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

Long Zhao, Deepayan Chakrabarti, and Kumar Muthuraman†
McCombs School of Business, University of Texas, Austin, TX, USA,
long.zhao@mccombs.utexas.edu,
deepayan.chakrabarti@mccombs.utexas.edu,
kumar.muthuraman@mccombs.utexas.edu

This paper focuses on the problem of poor portfolio performance when a minimum-variance portfolio is constructed using the sample estimates. This issue is well documented in the literature and has remained in the spotlight ever since Markowitz (1952). Estimation errors are mostly blamed for this problem. However, we argue that even small unbiased estimation errors can lead to significantly bad performance because the optimization step amplifies errors, that too in a non-symmetric 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 and handle them differently and appropriately. By using a single parameter held constant over all datasets and time periods, our method achieves excellent performance both empirically and in simulation. Finally, we show how to use information from the sample mean to construct mean-variance portfolios, which we demonstrate have higher out-of-sample Sharpe ratios.

Key words: portfolio choice, estimation error
History:

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 work is considered to be among the most fundamental works in the finance literature helping to initiate an era of mathematical analysis of financial problems. However, the out-of-sample performance of these mean-variance portfolios in the real-world often has been demonstrated to be unacceptable (Jobson and Korkie 1981, Frost and Savarino 1986, 1988,

* The authors thank Sheridan Titman, Stathis Tompaidis, Paat Rusmevichientong, Courtney Paulson, Victor DeMiguel, John Birge, Olivier Ledoit, Michael Wolf, Chanaka Edirisinghe, the two anonymous referees, the associated editor and the area editor Steven Kou for their help, feedback and suggestions.
† Corresponding Author

Jorion 1986, Michaud 1989). This poor performance stems largely from our inability to make precise parameter estimates. Even the simpler variance-minimizing portfolio has been shown to have a similarly unacceptable performance (Jagannathan and Ma 2003, DeMiguel et al. 2009b).

The most direct solution to the minimum-variance 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 the optimal portfolio. Both steps are relatively straightforward. However, even the best covariance estimation in the first step, though unbiased, has errors. When we use this estimate of the covariance matrix in place of the true covariance, the hope is that the resulting portfolio still is close to the true 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.

Indeed, this amplification is in a sense non-symmetric; that is, different kinds of errors are amplified in different ways. Hence, an unbiased initial error in estimation does not translate to portfolios that are unbiased estimates of the optimal portfolio. The compounding effect of the optimization-driven error amplification on the initial estimation errors is the primary cause of this unacceptable performance.

A plethora of research papers suggests 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. (We discuss several of these papers in section 1.3). However, DeMiguel et al. (2009b) consider 14 popular methods in the literature and show that none performs consistently better than the naive equal-weighted portfolio. They examine the methods across seven monthly datasets, evaluating the Sharpe ratio, the certainty-equivalent return, and turnover. Later, a few papers (Brodie et al. 2009, DeMiguel et al. 2009a, Fan et al. 2012) demonstrate that the *norm-constrained* portfolios, which belong to the second category that modifies the optimization, outperform several other competing methods on many real-world financial datasets. Instead of just minimizing portfolio variance in the second step, 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 themselves as large weights of some assets, and penalizing portfolio weights limits this problem. Norm-constrained portfolios have been shown to be mathematically equivalent to several other methods (see section 1.3), indicating that this same basic idea underpins many seemingly disparate models.

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 true optimal portfolio can have sizeable asset weights.

Hence, although norm constraints might help, they also might be wrong because they exclude the true optimal solutions, which involve 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 optimal choice of this parameter is often fine-tuned using out-of-sample data to demonstrate performance. Moreover, the optimal value of this parameter varies based on the particular financial dataset and the amount of historical data used, and it even changes over the time horizon encompassed by a dataset. These problems motivate us to search for a deeper understanding of the dynamics of error propagation in portfolio optimization and use this understanding to construct well-understood portfolios that have good out-of-sample performance.

1.1. Our Main Ideas

In this paper, we first try to tease out the reason for the poor performance of the minimum-variance portfolios in the real world. Limited data lead to estimation errors, which we argue are amplified by the optimization procedure to cause the unacceptable performance. Instead of trying to independently improve the estimation step or fix the optimization step for robustness, we try to disentangle the well-estimated aspects from the poorly-estimated of the covariance matrix and handle them differently and appropriately when constructing our portfolio.

Our approach has four steps. First, we begin by looking into the estimation of the covariance matrix. It turns out that some eigenvectors of the covariance matrix are easier to estimate than others.[1] We show that the portfolio weights that result from solving the optimization problem depend to a greater degree on the poorly-estimated eigenvectors, which suggests the need to split the set of eigenvectors into two groups: the well-estimated and the poorly-estimated. However, instead of splitting 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 directly construct a *signal-only* portfolio from the well-estimated signal eigenvectors. This portfolio by itself performs significantly better than the classical minimum-variance portfolio.

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. This observation is important because a portfolio from the noise space has the potential to improve

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

performance when combined with the signal-only portfolio. We show this by devising an upper bound on the true variance of any portfolio constructed from these noisy eigenvectors and use this upper bound to build a *conservative noise-only* portfolio.

Finally, we show how to combine the signal-only portfolio with the conservative noise-only portfolio to generate a single portfolio that outperforms the signal-only portfolio. Using simulated data and twelve standard datasets (listed in Table 1) with different rebalance frequencies and training lengths, we then show that our method yields portfolios that do well not only in simulation but also on these real-world datasets. Moreover, unlike norm-constrained portfolios, we use the same value of the scalar threshold parameter that defines the split for all datasets. 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 physically meaningful parameter whose value is invariant to the financial datasets and to the length and time of the historical data.

As can be noted from the description of our contributions, we explore the performance of the proposed methodology in multiple ways: testing on both the real-world and the simulated data and providing mathematical justifications. Performance on the real-world financial data is, of course, an important indicator, but it comes with two caveats. First, it does not allow us to understand the effects of modeling and estimation errors separately. Second, we also risk being at the mercy of a few sample paths, making it harder to establish that we are not being favored by certain datasets only. It is entirely understandable that performance in the real world is the ultimate price and this is the reason most methods in the literature use performance of the real-world datasets as their only evaluation mechanism. However, (a) testing on simulated data and (b) providing mathematical insights into why a specific method performs well, together with (c) testing on the real-world data, would make a much stronger case. Testing on simulated data allows for exhaustive tests and keeps the focus on estimation errors alone. From an academic perspective, understanding mathematically why a method does well is reassuring and will enable us to understand the method's limitations, enabling further improvement. Hence in this paper, we do all three.

1.2. Other contributions

Mean-variance portfolios, as opposed to minimum-variance portfolios, also use the estimated expected returns to construct portfolios with an expected return target. 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, DeMiguel et al. 2009a, Brodie et al. 2009, Fan et al. 2012) has mostly focused on the minimum-variance problem to bypass this issue. However, expected returns are important drivers of the Sharpe ratio. In section 4 and 6, 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. It turns out that by bounding the out-of-sample variance, our method is more tolerant to sample mean errors, allowing us to achieve a higher Sharpe ratio.

We also provide a detailed discussion on the connection between our method and the norm-constrained methods. Our analysis shows that the best performing norm-constrained portfolio corresponds to a wrong constraint, which in fact could render the true optimal portfolio infeasible. We show that a penalized norm avoids error amplification indirectly, which is why, rather paradoxically, such a wrong constraint can work.

1.3. Literature Review

The minimum-variance portfolio is 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 MinVar portfolio in place of the true MinVar portfolio. The poor out-of-sample performance of the estimated MinVar 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. The author 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 naive equal-weighted portfolio that spreads the budget equally among all assets performs better (DeMiguel et al. 2009b). We can group the papers trying to overcome this problem into three categories. The first category tries to develop methods that provide better covariance estimates than the sample covariance matrix. The second category combines the estimated MinVar portfolio with the equal-weighted portfolio to maximize a utility measure other than variance. The third category tries to modify 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

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

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 shrink 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 multi-factor 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 the simulated and the real-world datasets 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 minimum-variance portfolios on two real-world datasets.

Combining with the equal-weighted portfolio: The second category is inspired by the good performance of the equal-weighted portfolio documented in the literature (Jobson and Korkie 1980, DeMiguel et al. 2009b, Duchin and Levy 2009). With five reasonable assumptions, Frahm and Memmel (2010) prove that the portfolio constructed by carefully combining the estimated MinVar 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 dataset evaluation, they take the equal-weighted portfolio as the reference portfolio and demonstrate the benefit of combination. By minimizing the expected utility loss, Tu and Zhou (2011) estimate the combination level of each of four different portfolios and the equal-weighted portfolio. Using an exhaustive assessment of both the simulated and the real-world datasets, they show that the new portfolios perform better than the equal-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 datasets.

Modifying the optimization: The third category modifies the portfolio optimization 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 insightfully 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 the simulated and the real-world data. A weaker version of the no-shorting constraint involves penalizing a norm of the portfolio weights,

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

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. 2009a). Among these studies, Fan et al. (2012) is the only one that uses both the simulated and the 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 dataset is used. DeMiguel et al. (2009a) provide more comprehensive empirical results, and they show that the norm-constrained portfolios dominate the equal-weighted portfolio and the estimated MinVar portfolio in terms of both the out-of-sample variance and Sharpe ratio. They also build the connection between norm-constrained portfolios and Bayesian priors on the sample covariance matrix. Gotoh and Takeda (2011) find that the norm constraints are equivalent to 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.

Our approach is complementary to each of the three categories discussed. Estimation error 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. Our discussion of the causes of this amplification and the way to mitigate it are relevant here. The second category is based on the good performance of the equal-weighted portfolio. We provide theoretical reasons for its good performance and indeed show that a new way of combining portfolios can yield even better performance. The third category penalizes the norm of the portfolio weights, but the penalty factor and the norm must be chosen for each dataset. This parameter is typically chosen repeatedly via cross-validation and is quite sensitive. Even for the same dataset, different training subsets usually give different parameter choices. We show that the norm penalty avoids error amplification indirectly, and also show how the right penalty can be chosen using a single constant parameter value that applies to every period of every dataset. We explain the mathematical justification for our method and demonstrate its efficacy on both the simulated and twelve real-world datasets of varying sizes and characteristics.

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 vs. 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 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 our case. 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.

1.4. Outline

The rest of the paper is organized as follows. In section 2, by adopting an in-depth understanding of the estimation errors in the sample covariance estimation, we discuss how certain errors are amplified via the optimization solver, which results in poor portfolio performance. In section 3, by mitigating the error amplification, we construct the bound-noise portfolio. Section 4 demonstrates how to extend the bound-noise idea from minimizing the variance to maximizing the Sharpe ratio. In section 5, we provide the connection with several existing portfolio optimization methods. Section 6 provides exhaustive comparisons of our portfolios with eight other different portfolio construction methods, using twelve datasets with two different rebalancing frequencies and training lengths. Concluding remarks and future research directions are offered in section 7.

2. 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. This section describes the estimation errors and how they get amplified in the optimization step. While doing this, we also represent our signal and noise space and how portfolio in each space can be combined.

2.1. Estimation Error

Let's begin with Proposition 2.1 which shows that the relative errors in estimating the large eigenvalues of the true covariance matrix are small while 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. $||\cdot||_{op}$ stands for the operator norm. The sample size is n, and the number of assets is p.

Proposition 2.1 (Eigenvalue Concentration) *Let λ_i and $\hat{\lambda}_i$ represent the i^{th} largest eigenvalues of Σ and $\hat{\Sigma}$, respectively. Then we have:*

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

Estimation errors for the eigenvectors are a bit more complicated to characterize. Lemma 2.2 show that the estimation error not only depends on $\|\Sigma - \hat{\Sigma}\|_{op}$, but also how separated the eigenvalues are.

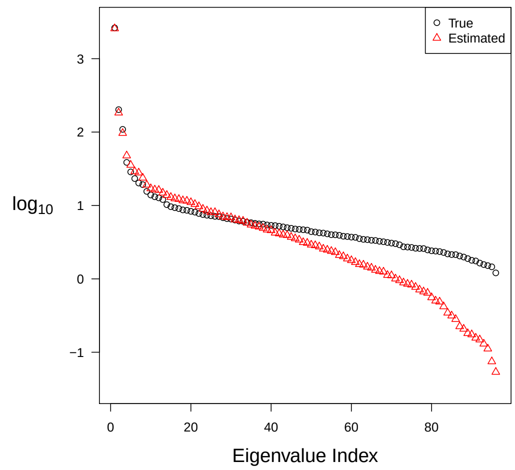


Figure 1 Distribution of True and Estimated Eigenvalues

Lemma 2.2 (Concentration of Eigenvectors (Yu et al. 2015)) *Let* $\Sigma, \hat{\Sigma} \in \mathbb{R}^{p\times p}$ *be symmetric, with eigenvalues* $\lambda_1 \geq \ldots \geq \lambda_p$ *and* $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p$*, respectively. Fix* $1 \leq r \leq s \leq p$*, and assume that* $\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}) > 0$*, where we define* $\lambda_0 = \infty$ *and* $\lambda_{p+1} = -\infty$*. Let* $d = s - r + 1$*. Let* $V = (\boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_s) \in \mathbb{R}^{p\times d}$ *and* $\hat{V} = (\hat{\boldsymbol{v}}_r, \hat{\boldsymbol{v}}_{r+1}, \ldots, \hat{\boldsymbol{v}}_s) \in \mathbb{R}^{p\times d}$ *have orthogonal columns satisfying* $\Sigma \boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j$ *and* $\hat{\Sigma}\hat{\boldsymbol{v}}_j = \hat{\lambda}_j \hat{\boldsymbol{v}}_j$*; then there exists an orthogonal matrix* $\hat{O} \in \mathbb{R}^{d\times d}$ *such that*

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

Vershynin (2011) gives a nice 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 $\sqrt{p/n}$. Thus, for a given number of assets p, the difference decays when more observations are available, as expected.

Previous work on financial datasets 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) while the bulk of the eigenvalues are small (so their eigenvectors just have a small contribution to the variance of asset returns).

This intuition is supported by the observations from a historical covariance matrix constructed from the monthly returns of the Fama-French value-weighted dataset 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 these 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).

In summary, we can separate the eigenvalues and the corresponding eigenvectors into two parts. *The largest eigenvalues and related eigenvectors in the sample covariance $\hat{\Sigma}$ can relatively well approximate the corresponding eigenvalues and eigenvectors of the true covariance matrix Σ. We call these pairs the signal. The smaller eigenvalues and the corresponding eigenvectors are poor estimations. We call these pairs the noise.*

2.2. Error Amplification in Portfolio Optimization

The previous discussion showed that the estimation errors affect the smaller eigenvalues and eigenvectors more than the larger ones. To understand how these differences influence portfolio optimization, we first give a new characterization of the true MinVar portfolio.

Separate the true eigenvectors $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$ into two sets: from index 1 to k, and from $k+1$ to p. Intuitively, we expect the first set to contain better-estimated eigenvectors than the second set. Denote the space spanned by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ as $\mathcal{S}$ and the space spanned by the other eigenvectors as $\mathcal{N}$.

Lemma 2.3 (Portfolio Decomposition) *For any separation ($\mathcal{S}$, $\mathcal{N}$), the optimal portfolio $\boldsymbol{w}^*$ can be expressed as*

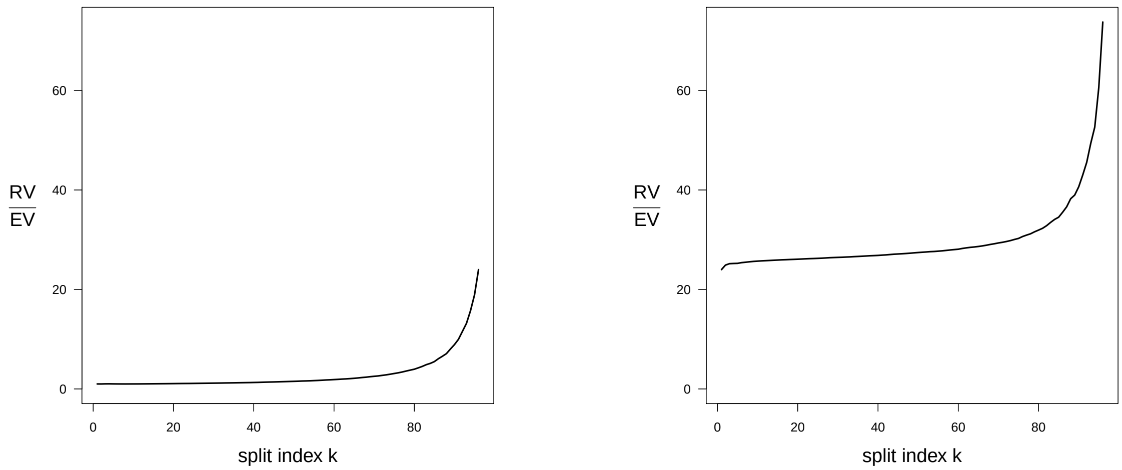
$$\boldsymbol{w}^* = \alpha \boldsymbol{w}_S^* + (1-\alpha)\boldsymbol{w}_N^* \tag{2}$$

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

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

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

Namely, $\boldsymbol{w}_S^*$ and $\boldsymbol{w}_N^*$ are the solution of the minimum-variance problem restricted being a linear combination of the first k eigenvectors (the vectors that span $\mathcal{S}$) and other eigenvectors, respectively.



(a) The signal-only portfolio $\hat{\boldsymbol{w}}_S^*$ (b) The aggressive noise-only portfolio $\hat{\boldsymbol{w}}_N^*$

Figure 2 The ratio between RV and EV

In the above, $RV(\boldsymbol{w})$ is the expected out-of-sample variance (henceforth, the realized variance[3]*) of $\boldsymbol{w}$, namely,*

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

Thus, the true MinVar 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 MinVar portfolio. It can be expressed in the same form as in Lemma 2.3, but now the true parameters are replaced with their estimated counterparts. In particular, the eigenspace $\mathcal{S}$ is replaced by $\hat{\mathcal{S}} = \text{span}(\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_k)$; $\mathcal{N}$ is replaced by $\hat{\mathcal{N}} = \text{span}(\hat{\boldsymbol{v}}_{k+1}, \ldots, \hat{\boldsymbol{v}}_p)$; 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 $\hat{\boldsymbol{w}}_S^*$ to $\hat{\boldsymbol{w}}_N^*$ in the overall portfolio $\hat{\boldsymbol{w}}^*$ (Eq. 3) is now driven by estimated variance instead of realized variance.

To further illustrate the difference between the realized variance and the estimated variance, we perform simulations on the previously mentioned Fama-French value-weighted dataset comprising 96 stocks. In the simulation, we assume that the true covariance matrix Σ and the true expected return $\boldsymbol{\mu}$ are the sample covariance matrix and the sample mean using all monthly data from July

[3] Our definition of realized variance is slightly different from some of 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 papers that use daily data. However, we use monthly data, and the sample mean is not negligible.

1963 to July 2015 (625 observations). We also assume that the returns follow a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ, and we draw 120 observations (10-year monthly data) from this distribution.

We calculate the realized variance and the estimated variance for various signal-noise splits. We obtain Figure 2 by repeating this experiment 100 times and calculating related averages. It shows the ratio of realized variance to estimated variance for $\hat{\boldsymbol{w}}_S^*$ and $\hat{\boldsymbol{w}}_N^*$. As discussed in the previous subsection, we expect the realized variance of $\hat{\boldsymbol{w}}_S^*$ to be similar to its estimated variance when k is small. Figure 2a supports this intuition. However, Figure 2b shows that for $\hat{\boldsymbol{w}}_N^*$, its realized variance is much larger than its estimated variance. Indeed, it is at least 20 times larger for any k. This underestimation means that $\hat{\boldsymbol{w}}_N^*$, which uses the poorly-estimated parameters, gets overweighted significantly. We call $\hat{\boldsymbol{w}}_N^*$ the *aggressive noise-only portfolio* and the ratio between the realized variance and the estimated variance the *amplification ratio*.

3. The Bounded-Noise Portfolio

The previous discussion shows the utility of separating a "signal" space $\hat{\mathcal{S}}$ (and the *signal-only portfolio* $\hat{\boldsymbol{w}}_S^*$) from a "noise" space $\hat{\mathcal{N}}$ (and the *aggressive noise-only portfolio* $\hat{\boldsymbol{w}}_N^*$) 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 3.1 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 3.2. To take advantage of the information contained in the noise space, in Section 3.3, 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 3.4. 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 3.5.

3.1. 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 2.2 shows the under-estimation of the realized variance by the estimated variance. This under-estimation leads to the aggressive noise-only portfolio being overweighted in the estimated Min-Var portfolio. Hence we characterize the signal space as all eigen pairs $(\hat{\lambda}_i, \hat{\mathbf{v}}_i)$ 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 6.5. Because the realized variance is unknown, ϕ_i needs to be estimated and we provide the procedure in Section 3.5. 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** *while 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:

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

where the last inequality follows from Proposition 2.1 and the definition of the operator norm.

3.2. 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, $\hat{\boldsymbol{w}}_S^*$, constructed from these sample eigenvectors should also be reliable. Mathematically speaking, this portfolio is equivalent to a PCA-based portfolio that ignores a certain number of the low eigenvalues of $\hat{\Sigma}$ and corresponding eigenvectors.

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

3.3. 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 3.1 (Bounding Realized Variance of any Portfolio from the Noise Space) *Let the noise space eigenvectors of* $\hat{\Sigma}$ *be* $\hat{\boldsymbol{v}}_{k+1}, \ldots, \hat{\boldsymbol{v}}_p$*, the space spanned by them be* $\hat{\mathcal{N}}$*, and the matrix whose columns are* $(\hat{\boldsymbol{v}}_{k+1}, \ldots, \hat{\boldsymbol{v}}_p)$ *be* $\hat{N}$*. For any* $\boldsymbol{w} \in \hat{\mathcal{N}}$*,*

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

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

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 3.5. The realized variance of any portfolio $\boldsymbol{w}$ from the noise space can be upper-bounded by a function $BRV(\boldsymbol{w})$, which is defined as $BRV(\boldsymbol{w}) = EV(\boldsymbol{w}) + m||\boldsymbol{w}||_2^2$. 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:

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

This portfolio is not necessarily close to the optimal noise portfolio $\boldsymbol{w}_N^*$. 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 $\hat{\boldsymbol{w}}_N^{BN}$.

3.4. Combining the Two Portfolios

Finally, we must combine the signal-only portfolio, $\hat{\boldsymbol{w}}_S^*$, with the conservative noise-only portfolio, $\hat{\boldsymbol{w}}_N^{BN}$, 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:

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

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:

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

3.5. Estimating k^* and m

Both k^* and m are functions of Σ and $\hat{\Sigma}$. Since 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 3.1. 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.

[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.

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

$$\hat{k} = \max\left\{k \mid \text{median}\{\tilde{\phi}_{ji} \mid j \in [1, L]\}) \leq 1 + \gamma = 1 + 0.25 = 1.25,\ \forall i \leq k\right\}.$$

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

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

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.

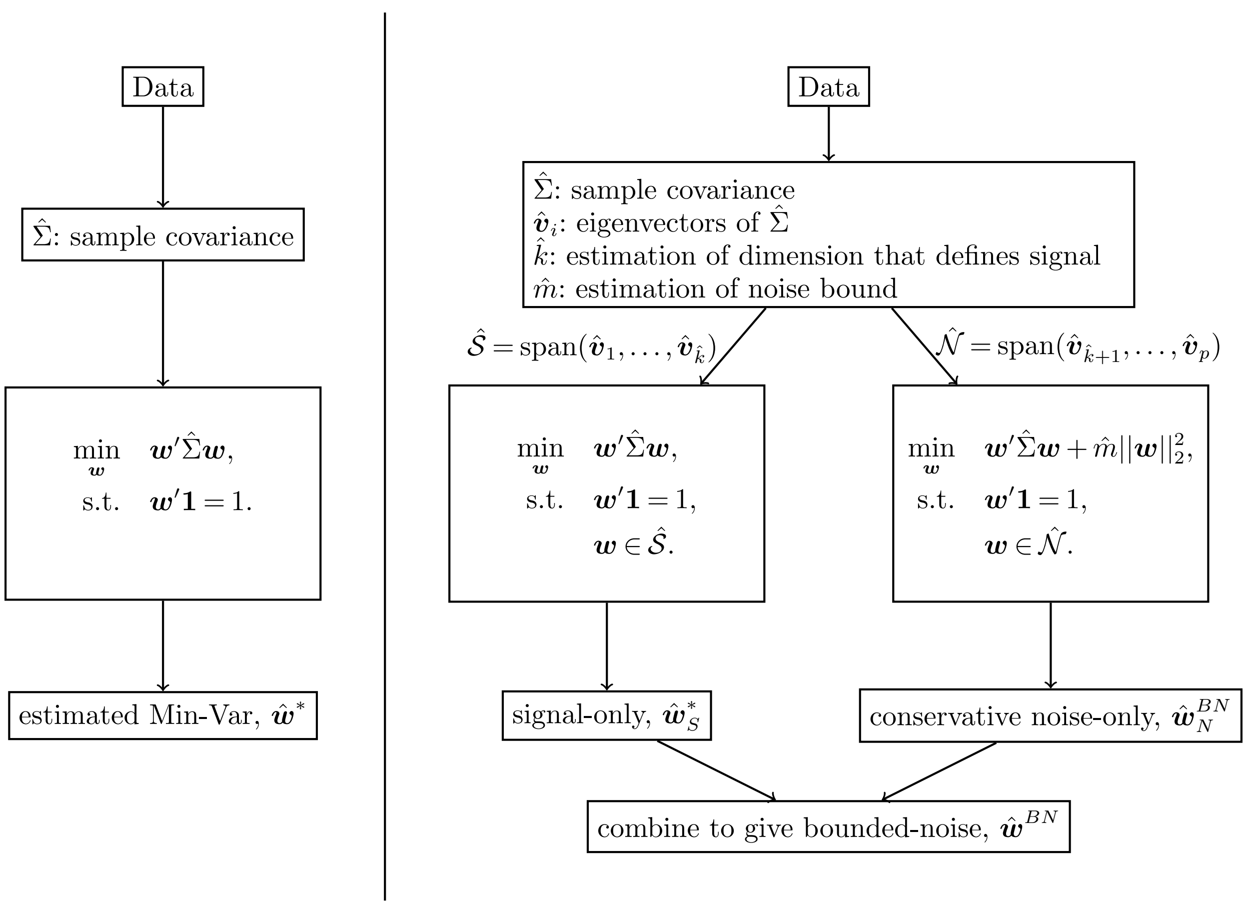


Figure 3 Diagram of the Estimated Min-Var Portfolio Compared to the Bounded-Noise Portfolio

4. Bounded-Noise Portfolios for Mean-Variance Optimization

Our entire discussion up to this point and much of the related literature(Jagannathan and Ma 2003, DeMiguel et al. 2009a, Brodie et al. 2009, Fan et al. 2012) focus on the minimum-variance 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 realized Sharpe ratio is that stretching for higher estimated expected returns often requires aggressive positions which can cause unexpected increases in the realized standard deviation ($RSD(\boldsymbol{w}) = \sqrt{RV(\boldsymbol{w})}$) 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 standard deviation, leading to a Sharpe ratio lower than even the estimated MinVar 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:

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

Here $c \geq 1$ is a constant and $\sigma^2_{\min}$ is the optimal objective value of the BN optimization problem (Eq. 9). Clearly, if $c = 1$, we recover the BN portfolio. If $c > 1$, Eq. 10 yields a portfolio (the BNvar portfolio) that maximizes the estimated expected return by tolerating the chance of a higher realized variance than the BN portfolio has. Crucially, the inequality in Eq. 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 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 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 third year, we calculate the Sharpe ratio of the previous three years' monthly returns for various $c \in [1, 1.5]$. Then we use the c that gives the highest Sharpe ratio to get the BNvar portfolio weights to hold for the next period.

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

5. Connections to Existing Portfolio Optimization Methods

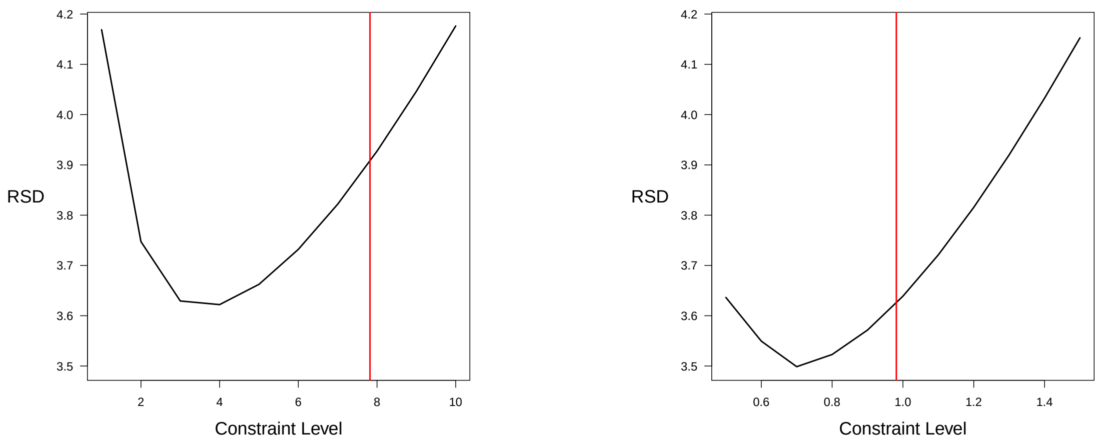
Empirical studies have shown that the norm-constrained portfolios work very well in practice (DeMiguel et al. 2009a). The preferred reasoning for its good performance is that the norm penalties on portfolio weights prevent large weights, which are often the result of estimation errors. Apart from the obvious issue of disallowing large portfolio weights even when the true MinVar portfolio might have them (Green and Hollifield 1992), this approach only fixes one particular effect of estimation error. Moreover, the choice of the norm and the magnitude of the penalty term are unclear. Because the BN portfolio deals with the underlying estimation problems directly, we now seek to understand its connections with the norm-constrained portfolio.

We first show that the norm-constrained portfolios impose the "wrong" constraints. Coupled with the idea of signal and noise split, we show that the norm-constrained portfolios avoid error amplification indirectly. We also discuss the relationship between the BN portfolio and the equal-weighted portfolio. Finally, we provide an interpretation that allows the BN portfolio to be considered as an innovative way to combine the estimated MinVar portfolio and the equal-weighted portfolio. This understanding offers a connection with those methods that alleviate the estimation and performance issue by combining with the equal-weighted portfolio, as discussed in section 1.3.

5.1. Imposing the Wrong Constraints to Combat Estimation Error

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

Figure 4 shows how the realized standard deviation varies with different constraint levels δ, for the $\mathbb{L}_1$ and $\mathbb{L}_2$ constrained portfolios under the simulations using the Fama-French value weighted dataset with 96 assets. In both cases, as expected, the realized standard deviation is too high at the extremes, because the constraints become either too strict or too weak. However, the optimum realized standard deviation is achieved for a constraint level at which the true 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 showed that the optimal portfolio could have large weights. Thus, the norm-constrained methods can achieve a low realized standard deviation only by imposing the wrong constraints, and they cannot be justified simply as a means of capping the effects of estimation error.



(a) L_1-norm, with the red line showing $\delta = ||\boldsymbol{w}^*||_1$ (b) L_2-norm, with the red line showing $\delta = ||\boldsymbol{w}^*||_2$

Figure 4 Realized standard deviation (RSD) with respect to different norm-constraint levels

5.2. Norm-Contrained Portfolios Avoid Error Amplification Indirectly

Recall that the BN portfolio is the solution to the minimum-variance problem using a modified covariance matrix, where we add the matrix $M = m\hat{N}\hat{N}'$ to the estimated covariance matrix $\hat{\Sigma}$ (Eq. 9). If all eigenvectors are noise, we have $M = mI$, which is precisely the *norm-constrained* portfolio optimization, using an $\mathbb{L}_2$-norm penalty and with the regularization parameter set to the noise bound, m. Thus, $\mathbb{L}_2$ norm-constrained portfolio is a special case of our solution. To get further insights, let's use simulation and project the norm-constrained portfolio into signal and noise space. Given the estimated signal vs. noise split, $\hat{k}$, we can compute the signal-portfolio and the noise-portfolio corresponding to any portfolio as shown in the following lemma.

Lemma 5.1 (Projection Portfolios) *Denote the eigenvectors of $\hat{\Sigma}$ by $\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_p$. Let $\hat{\mathcal{S}} = span(\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_{\hat{k}})$, and $\hat{\mathcal{N}} = span(\hat{\boldsymbol{v}}_{\hat{k}+1}, \ldots, \hat{\boldsymbol{v}}_p)$. Also introduce matrix $\hat{S} = (\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_{\hat{k}})$, and matrix $\hat{N} = (\hat{\boldsymbol{v}}_{\hat{k}+1}, \ldots, \hat{\boldsymbol{v}}_p)$. 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

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

This lemma informs that any portfolio corresponds not only to a particular choice for the signal and noise portfolios but also to a specific mixing proportion θ by which they are combined and where the error amplification happens. Thus, to investigate the actions against error amplification, we can compare one portfolio against another portfolio that combines these same signal and noise portfolio, but with the mixing proportion set using their estimated variances, as in the estimated MinVar portfolio. Specifically, if we choose the $\mathbb{L}_1$ ($\mathbb{L}_2$) norm-constrained portfolio with cross-validated[7] regularization parameter, denoted as $\mathbb{L}_1 CV$ ($\mathbb{L}_2 CV$), as the original portfolio, we can define the corresponding portfolio $\mathbb{L}_1$Mix ($\mathbb{L}_2$Mix) as

$$\boldsymbol{w}^{\mathbb{L}_1\text{Mix}} = \hat{\theta}^{L1CV}\boldsymbol{w}_S^{L1CV} + \left(1-\hat{\theta}^{L1CV}\right)\boldsymbol{w}_N^{L1CV} \qquad \hat{\theta}^{L1CV} = \frac{1/EV(\boldsymbol{w}_S^{L1CV})}{1/EV(\boldsymbol{w}_S^{L1CV})+1/EV(\boldsymbol{w}_N^{L1CV})} \tag{13}$$

$$\boldsymbol{w}^{\mathbb{L}_2\text{Mix}} = \hat{\theta}^{L2CV}\boldsymbol{w}_S^{L2CV} + \left(1-\hat{\theta}^{L2CV}\right)\boldsymbol{w}_N^{L2CV} \qquad \hat{\theta}^{L2CV} = \frac{1/EV(\boldsymbol{w}_S^{L2CV})}{1/EV(\boldsymbol{w}_S^{L2CV})+1/EV(\boldsymbol{w}_N^{L2CV})}. \tag{14}$$

We also define LiftWRTopt as the percentage improvement of the BN portfolio over an alternative portfolio using the true MinVar portfolio, $\boldsymbol{w}^*$, as the baseline:

$$\text{LiftWRTopt}(\boldsymbol{w}) = \frac{Avg.\ RSD(\boldsymbol{w}) - Avg.\ RSD(\hat{\boldsymbol{w}}^{\text{BN}})}{Avg.\ RSD(\boldsymbol{w}) - RSD(\boldsymbol{w}^*)}.$$

Table 1 compares $\mathbb{L}_1$CV and $\mathbb{L}_2$CV with $\mathbb{L}_1$Mix and $\mathbb{L}_2$Mix, as well as with the BN portfolio and the true MinVar portfolio, Opt, using the true covariance matrix. Also, we consider the signal-only portfolio, identified as SignalOnly, which ignores the noise space. We see that $\mathbb{L}_1$Mix ($\mathbb{L}_2$Mix) is much worse than $\mathbb{L}_1$CV ($\mathbb{L}_2$CV), showing that the norm-constrained portfolios avoid error amplification indirectly. Indeed, both $\mathbb{L}_1$Mix and $\mathbb{L}_2$Mix are worse than the signal-only portfolio, indicating that using the correct mixing proportion between signal and noise is essential. The inferred mixing proportion θ (from Lemma 5.1) is, on average, 1.65 times larger for $\mathbb{L}_1$CV compared to the mixing proportion $\hat{\theta}^{L1}$ of $\mathbb{L}_1$Mix (Eq. 13). The corresponding ratio is 2.09 for $\mathbb{L}_2$CV versus $\mathbb{L}_2$Mix. Thus, both the $\mathbb{L}_1$CV portfolio and the $\mathbb{L}_2$CV portfolio reduce the importance of the noise portfolios when combining it with their signal portfolios, thus avoiding the problem of overweighting the noise portfolios.

Table 1 Average RSD of Portfolios

	Opt	BN	$\mathbb{L}_1$CV	$\mathbb{L}_1$Mix	$\mathbb{L}_2$CV	$\mathbb{L}_2$Mix	SignalOnly
Avg. RSD	3.014	3.488	3.700	4.215	3.531	3.979	3.696
LiftWRTopt	NA	0	30.89%	60.50%	8.26%	50.82%	30.391%

We also see that the BN portfolio outperforms all norm-constrained portfolios. The dominance of BN is shown in Table 1 by an improvement of 8.26% over $\mathbb{L}_2$CV and larger improvements over all the others.

[7] Following DeMiguel et al. (2009a), we use the leave-one-out cross-validation approach through this paper. 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.

5.3. Connection to the Equal-Weighted Portfolio

The estimated MinVar portfolio fails because it takes the eigenvectors and eigenvalues from the noise space at face value. The BN portfolio rectified this problem by picking the conservative noise-only portfolio, $\hat{\boldsymbol{w}}_N^{BN}$, which minimized an upper bound of the realized variance (Proposition 3.1). 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:

$$\begin{aligned} \min_{\boldsymbol{w}} \max_{\Psi \in \mathcal{U}} \quad & \boldsymbol{w}'\Psi\boldsymbol{w}, \\ \text{subject to} \quad & \boldsymbol{w}'\mathbf{1} = 1 \\ & \boldsymbol{w} \in \hat{\mathcal{N}}, \end{aligned} \tag{15}$$

where $\mathcal{U}$ is the uncertainty set of all possible covariance matrices Ψ that have the same signal eigenvectors and eigenvalues as $\hat{\Sigma}$. Since Eq. 15 considers only $\boldsymbol{w} \in \hat{\mathcal{N}}$, we can use the following uncertainty set:

$$\mathcal{U} = \{\Psi \mid \hat{N}'\Psi\hat{N} \preceq bI_{n-\hat{k}+1}\}, \tag{16}$$

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 equal-weighted portfolio that invests $1/p$ in each of the p available stocks. This strategy is the right one in the extreme case where *no* sample is available. However, given sample, applying this idea just to the noise space is reasonable. Indeed, the projection of the equal-weighted portfolio on the noise space yields precisely the portfolio of Eq. 15, as we show next.

Lemma 5.2 *The solution to the robust optimization problem Eq. 15 with uncertainty set defined in Eq. 16 is the projection portfolio of the equal-weighted portfolio on $\hat{\mathcal{N}}$.*

Lemma 5.2 provides an alternative to the conservative noise-only portfolio, but we still need a way to combine it with the signal-only portfolio. To avoid overweighting the noise portfolio, we would have to use the bound on its realized variance in computing the mixing proportion between the signal and noise portfolios. Thus, we need to use Proposition 3.1 anyway. Therefore, we prefer using the conservative noise-only portfolio (Eq. 6), which also provides the bound, instead of the projection of the equal-weighted portfolio (Lemma 5.2) as the portfolio constructed from the noise space.

Though we have a preference of the conservative noise-only portfolio, $\hat{\boldsymbol{w}}_N^{BN}$, over the projection of the equal-weighted portfolio on the noise portfolio, $\boldsymbol{w}_N^{PEW}$, those two portfolios might be almost identical in some cases. For example, in simulation, the average inner product between their weight vectors is 0.991. This happens when $\hat{m} \gg \hat{\lambda}_i, \ \forall i \geq \hat{k}+1$, because

$$\boldsymbol{w}_N^{PEW} = \frac{\sum_{i=\hat{k}+1}^{p} (\hat{\boldsymbol{v}}_i' \mathbf{1}) \hat{\boldsymbol{v}}_i}{\sum_{i=\hat{k}+1}^{p} (\hat{\boldsymbol{v}}_i' \mathbf{1})^2}, \quad \hat{\boldsymbol{w}}_N^{BN} = \frac{\sum_{i=\hat{k}+1}^{p} \frac{\hat{\boldsymbol{v}}_i' \mathbf{1}}{\hat{\lambda}_i + \hat{m}} \hat{\boldsymbol{v}}_i}{\sum_{i=\hat{k}+1}^{p} \frac{(\hat{\boldsymbol{v}}_i' \mathbf{1})^2}{\hat{\lambda}_i + \hat{m}}}. \tag{17}$$

5.4. A New Way to Combine the Estimated MinVar and the Equal-Weighted Portfolio

Because the signal-only portfolio is the projection of the estimated MinVar portfolio on the signal space and the conservative noise-only portfolio is similar to the projection of the equal-weighted portfolio on the noise space, the BN portfolio provides an innovative way of combining the estimated MinVar portfolio and the equal-weighted portfolio. It takes the better one in both the signal and noise spaces separately and combines them efficiently. Thus, the BN portfolio has a close connection with the second category mentioned in the literature review (section 1.3) which tries to combine the estimated MinVar portfolio and the equal-weighted portfolio. Let's again use simulation to illustrate this insight.

Table 2 RSD of Projection Portfolios

	$\hat{\boldsymbol{w}}_S^*$	$\boldsymbol{w}_S^{PEW}$	$\hat{\boldsymbol{w}}_N^*$	$\hat{\boldsymbol{w}}_N^{\text{BN}}$	$\boldsymbol{w}_N^{PEW}$
Avg. RSD	3.696	5.168	7.687	4.917	4.948

Using Lemma 5.1, we compare the RSD of the projection portfolios of three portfolios via simulations: the BN portfolio, $\hat{\boldsymbol{w}}^{\text{BN}}$; the equal-weighted portfolio $\boldsymbol{w}^{EW}$; and the estimated MinVar portfolio, $\hat{\boldsymbol{w}}^*$. The results are shown in Table 2. Note that the signal projection portfolio of the estimated BN portfolio, $\hat{\boldsymbol{w}}^{\text{BN}}$, and the estimated MinVar portfolio, $\hat{\boldsymbol{w}}^*$, are the same. It is the signal-only portfolio, $\hat{\boldsymbol{w}}_S^*$. This portfolio dominates the projection of the equal-weighted portfolio on the signal space, $\boldsymbol{w}_S^{PEW}$. The opposite is true for the aggressive noise-only portfolio, $\hat{\boldsymbol{w}}_N^*$, and the equal-weighted portfolio's projection portfolio in the noise space, $\hat{\boldsymbol{w}}_N^{PEW}$. We also see that the RSD of the conservative noise-only portfolio, $\hat{\boldsymbol{w}}_N^{BN}$, is very close to the RSD of $\hat{\boldsymbol{w}}_N^{PEW}$, which supports the previous argument on their similarity. This simulation indicates that the BN portfolio approximately picks the better projection portfolio of the estimated MinVar portfolio and the equal-weighted portfolio on the signal and noise space, separately.

Table 3 List of Portfolios Considered in Empirical Experiments

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

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

Table 4 List of Datasets Considered

Dataset	*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

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

6. 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 existing literature (Table 3) across twelve different datasets (Table 4). All datasets except individual stocks dataset come from K.French's website[8]. The individual stocks datasets come from CRSP. The time period for all datasets is 07/1963 to 07/2015 which shares the same starting point as DeMiguel et al. (2009a). When sorting the market value of firms, we only include the stocks whose returns are available for the past ten years and the future one year. For one hundred Fama and French (1992) dataset, because there are missing values for four risky assets for an extended period, we deleted them, leaving 96 of the original 100 portfolios. The BN portfolio uses one parameter γ, which we set to 0.25 for all datasets and provide its sensitivity analysis in section 6.5. We use 1000 bootstrap samples in the estimation procedure.

Competing methods. We consider two naive portfolios, the equal-weighted (EW) and the value-weighted (VW) portfolio, as our benchmarks. Every asset of the EW portfolio is given equal weight when it is rebalanced. The VW portfolio, on the other hand, assigns the fraction of the market

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

capitalization to each asset as its portfolio weight. DeMiguel et al. (2009b) provide a thorough analysis for both portfolios. The EstMinVar portfolio, which is defined at the beginning of section 1.3, is the classical minimum-variance portfolio formulated in Markowitz (1952).

In addition to these standard benchmarks, we consider three others that add additional constraints or penalties to the minimum-variance portfolio optimization problem. The first one is the shortsale-constrained portfolio (Jagannathan and Ma 2003, section 1), which has a non-negativity constraint on the portfolio weights. We call it the NoShorting portfolio. The remaining two are norm-constrained portfolios with parameters set via cross-validation over standard deviation. These portfolios are detailed in DeMiguel et al. (2009a, section 3.1 and 3.2). The L_1-norm constrained portfolio is labeled as $\mathbb{L}_1$CV, and the L_2-norm constrained is labeled as $\mathbb{L}_2$CV.

Finally, we also include two relatively recent and well-performing benchmarks. The partial minimum-variance 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. (2009a, 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 out-of-sample Sharpe ratio. The turnover discussion can be seen in section 6.3. Following the convention of Brodie et al. (2009), DeMiguel et al. (2009a), 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 dataset. As in DeMiguel et al. (2009a), 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 strategy. We denote the weight vector as $\boldsymbol{w}_t^i$ for $t = n, \ldots, T-1$ and for each strategy i.

Following DeMiguel et al. (2009a), we hold the portfolio weight $\boldsymbol{w}_t^i$ for one 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:

$$
\begin{aligned}
(\hat{\sigma}^i)^2 &= \frac{1}{T-n-1}\sum_{t=n}^{T-1}\left((\boldsymbol{w}_t^i)'\boldsymbol{r}_{t+1} - \hat{\mu}^i\right)^2, \\
\text{where} \quad \hat{\mu}^i &= \frac{1}{T-n}\sum_{t=n}^{T-1}(\boldsymbol{w}_t^i)'\boldsymbol{r}_{t+1}, \\
\widehat{SR}^i &= \frac{\hat{\mu}^i}{\hat{\sigma}^i}.
\end{aligned}
$$

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

6.1. Comparison of Out-of-Sample Standard Deviation

Table 5 Out-of-Sample Standard Deviation in Percentage

Strategy	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
$\mathbb{L}_1$CV	**4.415**	4.058	3.720	3.680	3.758	3.790	3.754	3.605	3.902	3.757	3.602	3.487
$\mathbb{L}_2$CV	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 number in **bold** is the smallest standard deviation for one dataset. The p-value is calculated between the **BN** portfolio and other portfolios.
<u>One underline</u>, <u>two underlines</u>, and <u>three underlines</u> indicate that the related p-value is smaller than 0.1, 0.05, and 0.01, respectively.
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 MinVar portfolio portfolio.

Table 5 shows that the BN portfolio achieves the best out-of-sample standard deviation on five out of the six large[9] portfolio datasets and is second-best on the 48IndEW dataset. For all datasets, the BN portfolio is always significantly[10] better than the EW portfolio. Note that the BNVAR portfolio has a higher out-of-sample standard deviation than the BN portfolio precisely because it is expected to maximize the Sharpe ratio, and not to minimize the standard deviation. There is a challenge in creating and comparing stock portfolios, due to market issues like mergers, acquisitions, delistings, IPOs, 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. 2009a). We use this procedure annually and update our list by choosing the largest 100 or 500 stocks, as measured by their market value[11]. Updating the stock list selection annually facilitates our turnover investigations as well (section 6.3). The results for these stock portfolios have to be interpreted with caution since it can be argued that these are aggregates over not perfectly

[9] We use the phrase large datasets when the number of assets, p, is larger than ten.

[10] p-value is less than 0.05.

[11] The number of asset changes for each update is 2.5 and 50 on average for the 100 and 500 stock dataset, respectively

comparable stock datasets. The BNVAR portfolio doesn't exist for these two stocks datasets because of the changing universe of stocks.

On the small datasets, the out-of-sample standard deviation of the ESTMINVAR portfolio is only about 1% larger than the best portfolio. This relationship indicates 120 observations are enough for the small datasets to have the whole eigenspace as the signal space. Thus, the BN portfolio shouldn't differ much from the ESTMINVAR portfolio, and indeed the correlation between their returns turns out to be more than 0.99. For the same reason, we expect cross-validation to choose very loose norm constraints for all the norm-constrained methods. Thus, their corresponding portfolios should be essentially the same as ESTMINVAR. This result is again supported by the extremely 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. But it does better on some big datasets, where its constraint helps to avoid the effects of covariance estimation errors.

6.2. Discussion of Out-of-Sample Sharpe Ratio

Table 6 Out-of-Sample Sharpe Ratio

Strategy	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
$\mathbb{L}_1$CV	0.395	0.329	0.290	0.278	0.427	0.345	0.272	0.244	0.399	0.364	0.242	0.259
$\mathbb{L}_2$CV	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 dataset. 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.
One underline, two underlines, and three underlines indicate that the related p-value is smaller than 0.1, 0.05, and 0.01, respectively.
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 MinVar portfolio portfolio.

Table 6 shows that except for dataset 48IndVW, the portfolio that has the highest Sharpe ratio is not the portfolio that has the lowest standard deviation. The BNVAR portfolio has the best out-of-sample Sharpe ratio for six of ten portfolio datasets, and the dominance on these six datasets is both statistically and economically significant. These results show that we are indeed able to increase the out-of-sample Sharpe ratio for most datasets by allowing a higher variance level.

The upper bound on the out-of-sample variance was critical in achieving this result; without it, a small change in the estimated variance can translate into substantial increases in the out-of-sample variance resulting in a drastic decrease of the out-of-sample Sharpe ratio. The PARR portfolio achieves the best performance for four datasets which is consistent with DeMiguel et al. (2009a). However, the result here doesn't consider the transaction costs and taxation which are crucial when the turnover is high. Thus, in the next subsection, we will address the issue of turnover.

6.3. Robustness of Holding Length: a Discussion of Turnover

The comparison of results with transactions costs mostly serves the purpose of evaluating turnovers since portfolios with large turnovers get significantly penalized for high transaction costs. To get a sense of the sensitivities of portfolios' performance to turnover, we compare the performance of the earlier monthly-rebalanced portfolios with the annually-rebalanced portfolios (Brodie et al. 2009) which we construct in this subsection. This would allow us to evaluate the effects of turnover without making the results sensitive to the type and 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 one 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 reduce the turnover sharply without sacrificing much in performance.

Table 7 Hold for 1 year, Out-of-Sample Standard Deviation in Percentage

Strategy	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
$\mathbb{L}_1$CV	4.860	4.607	3.746	3.642	4.034	3.935	4.372	3.682	4.126	4.006	3.789	3.357
$\mathbb{L}_2$CV	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 dataset. The p-value is calculated between the **BN** portfolio and other portfolios.

One underline, two underlines, and three underlines indicate that the related p-value is smaller than 0.1, 0.05, and 0.01, respectively.

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 MinVar portfolio portfolio.

Compared to Table 5 and 6, Table 7 and 8 show that the performance of the low turnover portfolios (EW, VW, and NOSHORTING) remains similar. The previous winners in terms of Sharpe ratio, namely the BNVAR portfolio and the PARR portfolio, see a huge decrease in Sharpe ratio and are no longer the best. This happens because both have high turnovers. The BNvar portfolio, the NoShorting portfolio, and the $\mathbb{L}_1$CV portfolio are the only ones that have a larger Sharpe ratio than the equal-weighted portfolio across all the portfolio datasets. Right now, there is no clear winner in terms of Sharpe ratio.

Table 8 Hold for 1 year, Out-of-Sample Sharpe Ratio

Strategy	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
$\mathbb{L}_1$CV	0.362	**0.302**	0.260	0.285	0.402	0.345	0.245	0.258	0.395	0.356	0.249	0.271
$\mathbb{L}_2$CV	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 dataset. 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.
One underline, two underlines, and three underlines indicate that the related p-value is smaller than 0.1, 0.05, and 0.01, respectively.
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 MinVar portfolio portfolio.

6.4. Robustness of Training Length

In this subsection, following Brodie et al. (2009), we show the results using the earlier datasets 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[12]. Even when $n > p$, the estimated covariance might be close to singular (i.e., its smallest eigenvalue could be nearly zero). Especially since 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 problems for datasets 96FFEW, 96FFVW, 100, and 500 necessarily are singular.

From Table 9, the BN portfolio is the best on eight out of ten portfolio datasets, including five (of six) large portfolio datasets and the second-best for the sixth. Considering Table 5 and Table 9 together, we find that the out-of-sample standard deviation of the BN portfolio is quite 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 eight (of ten) portfolio datasets (Table 10). The natural

[12] 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.

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

Strategy	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
$\mathbb{L}_1$CV	4.399	4.121	3.800	3.723	3.912	3.942	3.900	4.031	4.286	4.418	3.928	3.462
$\mathbb{L}_2$CV	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 dataset. The p-value is calculated between the **BN** portfolio and other portfolios.

One underline, two underlines, and three underlines indicate that the related p-value is smaller than 0.1, 0.05, and 0.01, respectively.

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 MinVar portfolio portfolio.

reason for such robust performance is that the BN portfolio and the BNVAR portfolio become more cautious when training length becomes smaller. It turns out that the signal space becomes smaller and the noise bound, m, becomes larger as fewer observations are available. This happens because, for any chosen γ, the signal to noise split dictated by definition 1 makes the signal space smaller leaving us with a larger noise space and a larger noise bound m.

The out-of-sample standard deviations of the $\mathbb{L}_1$CV portfolio and $\mathbb{L}_2$CV portfolio (Table 9) increase significantly compared to Table 5. This change increases the margin between the standard deviations of the BN portfolio and other portfolios. For example, on the dataset 96FFVW, the standard deviation of the BN portfolio is 6% better than that of the $\mathbb{L}_2$CV portfolio and 11% better than that of the $\mathbb{L}_1$CV portfolio. The intuitive reason is that, unlike the bound-noise procedure, cross-validation is unable to force a more conservative portfolio when there are fewer data. In fact, in about 36% of the time periods, the penalty parameter (Eq. 1) with 60 observations η_{60} is smaller than η_{120}.

6.5. Robustness of Model Parameters

There are two model parameters in the BN portfolio: the number of bootstraps, L, and the cutoff of amplification ratio, γ. We find that $L = 100$ generates almost identical result as $L = 1000$ whose result is reported in the previous subsections.

Table 11 reports the sensitivity analysis of γ with $\gamma = 0.25$ as benchmark case. For the BN portfolio, the differences between $\gamma = 0.15$ or $\gamma = 0.4$ with the benchmark case are around 1%. For the BNVAR portfolio, the differences are slightly larger, but most of them are smaller than 2%.

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

Strategy	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
$\mathbb{L}_1$CV	0.422	0.327	0.314	0.264	0.402	0.310	0.299	0.236	0.295	0.295	0.185	0.263
$\mathbb{L}_2$CV	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 dataset. 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.
One underline, two underlines, and three underlines indicate that the related p-value is smaller than 0.1, 0.05, and 0.01, respectively.
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 MinVar portfolio portfolio.

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

Strategy	6FFEW	6FFVW	10IndEW	10IndVW	25FFEW	25FFVW	48IndEW	48IndVW	96FFEW	96FFVW	100	500
Out-of-Sample 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 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 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 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

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.

7. 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 fixing the optimization step in an arbitrary manner that happens to reduce the impact of estimation errors. Instead of either of these, we disentangle the covariance matrix into two parts. The part that behaves well in the optimization step, we call the signal part and the part that does not, we call the noise part. We detailed and discussed 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, connections to other methods, extension that allow constructing mean-variance portfolios and finally, evidence of superior performance using both the simulated and the real-world data.

Regarding next steps, there are several aspects that could benefit from further investigation. The signal/noise split and the related optimal portfolios rely heavily on the investor having no additional constraint. Extending the splitting idea in the context of optimal portfolios with additional constraints is very valuable, though challenging. Pushing the ideas in the paper, more along the mean-variance direction would be another good direction for future work. The method we have described in this paper for constructing mean-variance portfolio 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, though 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 upper-bound parameter M via a median of bootstrap samples. A more careful analysis could use the full distribution of M derived from these samples.

References

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

Brodie J, Daubechies I, De Mol C, Giannone D, Loris I (2009) Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences* 106(30):12267–12272.

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

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

Chopra VK, Ziemba WT (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of 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, Garlappi L, Nogales FJ, Uppal R (2009a) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55(5):798–812.

DeMiguel V, Garlappi L, Uppal R (2009b) Optimal versus naïve diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* 22(5):1915–1953.

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

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

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

Duchin R, Levy H (2009) Markowitz versus the Talmudic portfolio diversification strategies. *Journal of Portfolio Management* 35(2):71.

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

Fama EF, French KR (2015) A five-factor asset pricing model. *Journal of Financial Economics* 116(1):1 – 22, ISSN 0304-405X.

Fan J, Zhang J, Yu K (2012) Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association* 107(498):592–606.

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

Frost PA, Savarino JE (1986) An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis* 21(03):293–305.

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

Gotoh Jy, Takeda A (2011) On the role of norm constraints in portfolio selection. *Computational Management Science* 8(4):323–353.

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

Hansen PR, Lunde A (2006) Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24(2):127–161.

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

Jobson JD, Korkie B (1980) Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association* 75(371):544–554.

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

Jorion P (1986) Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21(03):279–292.

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

Laloux L, Cizeau P, Potters M, Bouchaud JP (2000) Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance* 3(03):391–397.

Lauprête GJ (2001) *Portfolio risk minimization under departures from normality*. Ph.D. thesis, Massachusetts Institute of Technology.

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

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

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

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

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

Levene H (1960) Robust tests for equality of variances1. *Contributions to probability and statistics: Essays in honor of Harold Hotelling* 2:278–292.

Lim TS, Loh WY (1996) A comparison of tests of equality of variances. *Computational Statistics & Data Analysis* 22(3):287–301.

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

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

Michaud RO (1989) The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts Journal* 45(1):31–42.

Mihail M, Papadimitriou C (2002) On the eigenvalue power law. *International Workshop on Randomization and Approximation Techniques in Computer Science*, 254–262.

Olivares-Nadal AV, DeMiguel V (2018) Technical note: A robust perspective on transaction costs in portfolio optimization. *Operations Research* 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. *Physical Review E* 65(6):066126.

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

Vershynin R (2011) How Close is the Sample Covariance Matrix to the Actual Covariance Matrix? *Journal of Theoretical Probability* 25(3):655–686.

Welsch RE, Zhou X (2007) Application of robust statistics to asset allocation models. *REVSTAT–Statistical Journal* 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.

Appendix A: Absolute Versus Relative Differences in Eigenvalues

Lemma 2.2 showed that the bottom eigenvectors could not be well estimated because they are not well separated (i.e., the absolute differences between their eigenvalues are small). This observation might seem to suggest that the corresponding eigenvectors are almost interchangeable and that these errors have a limited effect on the performance of the aggressive noise-only portfolio. However, this intuition is false, because the optimal portfolio depends on the *relative* differences between eigenvalues, which can still be large. We demonstrate this understanding with some examples.

EXAMPLE 1. Suppose the eigenvalues vary as a (heavy-tailed) power-law, with $\lambda_i = \xi \cdot \beta^{p-i-1}$ for some $\xi > 0$ and $\beta > 1$. The absolute difference between consecutive eigenvalues is $\lambda_i - \lambda_{i+1} = \xi \cdot \beta^{p-i-1} \cdot (1 - 1/\beta)$, which decreases with increasing i and is at most ξ for the last two eigenvalues. Thus, with a large enough β and a small enough ξ, every consecutive pair of eigenvectors is well separated, except for the bottom two eigenvectors. Under these conditions, for some number of samples n, we can expect the top $p-2$ eigenvectors to be well estimated, and the last two to be poorly estimated. Let us also assume for simplicity that $\boldsymbol{v}_{p-1}'\mathbf{1} = \boldsymbol{v}_p'\mathbf{1} = \rho \neq 0$.

In general, the bottom eigenvalues are poorly estimated, as in Figure 1. However, let us consider the best-case scenario for estimation: Suppose that the top $p-2$ eigenvectors are estimated perfectly, as are all eigenvalues. Let us take $\hat{\mathcal{S}}$ to be the span of the first $p-2$ sample eigenvectors. Because these eigenvectors are perfectly estimated, we have $\hat{\mathcal{S}} = \mathcal{S}$. Let $\hat{\mathcal{N}}$ and $\mathcal{N}$ denote the spans of the last two sample eigenvectors and true eigenvectors, respectively. Note that $\hat{\mathcal{N}} = \mathcal{N}$ because each is simply the space orthogonal to $\hat{\mathcal{S}} = \mathcal{S}$. Thus, the only error is in the orientation of the bottom two eigenvectors $\hat{\boldsymbol{v}}_{p-1}$ and $\hat{\boldsymbol{v}}_p$.

In other words,

$$\hat{\boldsymbol{v}}_{p-1} = \boldsymbol{v}_{p-1} \cdot \cos\theta - \boldsymbol{v}_p \cdot \sin\theta \qquad \hat{\boldsymbol{v}}_p = \boldsymbol{v}_{p-1} \cdot \sin\theta + \boldsymbol{v}_p \cdot \cos\theta$$

for some random angle θ. Because we can always reverse the direction of these four eigenvectors without loss of generality, we confine θ to $[0, \pi]$. Then, we can show

$$RV(\boldsymbol{w}_N^*) = \frac{\xi}{\rho^2 \cdot (\beta + 1)} \tag{18}$$

$$RV(\hat{\boldsymbol{w}}_N^*) \approx EV(\hat{\boldsymbol{w}}_N^*) \cdot \left(1 + (\beta - 1) \cdot (\sin\theta)^2\right) \quad \text{for } \beta \gg 1 \tag{19}$$

$$RV(\hat{\boldsymbol{w}}_N^*) \approx RV(\boldsymbol{w}_N^*) \cdot \frac{\beta \cdot (\sin\theta)^2}{(\cos\theta + \sin\theta)^2} \quad \text{for } \beta \gg 1. \tag{20}$$

The approximations hold when $\theta \neq \frac{3\pi}{4}$, which is true with probability 1. Thus, the aggressive noise-only portfolio is considered to be a far better portfolio than it is actually is ($EV(\hat{\boldsymbol{w}}_N^*) \ll RV(\hat{\boldsymbol{w}}_N^*)$) and performs poorer than the optimal portfolio from the noise space ($RV(\hat{\boldsymbol{w}}_N^*) \gg RV(\boldsymbol{w}_N^*)$). □

Recall that the individual bottom eigenvectors might be poorly estimated, but the *span* of these eigenvectors is well estimated (i.e., space $\hat{\mathcal{N}}$ itself is well estimated). This reasoning is simply that $\hat{\mathcal{N}}$ is the space that is orthogonal to $\hat{\mathcal{S}}$, which is well estimated. Thus, we can expect good performance from a portfolio that depends only on space $\hat{\mathcal{N}}$ while being invariant to the precise configuration of the eigenvectors in $\hat{\mathcal{N}}$. The following example illustrates the case.

Example 2 (An extension of Example 1). From Eq. 17, we know that the projected equal-weighted portfolio on the noise space weights the bottom two eigenvectors as follows

$$\boldsymbol{w}_N^{PEW} = \frac{\sum_{j=p-1}^{p}(\hat{\boldsymbol{v}}_j'\mathbf{1})\hat{\boldsymbol{v}}_j}{\sum_{j=p-1}^{p}(\hat{\boldsymbol{v}}_j'\mathbf{1})^2}.$$

Note that this definition does not refer to eigenvalues at all. One property of this portfolio is that it is invariant to θ:

$$\begin{aligned}
\boldsymbol{w}_N^{PEW} &= \frac{\hat{\boldsymbol{v}}_{p-1}'\mathbf{1}}{(\hat{\boldsymbol{v}}_{p-1}'\mathbf{1})^2+(\hat{\boldsymbol{v}}_p'\mathbf{1})^2}\hat{\boldsymbol{v}}_{p-1} + \frac{\hat{\boldsymbol{v}}_p'\mathbf{1}}{(\hat{\boldsymbol{v}}_{p-1}'\mathbf{1})^2+(\hat{\boldsymbol{v}}_p'\mathbf{1})^2}\hat{\boldsymbol{v}}_p \\
&= \frac{\rho(\cos\theta-\sin\theta)}{\rho^2(\cos\theta-\sin\theta)^2+\rho^2(\cos\theta+\sin\theta)^2}(\boldsymbol{v}_{p-1}\cdot\cos\theta-\boldsymbol{v}_p\cdot\sin\theta) \\
&+ \frac{\rho(\cos\theta+\sin\theta)}{\rho^2(\cos\theta-\sin\theta)^2+\rho^2(\cos\theta+\sin\theta)^2}(\boldsymbol{v}_{p-1}\cdot\sin\theta+\boldsymbol{v}_p\cdot\cos\theta) \\
&= \frac{1}{2\rho}\left(\boldsymbol{v}_{p-1}\cdot(\cos^2\theta+\sin^2\theta)+\boldsymbol{v}_p\cdot(\sin^2\theta+\cos^2\theta)\right) \\
&= \frac{1}{2\rho}(\boldsymbol{v}_{p-1}+\boldsymbol{v}_p) \\
&= \frac{\boldsymbol{v}_{p-1}'\mathbf{1}}{(\boldsymbol{v}_{p-1}'\mathbf{1})^2+(\boldsymbol{v}_p'\mathbf{1})^2}\boldsymbol{v}_{p-1} + \frac{\boldsymbol{v}_p'\mathbf{1}}{(\boldsymbol{v}_{p-1}'\mathbf{1})^2+(\boldsymbol{v}_p'\mathbf{1})^2}\boldsymbol{v}_p
\end{aligned}$$

Using Eq. 18, we find that

$$RV(\boldsymbol{w}_N^{PEW}) = \frac{1}{4\rho^2}\left(\xi+\frac{\xi}{\beta}\right) \approx RV(\boldsymbol{w}_N^*)\cdot\frac{\beta}{4} \quad \text{for } \beta \gg 1.$$

Again, the realized variance of $\boldsymbol{w}_N^{PEW}$ is invariant to θ. We find that $\boldsymbol{w}_N^{PEW}$ is comparable to the aggressive noise-only portfolio (Eq. 20) in terms of realized variance, and that it can, in fact, be better than $\hat{\boldsymbol{w}}_N^*$ when $\frac{\pi}{4} \leq \theta \leq \frac{3\pi}{4}$. This finding makes sense because $\frac{\pi}{4} \leq \theta \leq \frac{3\pi}{4}$ means that $\hat{\boldsymbol{v}}_{p-1}$ is closer to $\boldsymbol{v}_p$ than $\boldsymbol{v}_{p-1}$. □

Appendix B: Proofs

Proposition 2.1(Eigenvalue Concentration)

Proof. By Weyl's inequality, $|\lambda_i - \hat{\lambda}_i| \leq \|\Sigma - \hat{\Sigma}_n\|_{op}$. Dividing both sides by λ_i proves the proposition. □

Lemma 5.1 (Portfolio Decomposition)

Proof. Using the Lagrangian multiplier method, we can easily find:

$$\boldsymbol{w}^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = \frac{\sum_i \frac{\boldsymbol{v}_i'\mathbf{1}}{\lambda_i}\boldsymbol{v}_i}{\sum_i \frac{(\boldsymbol{v}_i'\mathbf{1})^2}{\lambda_i}},$$

where we use $\Sigma^{-1} = \sum_i (1/\lambda_i)\boldsymbol{v}_i\boldsymbol{v}_i'$. Similarly, we have

$$\boldsymbol{w}_S^* = \frac{\sum_{j=1}^k \frac{\boldsymbol{v}_j'\mathbf{1}}{\lambda_j}\boldsymbol{v}_j}{\sum_{j=1}^k \frac{(\boldsymbol{v}_j'\mathbf{1})^2}{\lambda_j}}, \qquad RV(\boldsymbol{w}_S^*) = \frac{1}{\sum_{j=1}^k \frac{(\boldsymbol{v}_j'\mathbf{1})^2}{\lambda_j}}, \qquad \frac{1}{RV(\boldsymbol{w}_S^*)}\boldsymbol{w}_S^* = \sum_{j=1}^k \frac{\boldsymbol{v}_j'\mathbf{1}}{\lambda_j}. \tag{21}$$

Repeat this process for $\boldsymbol{w}_N^*$, and some algebraic manipulations yield Eq. 2. □

Proposition 3.1(Bounding Realized Variance of Any Portfolio From the Noise Space)

Proof. Because the noise space $\hat{\mathcal{N}}$ is spanned by $\hat{\boldsymbol{v}}_{k+1}, \ldots, \hat{\boldsymbol{v}}_p$, any vector $\boldsymbol{w}$ from $\hat{\mathcal{N}}$ can be presented as a linear combination of this basis, namely,

$$\boldsymbol{w} = (\hat{\boldsymbol{v}}_{k+1}, \ldots, \hat{\boldsymbol{v}}_p)\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \hat{N}\boldsymbol{a}.$$

From the orthonormality of eigenvectors, we have

$$||\boldsymbol{w}||_2^2 = \boldsymbol{w}'\boldsymbol{w} = \boldsymbol{a}'\hat{N}'\hat{N}\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{a} = ||\boldsymbol{a}||_2^2. \tag{22}$$

Meanwhile, the definition of the noise bound, m, guarantees that the following inequality holds for any vector $\boldsymbol{b} \in \mathbb{R}^n$ such that $||\boldsymbol{b}||_2 = 1$,

$$\boldsymbol{b}'\left(\hat{N}'(\Sigma - \hat{\Sigma})\hat{N}\right)\boldsymbol{b} \leq m. \tag{23}$$

Plugging $\boldsymbol{a}/||\boldsymbol{a}||_2$ into the previous inequality, we have

$$\left(\frac{\boldsymbol{a}}{||\boldsymbol{a}||_2}\right)'\left(\hat{N}'(\Sigma - \hat{\Sigma})\hat{N}\right)\left(\frac{\boldsymbol{a}}{||\boldsymbol{a}||_2}\right) \leq m.$$

Rearranging, we get

$$(\hat{N}\boldsymbol{a})'\Sigma(\hat{N}\boldsymbol{a}) \leq (\hat{N}\boldsymbol{a})'\hat{\Sigma}(\hat{N}\boldsymbol{a}) + m||\boldsymbol{a}||_2^2.$$

Substituting $\boldsymbol{w} = \hat{N}\boldsymbol{a}$ and $||\boldsymbol{w}||_2^2 = ||\boldsymbol{a}||_2^2$ into the preceding inequality proves the proposition. □

Lemma 5.1 (Projection Portfolios)

Proof. Because $\boldsymbol{w} \in \hat{\mathcal{N}}$, we have $\boldsymbol{w} = \hat{N}\boldsymbol{a}$. Thus,

$$\max_{\Psi \in \mathcal{U}} \boldsymbol{w}'\Psi\boldsymbol{w} = \max_{\Psi \in \mathcal{U}} \boldsymbol{a}'\hat{N}'\Psi\hat{N}\boldsymbol{a} = b\boldsymbol{a}'I_{n-k+1}\boldsymbol{a}.$$

The last equality holds because of the definition of the uncertainty set. Then Eq. 15 becomes

$$\begin{aligned} \min_{\boldsymbol{a}} \quad & b\boldsymbol{a}'I_{n-k+1}\boldsymbol{a}, \\ \text{subject to} \quad & \boldsymbol{a}'(\hat{N}'\mathbf{1}) = 1. \end{aligned}$$

Its solution is

$$\boldsymbol{a}^* = \frac{\hat{N}'\mathbf{1}}{\mathbf{1}'\hat{N}\hat{N}'\mathbf{1}},$$

which implies that the solution to the robust optimization is

$$\hat{N}\boldsymbol{a}^* = \frac{\hat{N}\hat{N}'\mathbf{1}}{\mathbf{1}'\hat{N}\hat{N}'\mathbf{1}}.$$

From Eq. 12, then, the projection portfolio of the equal-weighted portfolio on $\hat{\mathcal{N}}$ is

$$\boldsymbol{w}_N^{PEW} = \frac{\hat{N}\hat{N}'(\mathbf{1}/p)}{(\mathbf{1}/p)'\hat{N}\hat{N}'\mathbf{1}} = \frac{\hat{N}\hat{N}'\mathbf{1}}{\mathbf{1}'\hat{N}\hat{N}'\mathbf{1}} = \hat{N}\boldsymbol{a}^*. \quad \square$$