1$ and $\mathbb{L}_2$ portfolios do not explicitly construct a signal/noise split, they indirectly use the signals just as effectively.

The noise-space projections of the $\mathbb{L}_1$ and $\mathbb{L}_2$ portfolios achieve a much lower RSD than the aggressive noise-only portfolio $\hat{w}_N^*$. Thus, norm-based penalties indirectly lead to improved noise-space portfolios.

To investigate the effect of the mixing proportion, we create new portfolios $\tilde{\mathbb{L}}_1$ and $\tilde{\mathbb{L}}_2$ that have the same projection portfolios as the $\mathbb{L}_1$ and $\mathbb{L}_2$ portfolios, respectively, but where the mixing proportion is calculated by using estimated variances [Equation (13)]. Table 2 shows that the $\tilde{\mathbb{L}}_1$ and $\tilde{\mathbb{L}}_2$ portfolios are much worse than the $\mathbb{L}_1$ and $\mathbb{L}_2$ portfolios, respectively. In fact, they are even worse than the signal-only portfolio. This indicates that, even with improved noise-space projection portfolios, finding the right mixing proportion is important. The inferred mixing proportion θ (from Lemma 6.1) for the $\mathbb{L}_1$ portfolio is, on average, 1.65 times as large as it for the $\tilde{\mathbb{L}}_1$ portfolio. The corresponding ratio is 2.09 for the $\mathbb{L}_2$ portfolio versus the $\tilde{\mathbb{L}}_2$ portfolio. This shows that norm-constrained portfolios avoid overweighting the noise-space projection portfolios, and hence escape the error amplification trap.

6.3. Relation to the Equally Weighted Portfolio

The estimated Min-Var portfolio mainly fails because the eigenvectors and eigenvalues from the noise space are taken at face value. In the BN portfolio, this problem is rectified because we pick the conservative noise-only portfolio, $\hat{w}_N^{BN}$, that minimized the upper bound of the realized variance (Proposition 4.1). An alternative approach to robustness would be to pick a portfolio from the noise space that has the best "worst-case" realized variance (i.e., the portfolio that is robust against all possible configurations of eigenvectors that span the noise space $\hat{\mathcal{N}}$ and is also robust against their eigenvalues). This solution is completely independent of $\hat{\Sigma}$, apart from the estimated signal/noise split, $\hat{k}$. We could achieve this solution by solving the following optimization problem:

$$\begin{aligned} \min_{w}\max_{\Psi\in\mathcal{U}} \quad & w'\Psi w, \\ \text{subject to} \quad & w'\mathbf{1} = 1 \\ & w \in \hat{\mathcal{N}}, \end{aligned} \tag{14}$$

where $\mathcal{U}$ is the uncertainty set of all possible covariance matrices Ψ that have the same signal eigenvectors and eigenvalues as $\hat{\Sigma}$. Because Equation (14) considers only $w \in \hat{\mathcal{N}}$, we can use the following uncertainty set:

$$\mathcal{U} = \{\Psi | \hat{N}'\Psi\hat{N} \preceq bI_{n-\hat{k}+1}\}, \tag{15}$$

where b is a constant and $I_{n-\hat{k}+1}$ is a $(n-\hat{k}+1)\times(n-\hat{k}+1)$ identity matrix.

The idea of a robust portfolio has been expressed previously in the literature in the form of the equally weighted portfolio. This strategy is the right one in the extreme case where no historical data are available. Otherwise, applying this idea just to the noise space is reasonable. Indeed, the projection of the equally weighted portfolio on the noise space yields precisely the portfolio of Equation (14), as shown in Lemma 6.2.

Lemma 6.2 (The Solution to the Robust Optimization). *The solution to the robust optimization problem* Equation (14) *with the uncertainty set defined in* Equation (15) *is the projection portfolio of the equally weighted portfolio on* $\hat{\mathcal{N}}$.

This shows that the projection w_N^{EW} of the equally weighted portfolio w^{EW} is a good candidate for a conservative noise-space portfolio, just like $\hat{w}_N^{BN}$. Indeed, if $\hat{m} \gg \hat{\lambda}_{\hat{k}+1}$, then $\hat{w}_N^{BN} \approx w_N^{EW}$, because

$$\hat{w}_N^{BN} = \frac{\sum_{i=\hat{k}+1}^{p} \frac{\hat{v}_i'\mathbf{1}}{\hat{\lambda}_i+\hat{m}}\hat{v}_i}{\sum_{i=\hat{k}+1}^{p} \frac{(\hat{v}_i'\mathbf{1})^2}{\hat{\lambda}_i+\hat{m}}} \approx \frac{\sum_{i=\hat{k}+1}^{p}(\hat{v}_i'\mathbf{1})\hat{v}_i}{\sum_{i=\hat{k}+1}^{p}(\hat{v}_i'\mathbf{1})^2} = w_N^{EW}. \tag{16}$$

In simulation, the average inner product between $\hat{w}_N^{BN}$ and w_N^{EW} is 0.991, and their RSDs are similar as well (4.917 versus 4.948). Because the BN portfolio combines $\hat{w}_S^{BN}$ $(=\hat{w}_S^*)$ and $\hat{w}_N^{BN}$ $(\approx w_N^{EW})$, it can be interpreted as a principled method to combine the estimated Min-Var portfolio and the equally weighted portfolio.

Table 2. Effects of Mixing Proportion on RSD

Portfolio weights	$\hat{w}^{BN}$	$\hat{w}^{\mathbb{L}_1}$	$\hat{w}^{\tilde{\mathbb{L}}_1}$	$\hat{w}^{\mathbb{L}_2}$	$\hat{w}^{\tilde{\mathbb{L}}_2}$	$\hat{w}_S^*$
Mean monthly RSD (%)	3.488	3.700	4.215	3.531	3.979	3.696

Table 3. List of Portfolios Considered in Empirical Experiments

Model	Abbreviation
Bounded-noise portfolios	
Min-Var portfolio	BN
Mean-variance portfolio	BNVAR
Equally-weighted portfolio	EW
Value-weighted portfolio	VW
Min-Var portfolio with sample covariance	EstMinVar
Min-Var portfolio with sample covariance and shortsale constrained	NoShorting
L_1-norm-constrained Min-Var portfolio	$\mathbb{L}_1$
L_2-norm-constrained Min-Var portfolio	$\mathbb{L}_2$
Partial Min-Var portfolio with parameter calibrated by maximizing portfolio return in previous period	PARR
Min-Var portfolio with nonlinear shrunk covariance	NonLin

Note. The penalty parameter for the norm-constrained portfolios is chosen by cross-validation over standard deviation.

7. Empirical Results

In this section, we compare the out-of-sample performance of the BN portfolio and the BNVAR portfolio to eight other portfolios from the literature (Table 3) across 12 different data sets (Table 4). The time period for all data sets is July 1963 to July 2015, which shares the same starting point as DeMiguel et al. (2009b). All data sets, except the ones for individual stocks, come from Kenneth French's website.[8] For the 100 Fama and French (1992) data set, because there are missing values for four risky assets for an extended period, we deleted them, leaving 96 of the original 100 portfolios. The individual stocks data sets come from the Center for Research in Security Prices. There is a challenge in creating the stocks data sets due to market issues like mergers, acquisitions, delistings, initial public offerings, etc. Ledoit and Wolf (2017) use a procedure that provides a more stable collection of stocks than random selections (Jagannathan and Ma 2003, DeMiguel et al. 2009b). We use this procedure annually and update our list by choosing the largest 100 or 500 stocks,[9] as measured by their market value.[10] Updating the stock list selection annually facilitates our turnover investigations as well (Section 7.3). The only parameter γ, which is used in the BN portfolio, is set to 0.25 for all data sets. Its sensitivity analysis is in Section 7.5. We use $L = 1,000$ bootstrap samples in the estimation procedure.

Competing Methods. We consider two naïve portfolios, the equally weighted (EW) and the value-weighted (VW) portfolios, as our benchmarks. Every asset in the EW portfolio is given equal weight when it is rebalanced. For the VW portfolio, the fraction of the market capitalization is assigned to each asset as its portfolio weight. DeMiguel et al. (2009a) provide a thorough analysis for both portfolios. The EstMinVar portfolio, which is defined at the beginning of Section 2, is the estimated Min-Var portfolio formulated in Markowitz (1952).

In addition to these standard benchmarks, we consider three others that add additional constraints or penalties to the Min-Var portfolio-optimization problem. The first one is the short-sale-constrained portfolio (Jagannathan and Ma 2003, section 1), which has a nonnegativity constraint on the portfolio weights. We call it the NoShorting portfolio. The remaining two are norm-constrained portfolios, with parameters set

Table 4. List of Data Sets Considered

Data set	Abbreviation	p
Six Fama and French (1992) portfolios of firms sorted by size and book-to-market	6FFEW, 6FFVW	6
Ten industry portfolios representing U.S. stock market	10IndEW, 10IndVW	10
Twenty-five Fama and French (1992) portfolios of firms sorted by size and book-to-market	25FFEW, 25FFVW	25
Forty-eight industry portfolios representing U.S. stock market	48IndEW, 48IndVW	48
One hundred Fama and French (1992) portfolios of firms sorted by size and book-to-market	96FFEW, 96FFVW	96
Top 100 market-value individual stocks with annual updates	100	100
Top 500 market-value individual stocks with annual updates	500	500

Note. We use EW (equally-weighted) and VW (value-weighted) to indicate the corresponding weighting type in the abbreviation.

via cross-validation over standard deviation. These portfolios are detailed in DeMiguel et al. (2009b, sections 3.1 and 3.2). The L_1-norm constrained portfolio is labeled as $\mathbb{L}_1$, and the L_2-norm constrained portfolio is labeled as $\mathbb{L}_2$.

Finally, we also include two relatively recent and well-performing benchmarks. The partial Min-Var portfolio, whose parameter is calibrated by maximizing the portfolio return in the previous period, is labeled as PARR and is detailed in DeMiguel et al. (2009b, section 3.3). Ledoit and Wolf (2017, section 3.4) introduce the nonlinear shrinkage method, which provides an excellent estimation of the covariance matrix. We call the corresponding portfolio the NonLin portfolio.

Evaluation Method. We report two performance measures: the out-of-sample standard deviation and the out-of-sample Sharpe ratio. The turnover discussion can be seen in Section 7.3. Following the convention of Brodie et al. (2009), DeMiguel et al. (2009b), and Fan et al. (2012), we use the "rolling-horizon" procedure, which uses a fixed-length training period to estimate. We denote the length of training period as $n < T$, where T is the total number of observations in the data set. As in DeMiguel et al. (2009b), we use $n = 120$ (10-year monthly return data). We construct various portfolios using the same training data. Then, we roll over to the next month, dropping the earliest month from the previous training window. This procedure yields $T - n$ portfolio-weight vectors for each portfolio. We denote the weight vector as $\boldsymbol{w}_t^i$ for $t = n, \ldots, T-1$ and for each portfolio i.

Following DeMiguel et al. (2009b), we hold the portfolio weight $\boldsymbol{w}_t^i$ for 1 month. This approach generates the out-of-sample return for time $t+1$: $r_{t+1}^i = (\boldsymbol{w}_t^i)'\boldsymbol{r}_{t+1}$, where $\boldsymbol{r}_{t+1}$ denotes the asset returns at time $t+1$. We use the time series of returns and weights to calculate the out-of-sample standard deviation and the out-of-sample Sharpe ratio:

$$(\hat{\sigma}^i)^2 = \frac{1}{T-n-1}\sum_{t=n}^{T-1}((\boldsymbol{w}_t^i)'\boldsymbol{r}_{t+1} - \hat{\mu}^i)^2, \quad \text{where}$$

$$\hat{\mu}^i = \frac{1}{T-n}\sum_{t=n}^{T-1}(\boldsymbol{w}_t^i)'\boldsymbol{r}_{t+1},$$

$$\widehat{SR}^i = \frac{\hat{\mu}^i}{\hat{\sigma}^i}.$$

We use Levene's test (Levene 1960) to calculate the statistical significance of the difference in the standard deviation. This test, with the sample median as an estimation of the location parameter, is favored in the literature because of its power and robustness against nonnormality (Brown and Forsythe 1974, Conover et al. 1981, Lim and Loh 1996). For the Sharpe ratio, we use the bootstrapping methodology proposed in Ledoit and Wolf (2008).

7.1. Out-of-Sample Standard Deviation

Table 5 shows that the BN portfolio achieves the best out-of-sample standard deviation on five out of the six large[11] portfolio data sets and is second-best on the 48IndEW data set. For all data sets, the BN portfolio is always significantly[12] better than the EW portfolio. Note that the BNvar portfolio has a higher out-of-sample standard deviation than the BN portfolio because it is expected to maximize the Sharpe ratio, and not to minimize the standard deviation. The results for the stock portfolios should be interpreted with caution, as these are aggregates over not perfectly comparable stock data sets. The BNvar portfolio does not exist for these two stocks data sets also because of the changing universe of stocks. The formulation of the BNvar portfolio includes a constant c [see Equation (10)]. The best choice of this c is on the entire historical data and not just the training set only. With a changing universe of stocks over the entire history, a consistent and justifiable choice would not be possible.

For the small data sets, the out-of-sample standard deviation of the EstMinVar portfolio is only about 1% larger than the best portfolio. This relationship indicates that 120 observations are enough for the small data sets to have the whole eigenspace as the signal space. Hence, the BN portfolio should not differ much from the EstMinVar portfolio, and indeed the correlation between their returns is more than 0.99. For the same reason, we expect cross-validation to determine very loose norm constraints for all the norm-constrained methods. Thus, their corresponding portfolios should be essentially the same as the EstMinVar portfolio. This result is again supported by the high correlation (about 0.99) between the returns of the norm-constrained portfolios and the EstMinVar portfolio. Meanwhile, the NoShorting portfolio's constraint cannot be relaxed, and, as expected, its performance suffers because its constraint interferes with portfolio selection using a well-estimated covariance matrix. However, it does better on some big data sets, where its constraint helps to avoid the effects of covariance estimation errors.

7.2. Discussion of Out-of-Sample Sharpe Ratio

Table 6 shows that, except for data set 48IndVW, the portfolio that has the highest Sharpe ratio is not the portfolio that has the lowest standard deviation.

Table 5. Out-of-Sample Monthly Standard Deviation in Percentage

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	4.473	4.062	3.571	3.604	**3.651**	**3.686**	3.640	**3.522**	**3.675**	**3.607**	3.477	3.302
BNVAR	5.105**	4.423*	3.984**	3.604	3.956	4.067**	3.964**	3.540	3.885	3.976*	NA	NA
EW	5.418***	4.916***	5.732***	4.308***	5.348***	5.107***	5.712***	4.900***	5.414***	5.204***	4.624***	4.795***
VW	5.133**	4.453	5.817***	4.031*	4.814***	4.409***	5.321***	4.347***	4.746***	4.424***	4.388***	4.386***
EstMinVar	4.474	4.059	3.559	3.609	3.858	3.878	5.984***	9.978**	7.172***	7.077***	6.499***	NA
NoShorting	4.870	4.377	3.605	3.615	4.614***	4.293**	**3.597**	3.694	4.506**	4.267**	3.482	3.332
$\mathbb{L}_1$	**4.415**	4.058	3.720	3.680	3.758	3.790	3.754	3.605	3.902	3.757	3.602	3.487
$\mathbb{L}_2$	4.468	4.066	**3.514**	**3.574**	3.703	3.697	3.697	3.588	3.723	3.651	**3.410**	3.133
PARR	4.652	4.154	4.518***	3.792	4.101**	3.981	4.783**	4.291**	5.244**	5.186**	5.157**	3.546
NonLin	4.469	**4.044**	3.545	3.583	3.690	3.717	3.662	3.651	3.732	3.666	3.435	**3.047**

Notes. This table reports the monthly out-of-sample standard deviation as a percentage. The numbers in bold are the smallest standard deviation for one data set. The p-value is calculated between the BN portfolio and other portfolios. The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter c. Because the sample covariance is degenerate, there is an NA of the estimated Min-Var portfolio.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The BNVAR portfolio has the best out-of-sample Sharpe ratio for 6 of the 10 portfolio data sets, and the dominance on these six data sets is both statistically and economically significant. These results show that we are indeed able to increase the out-of-sample Sharpe ratio for most data sets by allowing a higher variance level. The PARR portfolio achieves the best performance for four data sets, which is consistent with the results DeMiguel et al. (2009b). However, the analysis does not take transaction costs or taxation into account, which are crucial when turnover is high.

7.3. Robustness of Holding Length and Turnover

To get a sense of how portfolio performance depends on turnover, we compare the performance of the earlier monthly rebalanced portfolios with the annually rebalanced portfolios (Brodie et al. 2009). This allows us to evaluate the effects of turnover without making the results sensitive to either the type or the magnitude of transaction costs. The primary benefit here is that the performance measure now coincides with the objective, making it a fair comparison. The secondary benefit is that, from a taxation perspective, holding a portfolio 1 year also reduces the taxation rate from short term to long term. Olivares-Nadal and DeMiguel (2018) show that by penalizing the turnover in the portfolio-construction procedure, it is possible to sharply reduce the turnover without sacrificing much in performance.

Compared with Tables 5 and 6, Tables 7 and 8 show that the performance of the low-turnover portfolios (EW, VW, and NoShorting) remains similar. The BNVAR portfolio and the PARR portfolio see a significant drop and are no longer the best. This happens because both have high turnovers. The BNVAR portfolio, the NoShorting portfolio, and the $\mathbb{L}_1$ portfolio are the only ones that have a larger Sharpe

Table 6. Out-of-Sample Monthly Sharpe Ratio

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	0.398**	0.327	0.268	0.291	0.433**	0.353**	0.284	**0.280**	0.391	0.351*	**0.271**	0.289
BNVAR	**0.444**	**0.345**	0.309	0.291	**0.485**	**0.411**	0.325	0.274	**0.428**	**0.389**	NA	NA
EW	0.239***	0.236***	0.226	0.242	0.240***	0.238***	0.225*	0.222	0.237***	0.239***	0.202**	0.230*
VW	0.226***	0.226***	0.231	0.249	0.234***	0.235***	0.269	0.249	0.230***	0.236***	0.195**	0.210**
EstMinVar	0.398**	0.328	0.258	0.298	0.436**	0.361**	0.108***	0.120**	0.167***	0.169***	0.142***	NA
NoShorting	0.264***	0.247***	0.304	0.284	0.261***	0.242***	0.310	0.257	0.266***	0.260***	0.250	0.292
$\mathbb{L}_1$	0.395**	0.329	0.290	0.278	0.427**	0.345**	0.272	0.244	0.399	0.364	0.242	0.259
$\mathbb{L}_2$	0.398*	0.324	0.269	0.295	0.422**	0.350**	0.271	0.256*	0.391	0.359	0.238**	0.276
PARR	0.405	0.335	**0.369**	**0.343****	0.408**	0.360	**0.343**	0.271	0.248***	0.282**	0.166**	**0.345***
NonLin	0.393**	0.324	0.269	0.295	0.434**	0.358**	0.267*	0.239***	0.400	0.362	0.234**	0.284

Notes. This table reports the monthly out-of-sample Sharpe ratio. The number in bold is the largest Sharpe ratio for one data set. If the BNVAR portfolio is available, the p-value is calculated between it and other portfolios. If not, it is between the BN portfolio and others. The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter c. Because the sample covariance is degenerate, there is an NA of the estimated Min-Var portfolio.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 7. Hold for 1 Year, Out-of-Sample Monthly Standard Deviation in Percentage

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	4.835	4.549	4.437	3.571	3.930	**3.816**	4.922	**3.553**	**3.942**	**3.771**	3.536	3.430
BNVAR	6.094**	5.629	4.127	3.571	4.775*	4.834***	4.289	3.557	4.550	4.327*	NA	NA
EW	5.388**	4.911*	5.695***	4.276***	5.320***	5.109***	5.661***	4.843***	5.372***	5.203***	4.501***	4.633***
VW	5.128	**4.450**	5.788***	4.040*	4.796***	4.404***	5.300***	4.340***	4.740***	4.443***	4.380***	4.379***
EstMinVar	4.835	4.606	4.513	3.577	4.130	3.950	27.439***	11.896***	7.397***	7.417***	7.232***	NA
NoShorting	4.908	4.469	**3.628***	3.630	4.653***	4.353**	**3.634****	3.761	4.597***	4.364***	**3.522**	3.382
$\mathbb{L}_1$	4.860	4.607	3.746	3.642	4.034	3.935	4.372	3.682	4.126	4.006	3.789	3.357
$\mathbb{L}_2$	4.835	4.613	4.198	**3.540**	**3.922**	3.824	4.835	3.664	4.027	3.864	3.523	3.243
PARR	4.985	4.821	4.427	3.738	4.291***	4.473***	4.833**	4.255***	5.505***	6.292***	5.463***	3.511
NonLin	**4.796**	4.561	4.411	3.560	3.970	3.839	4.847	3.705	4.034	3.825	3.573	**3.228**

Notes. This table reports the monthly out-of-sample standard deviation as a percentage. The number in bold is the smallest standard deviation for one data set. The p-value is calculated between the BN portfolio and other portfolios. The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter c. Because the sample covariance is degenerate, there is an NA of the estimated Min-Var portfolio.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 8. Hold for 1 Year, Out-of-Sample Monthly Sharpe Ratio

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	0.369	0.302	0.210	0.289	**0.415**	0.344	0.216	**0.299**	0.385	0.349	**0.269**	0.287
BNVAR	0.362	0.279	0.247	0.289	0.405	0.344	0.271	0.291	0.370	0.346	NA	NA
EW	0.242	0.238	0.235	0.245	0.242**	0.239*	0.234	0.228*	0.239**	0.240**	0.202**	0.231*
VW	0.227*	0.226	0.234	0.250	0.236**	0.237*	0.272	0.255	0.233**	0.238**	0.196**	0.211**
EstMinVar	**0.369**	0.300	0.201	0.295	0.412	**0.359**	-0.027*	0.119***	0.192**	0.177***	0.141***	NA
NoShorting	0.265	0.243	**0.304**	0.280	0.261**	0.244*	**0.314**	0.256	0.262*	0.259**	0.251	0.287
$\mathbb{L}_1$	0.362	**0.302**	0.260	0.285	0.402	0.345	0.245	0.258*	0.395	0.356	0.249	0.271
$\mathbb{L}_2$	0.367	0.294	0.222	**0.296**	0.409	0.343	0.214	0.260***	0.389	0.357	0.248	0.292
PARR	0.351	0.271	0.223	0.270	0.391	0.318	0.224	0.210**	0.272*	0.250*	0.184**	0.273
NonLin	0.368	0.297	0.212	0.292	0.410	0.351	0.206	0.253***	**0.399**	**0.364**	0.243*	**0.293**

Notes. This table reports the monthly out-of-sample Sharpe ratio. The number in bold is the largest Sharpe ratio for one data set. If the BNVAR portfolio is available, the p-value is calculated between it and other portfolios. If not, it is between the BN portfolio and others. The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter c. Because the sample covariance is degenerate, there are NAs of the estimated Min-Var portfolio.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 9. Out-of-Sample Monthly Standard Deviation in Percentage Using 60 Observations

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	**4.268**	3.972	**3.463**	**3.563**	**3.740**	**3.708**	3.741	**3.504**	**3.796**	**3.707**	3.582	3.384
BNVAR	4.869**	4.470**	4.170***	3.563	3.938	4.060*	4.284***	3.698	3.861	3.896	NA	NA
EW	5.418***	4.916***	5.732***	4.308***	5.348***	5.107***	5.712***	4.900***	5.414***	5.204***	4.624*	4.795***
VW	5.133***	4.453**	5.817***	4.031**	4.814***	4.409***	5.321***	4.347***	4.746***	4.424***	4.388*	4.386***
EstMinVar	4.292	3.992	3.611	3.719	4.447***	4.381***	7.489***	11.168***	NA	NA	NA	NA
NoShorting	4.741*	4.296	3.565	3.610	4.518***	4.262*	3.665	3.615	4.453*	4.202*	3.553	3.341
$\mathbb{L}_1$	4.399	4.121	3.800	3.723	3.912	3.942	3.900	4.031**	4.286*	4.418*	3.928	3.462
$\mathbb{L}_2$	4.278	3.973	3.505	3.635	3.775	3.726	3.836	3.742	4.047	3.955	3.669	3.119
PARR	4.572	4.129	4.286***	3.773	4.345***	4.167*	5.213***	5.209***	4.549*	4.722*	4.177***	3.538
NonLin	4.278	**3.947**	3.518	3.616	3.742	3.770	**3.607**	3.590	3.822	3.782	**3.485**	**3.078**

Notes. This table reports the monthly out-of-sample standard deviation as a percentage. The number in bold is the smallest standard deviation for one data set. The p-value is calculated between the BN portfolio and other portfolios. To allow for a fair comparison with the 120-observation case, we truncate the return to the same period. The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter c. Because the sample covariance is degenerate, there are NAs of the estimated Min-Var portfolio.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 10. Out-of-Sample Monthly Sharpe Ratio Using 60 Observations

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	0.396*	0.309	0.266**	0.286	0.390***	0.324***	0.280*	0.256	0.360*	0.335	**0.266**	0.298
BNvar	**0.447**	**0.346**	0.338	0.286	**0.460**	**0.397**	**0.353**	**0.262**	**0.396**	**0.364**	NA	NA
EW	0.239***	0.236***	0.226*	0.242	0.240***	0.238***	0.225**	0.222	0.237***	0.239***	0.202*	0.230*
VW	0.226***	0.226***	0.231*	0.249	0.234***	0.235***	0.269	0.249	0.230***	0.236***	0.195**	0.210*
EstMinVar	0.421	0.324	0.236***	0.277	0.406*	0.314***	0.121***	0.081***	NA	NA	NA	NA
NoShorting	0.274***	0.259**	0.304	0.276	0.259***	0.244***	0.319	0.260	0.268***	0.257***	0.239	0.281
$\mathbb{L}_1$	0.422	0.327	0.314	0.264	0.402**	0.310***	0.299	0.236	0.295**	0.295**	0.185***	0.263
$\mathbb{L}_2$	0.423	0.315	0.257**	0.282	0.394**	0.324***	0.277*	0.225	0.327**	0.312**	0.208**	0.268
PARR	0.399	0.328	**0.421***	**0.347****	0.389**	0.329*	0.308	0.205	0.325*	0.321	0.236	**0.314**
NonLin	0.410	0.320	0.242**	0.284	0.416*	0.330**	0.277**	0.225*	0.361*	0.330	0.224**	0.275

Notes. This table reports the monthly out-of-sample Sharpe ratio. The number in bold is the largest Sharpe ratio for one data set. If the BNvar portfolio is available, the p-value is calculated between it and other portfolios. If not, it is between the BN portfolio and others. To allow for a fair comparison with the 120-observation case, we truncate the return to the same period. The NAs of the BNvar portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter c. Because the sample covariance is degenerate, there is an NA of the estimated Min-Var portfolio.

$*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

ratio than the EW portfolio across all the portfolio data sets.

7.4. Robustness of Training Length

In this subsection, following Brodie et al. (2009), we show the results using the same data sets but with only 60 (5-year monthly data) observations as training data. When the length of rolling window n is not larger than the number of assets p, the sample covariance matrix is singular.[13] Especially because the portfolio-construction problem assumes stationarity over n periods, small values of n are common. Hence, assessing the performance of portfolio optimization in the degenerate case (i.e., $n \leq p$) is important. By using 60 observations, the sample covariance matrix for data sets 96FFEW, 96FFVW, 100, and 500 are singular.

The results in Table 9 show that the BN portfolio is the best on 8 out of 10 portfolio data sets, including five (of six) large portfolio data sets, and the second best for the sixth. Comparing Table 5 to Table 9, we find that the out-of-sample standard deviation of the BN portfolio is robust to the choice of training length. We can make the same observation regarding Sharpe ratios. In fact, BNvar has the best Sharpe ratio for 8 (of 10) portfolio data sets (Table 10). The reason for the robustness is that both the BN portfolio and the BNvar portfolio become more cautious when training length becomes smaller. Indeed, the signal space becomes smaller and the noise bound, m, becomes larger when fewer observations are available. This happens because, given γ, the signal/noise split dictated by Definition 1 makes the signal space smaller, resulting in a larger noise space and a larger noise bound m.

As shown in Table 9, the out-of-sample standard deviations of the $\mathbb{L}_1$ portfolio and $\mathbb{L}_2$ portfolio increase significantly compared with those in Table 5. This change increases the margin between the standard deviations of the BN portfolio and other portfolios. For example, for the data set 96FFVW, the standard deviation of the BN portfolio is 6% better than that of the $\mathbb{L}_2$ portfolio

Table 11. Sensitivity Analysis of γ: $\gamma = 0.25$ as the Benchmark

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
Out-of-sample monthly standard deviation, using 120 observations												
BN$\gamma = 0.15$	−1.08%	−0.08%	0.28%	−1.20%	0.87%	0.47%	1.00%	2.16%	−0.18%	−0.57%	−1.09%	0.41%
BN$\gamma = 0.40$	0.01%	−0.07%	−0.44%	0.15%	−0.36%	0.18%	−0.70%	1.26%	1.04%	0.60%	1.01%	−1.03%
Out-of-sample monthly standard deviation, using 60 observations												
BN$\gamma = 0.15$	1.63%	−0.12%	0.33%	−0.03%	3.26%	−0.07%	0.26%	0.63%	−0.19%	−0.25%	0.06%	1.68%
BN$\gamma = 0.40$	0.65%	0.58%	0.96%	1.19%	0.33%	0.09%	−0.36%	1.52%	−0.65%	0.03%	0.12%	−0.41%
Out-of-sample monthly Sharpe ratio, using 120 observations												
BNvar$\gamma = 0.15$	2.19%	−0.35%	−0.63%	0.13%	−1.54%	−0.86%	−2.92%	0.68%	0.13%	0.57%	NA	NA
BNvar$\gamma = 0.40$	−0.21%	0.42%	−0.83%	2.67%	−0.48%	1.34%	0.61%	2.39%	−2.49%	0.87%	NA	NA
Out-of-Sample Monthly Sharpe Ratio, Using 60 Observations												
BNvar$\gamma = 0.15$	−1.03%	−2.29%	−0.80%	0.80%	−1.63%	−3.29%	1.35%	−1.71%	−0.09%	−1.44%	NA	NA
BNvar$\gamma = 0.40$	0.37%	−0.92%	1.10%	−0.25%	−2.46%	0.09%	1.08%	0.37%	−3.31%	−6.16%	NA	NA

Note. The NAs of the BNvar portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter c.

Table 12. Out-of-Sample Monthly Kurtosis

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	5.356	4.729	5.766	4.114	5.170	4.768	7.031	5.907	5.546	5.536	4.937	6.183
BNVAR	6.265	**4.199**	**4.536**	4.114	**4.716**	4.601	**5.205**	5.878	5.243	5.874	NA	NA
EW	6.380	6.015	6.343	5.290	6.415	6.078	6.866	6.145	6.353	6.070	4.970	5.783
VW	5.427	5.087	5.898	4.952	5.406	5.113	6.480	5.687	5.221	5.130	4.755	**5.105**
EstMinVar	5.356	4.748	5.800	4.062	5.209	4.667	91.026	323.514	**3.900**	**4.315**	**3.725**	NA
NoShorting	6.045	5.280	6.742	4.207	5.901	5.692	7.722	**5.563**	5.287	5.653	4.344	5.233
$\mathbb{L}_1$	5.612	4.762	6.833	4.156	4.974	4.952	7.214	5.915	5.633	5.339	4.305	6.311
$\mathbb{L}_2$	**5.350**	4.716	6.411	4.237	5.287	4.969	6.618	6.536	5.734	5.777	4.888	5.791
PARR	5.448	4.336	8.264	**3.925**	5.346	**4.322**	8.675	7.254	4.663	4.324	5.197	5.636
NonLin	5.433	4.664	5.791	4.083	5.204	4.555	5.826	6.448	5.492	5.285	4.735	5.642

Notes. This table reports the monthly out-of-sample kurtosis. The number in bold is the smallest kurtosis for one data set. The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter *c*. Because the sample covariance is degenerate, there is an NA of the estimated Min-Var portfolio.

Table 13. Hold for 1 Year, Out-of-Sample Monthly Kurtosis

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	5.669	10.916	15.239	3.968	5.781	4.842	22.989	6.288	5.873	5.648	4.961	5.993
BNVAR	16.534	42.615	**4.113**	3.968	15.018	12.866	**5.186**	6.173	10.550	7.382	NA	NA
EW	6.349	6.026	6.130	5.331	6.424	6.064	6.604	6.118	6.329	6.042	4.936	5.799
VW	**5.467**	**5.135**	5.692	5.018	5.421	5.132	6.504	5.745	5.113	**5.065**	4.759	5.117
EstMinVar	5.669	13.045	16.512	4.001	5.863	5.066	191.482	173.946	**4.341**	10.777	5.872	NA
NoShorting	6.359	5.293	6.667	4.190	5.748	5.623	7.239	**5.700**	5.355	5.478	**4.071**	**4.828**
$\mathbb{L}_1$	5.985	13.127	6.009	4.131	**5.120**	5.036	11.351	7.520	5.564	5.873	5.039	5.234
$\mathbb{L}_2$	5.651	12.961	11.516	4.109	5.410	5.003	18.356	7.317	6.160	6.215	5.241	5.558
PARR	5.628	11.732	5.426	**3.854**	5.136	5.555	5.652	5.965	6.795	9.995	5.117	5.105
NonLin	5.585	12.203	15.135	4.002	5.891	**4.746**	17.817	6.935	6.112	5.713	5.080	6.103

Notes. This table reports the monthly out-of-sample kurtosis. The number in bold is the smallest kurtosis for one data set. The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter *c*. Because the sample covariance is degenerate, there is an NA of the estimated Min-Var portfolio.

Table 14. Out-of-Sample Monthly Kurtosis Using 60 Observations

Portfolio	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
BN	6.256	6.231	6.655	4.983	5.592	6.472	6.540	6.905	6.179	6.672	4.998	6.154
BNVAR	**5.091**	5.170	**4.995**	4.983	5.426	6.294	**4.811**	6.780	6.007	6.317	NA	NA
EW	6.380	6.015	6.343	5.290	6.415	6.078	6.866	6.145	6.353	6.070	4.970	5.783
VW	5.427	**5.087**	5.898	4.952	5.406	**5.113**	6.480	5.687	5.221	**5.130**	4.755	**5.105**
EstMinVar	5.645	5.984	6.275	4.632	**5.195**	5.122	39.597	207.774	NA	NA	NA	NA
NoShorting	6.015	6.539	6.884	4.721	5.832	6.769	7.490	6.197	6.019	6.776	4.843	5.670
$\mathbb{L}_1$	6.218	6.262	6.660	5.151	5.654	6.135	11.028	6.570	**4.678**	6.326	5.985	5.572
$\mathbb{L}_2$	5.944	5.987	7.112	5.047	6.359	6.649	11.523	6.854	5.802	6.691	5.206	6.657
PARR	6.031	5.489	5.752	**4.429**	5.551	5.386	8.483	**5.133**	5.546	5.376	**4.357**	5.546
NonLin	5.679	6.033	6.584	4.811	5.523	5.927	5.523	7.004	6.316	6.546	5.161	6.391

Notes. This table reports the monthly out-of-sample kurtosis. The number in bold is the smallest kurtosis for one data set.The NAs of the BNVAR portfolio occur because the universe of stocks is changing and there are not enough data to learn the parameter *c*. Because the sample covariance is degenerate, there are NAs of the estimated Min-Var portfolio.

and 11% better than that of the $\mathbb{L}_1$ portfolio. The intuitive reason is that, unlike the bounded-noise procedure, cross-validation is unable to generate a more conservative portfolio when there are fewer data. In fact, in about 36% of the time periods, the penalty parameter [Equation (1)] with 60 observations η_{60} is smaller than η_{120}.

7.5. Robustness of Model Parameters

There are two model parameters in the BN portfolio: the number of bootstraps, L, and the cutoff of the amplification ratio, γ. We find that $L = 100$ generates almost identical results as $L = 1{,}000$, whose results are reported in the previous subsections.

Table 11 reports the sensitivity with respect to γ by measuring the percentage differences in standard deviations and Sharpe ratios of portfolios obtained using our preferred values ($\gamma = 0.25$) and alternative values ($\gamma = 0.15$ and 0.40). For the BN portfolio, the differences are around 1%. For the BNVAR portfolio, the differences are still mostly smaller than 2%.

7.6. Robustness to Alternative Measures of Tail Risk

To verify that optimization for variance does not lead to unwanted increases in tail risk, we compared the kurtosis for all methods. Table 12 reports the monthly out-of-sample kurtosis by using 120 previous observations, whereas Tables 13 and 14 provide robustness checks for 1-year holding period and 60 observations, respectively. EstMinVar performs particularly poorly for the 48IndEW and 48IndVW data sets, but otherwise, all the methods are roughly similar. These results suggest that the outperformance of BN and BNVAR may not have come at the expense of large increases in tail risks.

8. Concluding Remarks

The essence of the paper lies in recognizing that the primary problem in constructing well-performing portfolios does not come from estimation alone. Errors in the estimation are amplified by the optimization step, resulting in even unbiased small errors causing biased and unacceptable errors in portfolio weights. The usual route to fix this is by either trying to improve estimation or adjusting the optimization step in an arbitrary manner, which may reduce the impact of estimation errors. Instead, we disentangle the covariance matrix into two parts: one that behaves well in the optimization step, which we call the signal part, and another that does not, which we call the noise part. We detail and discuss the way to split, how we can construct portfolios from each of these, why the noise is useful, how to combine the two portfolios, relevant mathematical justifications, relations to other methods, an extension that allows the construction of mean-variance portfolios, and, finally, evidence of superior performance using both the simulated and the real-world data.

There are several aspects of portfolio analysis that could benefit from further investigation. The signal/noise split and the related optimal portfolios rely heavily on the investor having no additional constraints. Extending the splitting idea in the context of optimal portfolios with additional constraints is very valuable, although challenging. Pushing the ideas in the paper, more along the mean-variance direction, would be another good direction for future work. The method we describe for constructing mean-variance portfolios does not directly deal with uncertainty in estimates of the mean returns. Another useful extension would be to allow the user to specify shocks or black-swan events and construct a portfolio that could be robust against such events. Finally, regarding extending the core idea, although we consider a hard split between signal and noise eigenvectors, there is a continuum. A careful characterization of every eigenvector along this continuum may lead to better performance. However, we believe that this would only have a second-order improvement. Similarly, we computed the split and noise bound via a median of bootstrap samples. A more careful analysis could use the full distributions derived from these samples.

Acknowledgments

The authors thank Sheridan Titman, Stathis Tompaidis, Paat Rusmevichientong, Courtney Paulson, Victor DeMiguel, John Birge, Olivier Ledoit, Michael Wolf, Chanaka Edirisinghe, Rui Gao; the two anonymous referees; the associate editor; the area editor; Steven Kou; and the Center for Analytics and Transformative Technologies at McCombs for all their help, feedback, suggestions, and support.

Endnotes

[1] Note that eigenvectors of the covariance matrix are precisely the principal components of the data (whose mean has been removed).

[2] For a more detailed discussion, please see Ledoit and Wolf (2012, 2017) and the references therein.

[3] Our definition of realized variance is slightly different from some in the literature. For example, Hansen and Lunde (2006) directly use the square of returns without subtracting the sample mean. This definition is reasonable when the sample mean is close to 0 and much smaller than the sample variance. This argument is validated in studies that use daily data. However, we use monthly data, and the sample mean is not negligible.

[4] If $\lambda_i = \lambda_{i+1}$, it is impossible to estimate v_i or v_{i+1} accurately. However, their amplification ratios can be close to 1.

[5] If one is extremely concerned about the portfolio from the noise space, one can assign infinity as the upper bound for all these portfolios. This leads to the signal-only portfolio.

[6] The performance suffers when the upper bound is less than 1.5 because we have not taken advantage of enough information. The result is almost the same for values both at and larger than 1.5.

[7] The penalty parameter is chosen by leave-one-out cross-validation, as in DeMiguel et al. (2009b). We do a bisection search within the interval $[10^{-4}, 10^4]$ to find the parameter with the lowest cross-validated standard deviation. This "best" parameter is then used to build a portfolio using the entire 120 monthly returns.

[8] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[9] We only include the stocks whose returns are available for the past 10 years and the future 1 year.

[10] The number of asset changes for each update is 2.5 and 50 on average for the 100 and 500 stock data sets, respectively.

[11] We use the phrase large data sets when the number of assets, p, is larger than 10.

[12] $p < 0.05$ (Levene's test).

[13] In the calculation of the sample covariance matrix, the sample mean is subtracted. Thus, when $n \leq p$, the rank of the sample covariance matrix is at most $n - 1$, which is smaller than p.

References

Black F, Litterman R (1992) Global portfolio optimization. *Financial Anal. J.* 48(5):28–43.

Brodie J, Daubechies I, De Mol C, Giannone D, Loris I (2009) Sparse and stable Markowitz portfolios. *Proc. Natl. Acad. Sci. USA* 106(30):12267–12272.

Brown MB, Forsythe AB (1974) The ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics* 30(4): 719–724.

Chakrabarti D, Faloutsos C (2006) Graph mining: Laws, generators, and algorithms. *ACM Comput. Survey* 38(1):2.

Chopra VK, Ziemba WT (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. *J. Portfolio Management* 19(2):249–257.

Conover WJ, Johnson ME, Johnson MM (1981) A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 23(4):351–361.

DeMiguel V, Nogales FJ (2009) Portfolio selection with robust estimation. *Oper. Res.* 57(3):560–577.

DeMiguel V, Garlappi L, Uppal R (2009a) Optimal vs. naïve diversification: How inefficient is the 1/n portfolio strategy? *Rev. Financial Stud.* 22(5):1915–1953.

DeMiguel V, Garlappi L, Nogales FJ, Uppal R (2009b) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Sci.* 55(5):798–812.

DeMiguel V, Martin-Utrera A, Nogales FJ (2013a) Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *J. Bank. Finance* 37(8):3018–3034.

DeMiguel V, Plyakha Y, Uppal R, Vilkov G (2013b) Improving portfolio selection using option-implied volatility and skewness. *J. Financial Quant. Anal.* 48(6):1813–1845.

Duchin R, Levy H (2009) Markowitz vs. the Talmudic portfolio diversification strategies. *J. Portfolio Management* 35(2):71–74.

Fama EF, French KR (1992) The cross-section of expected stock returns. *J. Finance* 47(2):427–465.

Fama EF, French KR (2015) A five-factor asset pricing model. *J. Financial Econom.* 116(1):1–22.

Fan J, Zhang J, Yu K (2012) Vast portfolio selection with gross-exposure constraints. *J. Amer. Statist. Assoc.* 107(498):592–606.

Frahm G, Memmel C (2010) Dominating estimators for minimum-variance portfolios. *J. Econometrics* 159(2):289–302.

Frost PA, Savarino JE (1986) An empirical Bayes approach to efficient portfolio selection. *J. Financial Quant. Anal.* 21(3):293–305.

Frost PA, Savarino JE (1988) For better performance: Constrain portfolio weights. *J. Portfolio Management* 15(1):29–34.

Gotoh J-y, Takeda A (2011) On the role of norm constraints in portfolio selection. *Comput. Management Sci.* 8(4):323–353.

Green R, Hollifield B (1992) When will mean-variance efficient portfolios be well diversified? *J. Finance* 47(5):1785–1809.

Hansen PR, Lunde A (2006) Realized variance and market microstructure noise. *J. Bus. Econom. Statist.* 24(2):127–161.

Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Finance* 58(4):1651–1684.

Jobson JD, Korkie B (1980) Estimation for Markowitz efficient portfolios. *J. Amer. Statist. Assoc.* 75(371):544–554.

Jobson JD, Korkie RM (1981) Putting Markowitz theory to work. *J. Portfolio Management* 7(4):70–74.

Jorion P (1986) Bayes-Stein estimation for portfolio analysis. *J. Financial Quant. Anal.* 21(3):279–292.

Laloux L, Cizeau P, Bouchaud JP, Potters M (1999) Noise dressing of financial correlation matrices. *Phys. Rev. Lett.* 83(7):1467.

Laloux L, Cizeau P, Potters M, Bouchaud JP (2000) Random matrix theory and financial correlations. *Internat. J. Theor. Appl. Finance* 3(3):391–397.

Lauprête GJ (2001) Portfolio risk minimization under departures from normality. PhD thesis, Massachusetts Institute of Technology, Cambridge.

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance* 10(5):603–621.

Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* 88(2): 365–411.

Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the Sharpe ratio. *J. Empirical Finance* 15(5):850–859.

Ledoit O, Wolf M (2012) Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* 40(2):1024–1060.

Ledoit O, Wolf M (2017) Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financial Stud.* 30(12):4349–4388.

Levene H (1960) Robust tests for equality of variances. Olkin I, Ghurye SG, Hoeffding W, Madow WG, Mann HB, eds. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (Stanford University Press, Stanford, CA), 278–292.

Lim TS, Loh WY (1996) A comparison of tests of equality of variances. *Comput. Statist. Data Anal.* 22(3):287–301.

Markowitz H (1952) Portfolio selection. *J. Finance* 7(1):77–91.

Merton RC (1980) On estimating the expected return on the market: An exploratory investigation. *J. Financial Econom.* 8(4):323–361.

Michaud RO (1989) The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Anal. J.* 45(1):31–42.

Mihail M, Papadimitriou C (2002) On the eigenvalue power law. *Internat. Workshop Randomization Approximation Techniques Comput. Sci.* (Springer, Berlin), 254–262.

Olivares-Nadal AV, DeMiguel V (2018) Technical note: A robust perspective on transaction costs in portfolio optimization. *Oper. Res.* 66(3):733–739.

Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE (2002) Random matrix approach to cross correlations in financial data. *Phys. Rev. E* 65(6):066126.

Tu J, Zhou G (2011) Markowitz meets Talmud: A combination of sophisticated and naïve diversification strategies. *J. Financial Econom.* 99(1):204–215.

Vershynin R (2011) How close is the sample covariance matrix to the actual covariance matrix? *J. Theoret. Probab.* 25(3):655–686.

Welsch RE, Zhou X (2007) Application of robust statistics to asset allocation models. *REVSTAT* 5(1):97–114.

Yu Y, Wang T, Samworth RJ (2015) A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika* 102(2):315–323.

Long Zhao is a PhD candidate at the McCombs School of Business, University of Texas at Austin. His research interest is data-driven decision making.

Deepayan Chakrabarti is an assistant professor in the Department of Information, Risk, and Operations Management at McCombs School of Business, University of Texas at Austin. He received his PhD from Carnegie Mellon University in 2005. His research includes statistical modeling and analysis of networks and decision making under uncertainty.

Kumar Muthuraman is the H. Timothy Harkins Centennial Professor in the Department of Information, Risk, and Operations Management and Department of Finance at McCombs School of Business, University of Texas at Austin. He received his PhD from Stanford University in 2003. His research interests are primarily in analytics and decision making.