A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms

Victor DeMiguel  Lorenzo Garlappi  Francisco J. Nogales  Raman Uppal

September 24, 2008

Abstract

We provide a general framework for finding portfolios that perform well out-of-sample in the presence of estimation error. This framework relies on solving the traditional minimum-variance problem but subject to the additional constraint that the norm of the portfolio-weight vector be smaller than a given threshold. We show that our framework nests as special cases the shrinkage approaches of Jagannathan and Ma (2003) and Ledoit and Wolf (2003, 2004), and the 1/N portfolio studied in DeMiguel, Garlappi, and Uppal (2007). We also use our framework to propose several new portfolio strategies. For the proposed portfolios, we provide a moment-shrinkage interpretation and a Bayesian interpretation where the investor has a prior belief on portfolio weights rather than on moments of asset returns. Finally, we compare empirically the out-of-sample performance of the new portfolios we propose to ten strategies in the literature across five datasets. We find that the norm-constrained portfolios often have a higher Sharpe ratio than the portfolio strategies in Jagannathan and Ma (2003), Ledoit and Wolf (2003, 2004), the 1/N portfolio, and other strategies in the literature, such as factor portfolios.

*We are very grateful to the Department Editor, David Hsieh, the Associate Editor, and two anonymous referees for detailed and extensive comments. We wish to thank Mike Chernov, Vito Gala, Francisco Gomes, Burton Hollifield, Garud Iyengar, Tongshu Ma, Igor Makarov, Spencer Martin, Javier Peña, Roberto Rigobon, Chester Spatt, Catalina Stefanescu, and Bruce Weber for comments. We are also grateful for suggestions to seminar participants at Carnegie Mellon University, INSEAD, 2008 meetings of the American Finance Association, 2007 INFORMS Annual Meeting, 2007 ICCOPT-MOPTA Conference, and Follow-Up Workshop on Optimization in Finance in Coimbra. DeMiguel and Uppal are from London Business School; Garlappi is from The University of Texas at Austin; Nogales is from the Universidad Carlos III de Madrid. F. J. Nogales is partially supported by the Ministry of Education and Science and Technology of Spain through CICYT Project MTM2007-63140.
1 Introduction

Markowitz (1952) showed that an investor who cares only about the mean and variance of static portfolio returns should hold a portfolio on the efficient frontier. To implement these portfolios in practice, one needs to estimate the means and covariances of asset returns. Traditionally, the sample means and covariances have been used for this purpose. But due to estimation error, the portfolios that rely on the sample estimates typically perform poorly out of sample.\footnote{For evidence of the poor performance of the Markowitz portfolio based on sample estimates of means and covariances, see Frost and Savarino (1986, 1988), Michaud (1989), Best and Grauer (1991), Chopra and Ziemba (1993), Broadie (1993), and Litterman (2003).} In this paper, we provide a general framework for determining portfolios with superior out-of-sample performance in the presence of estimation error. This general framework relies on solving the traditional minimum-variance problem (based on the sample covariance matrix) but subject to the additional constraint that the norm of the portfolio-weight vector be smaller than a given threshold.

It is well known that it is more difficult to estimate means than covariances of asset returns (see Merton (1980)), and also that errors in estimates of means have a larger impact on portfolio weights than errors in estimates of covariances. For this reason, recent academic research has focused on minimum-variance portfolios, which rely solely on estimates of covariances, and thus, are less vulnerable to estimation error than mean-variance portfolios.\footnote{Note that although expected returns cannot be forecasted reasonably well from historical data, if a portfolio manager has the skill to forecast expected returns, then one may wish to use a mean-variance portfolio rather than a minimum-variance portfolio.} Indeed, extensive empirical evidence shows that the minimum-variance portfolio usually performs better out of sample than any other mean-variance portfolio—even when using a performance measure that depends on both the portfolio mean and variance. For example, Jagannathan and Ma (2003) report:\footnote{For additional evidence, see Jorion (1985, 1986, 1991) and DeMiguel, Garlappi, and Uppal (2007).}

The estimation error in the sample mean is so large nothing much is lost in ignoring the mean altogether when no further information about the population mean is available. For example, the global minimum variance portfolio has as large an out-of-sample Sharpe ratio as other efficient portfolios when past historical average returns are used as proxies for expected returns. In view of this we focus our attention on global minimum variance portfolios in this study (pp. 1652–3).
Just like Jagannathan and Ma (2003), we too focus on minimum-variance portfolios, even though the general framework we develop applies also to mean-variance portfolios. But even the performance of the minimum-variance portfolio depends crucially on the quality of the estimated covariances, and although the estimation error associated with the sample covariances is smaller than that for sample mean returns, it can still be substantial.

In the literature, several approaches have been proposed to deal with the problem of estimating the large number of elements in the covariance matrix. One approach is to use higher-frequency data, say daily instead of monthly returns (see Jagannathan and Ma (2003)). A second approach is to impose some factor structure on the estimator of the covariance matrix (Chan, Karceski, and Lakonishok (1999), Green and Hollifield (1992)), which reduces the number of parameters to be estimated. A third approach has been proposed by Ledoit and Wolf (2003, 2004), who use as an estimator a weighted average of the sample covariance matrix and another estimator, such as the 1-factor covariance matrix or the identity matrix. A fourth approach, which is often used in practice, is to impose shortsale constraints on the portfolio weights (see Frost and Savarino (1988) and Chopra (1993)). Jagannathan and Ma (2003) show that imposing a shortsale constraint when minimizing the portfolio variance is equivalent to shrinking the extreme elements of the covariance matrix. This simple remedy for dealing with estimation error performs very well. In fact, Jagannathan and Ma (2003) find that “the sample covariance matrix [with shortsale constraints] performs almost as well as those [covariance matrices] constructed using factor models, shrinkage estimators or daily returns” (p. 1654). Finally, DeMiguel, Garlappi, and Uppal (2007) demonstrate that even constraining shortsales may not mitigate completely the error in estimating the covariance matrix, and thus, an investor may be better off (in terms of Sharpe ratio, certainty-equivalent returns, and turnover) ignoring data on asset returns altogether and using the naive $1/N$ rule to allocate an equal proportion of wealth across each of the $N$ assets.

In this paper, we develop a new approach for determining the optimal portfolio weights in the presence of estimation error. Following Brandt (1999) and Britten-Jones (1999), we treat the weights rather than the moments of assets returns as the objects of interest to be estimated. So, rather than shrinking the moments of asset returns, we introduce the constraint that the norm of the portfolio-weight vector be smaller than a given threshold.
Our paper contributes to the literature on optimal portfolio choice in the presence of estimation error in several ways. One, we show that our framework nests as special cases the shrinkage approaches of Jagannathan and Ma (2003) and Ledoit and Wolf (2003, 2004). In particular, we prove that if one solves the minimum-variance problem subject to the constraint that the sum of the absolute values of the weights (1-norm) be smaller than or equal to 1, then one retrieves the shortsale-constrained minimum-variance portfolio considered by Jagannathan and Ma (2003). If one constrains the sum of the squares of the portfolio weights (2-norm) to be smaller than a given threshold, then one recovers the class of portfolios considered by Ledoit and Wolf (2004). Similarly, imposing a particular quadratic constraint allows one to recover the portfolios in Ledoit and Wolf (2003). Finally, constraining the squared 2-norm of the portfolio-weight vector to be smaller than or equal to $1/N$ gives the $1/N$ portfolio studied in DeMiguel, Garlappi, and Uppal (2007).

Two, we use this unifying framework to develop new portfolio strategies. For example, we show how the shortsale-constrained portfolio considered in Jagannathan and Ma (2003) can be generalized. In particular, we show that by imposing the constraint that the 1-norm of the portfolio-weight vector be smaller than a threshold that is strictly larger than 1, one obtains a new class of shrinkage portfolios that limit the total amount of shortselling in the portfolio, rather than limiting the shorting of each asset, as in the traditional shortsale-constrained portfolio. To the best of our knowledge, this kind of portfolio has not been analyzed before in the academic literature, although it corresponds closely to the actual portfolio holdings allowed in personal margin accounts.\footnote{These portfolios are quite popular among practitioners—see the articles in The Economist (Buttonwood (2007)) and The New York Times (Hershey Jr. (2007)) that describe “130–30” portfolios where investors are long 130% and short 30% of their wealth.} We also propose a class of portfolios that we term “partial minimum-variance portfolios.” These portfolios are obtained by applying the classical conjugate-gradient method (see Nocedal and Wright (1999)) to solve the minimum-variance problem. We show that these portfolios are related to the portfolios obtained by imposing a constraint on the 2-norm of the portfolio-weight vector.

Three, we give a Bayesian interpretation for the norm-constrained portfolios and also for the portfolios proposed by Jagannathan and Ma (2003) and Ledoit and Wolf (2003, 2004) that is in terms of a certain prior belief on portfolio weights rather than on moments of asset returns.
Four, our approach to minimum-variance portfolio selection is related also to a number of approaches proposed in the statistics and chemometrics literature to reduce estimation error in regression analysis. It is known in the literature that optimal portfolio weights in an unconstrained mean- or minimum-variance problem can be thought of as coefficients of an OLS regression (see, for example, Britten-Jones (1999)). It then follows that, in general, constrained weights are the outcome of similarly specified restricted regressions. In particular, constraining the 1-norm of the portfolio vector to be less than a certain threshold is analogous to the statistical technique for regression analysis known as “least absolute shrinkage and selection operator” (lasso) (Tibshirani (1996)); constraining the 2-norm of the portfolio vector to be less than a certain threshold corresponds to the statistical technique known as “ridge regression” (Hoerl and Kennard (1970)); and the “partial minimum-variance portfolio” corresponds to the technique developed in chemometrics known as “partial least squares” (Wold (1975); Frank and Friedman (1993)). These regression techniques and the distribution theory associated with them have been used extensively in the statistics literature.

Five, the generalized framework allows one to calibrate the model using historical data in order to improve its out-of-sample performance. We compare empirically the out-of-sample performance of the norm-constrained portfolios to ten strategies in the literature for five different datasets. The datasets we consider are listed in Table 1 and the portfolios we evaluate are listed in Table 2. In terms of out-of-sample variance, the norm-constrained portfolios often have a lower variance than the shortsale-constrained minimum-variance portfolio studied in Jagannathan and Ma (2003), the $1/N$ portfolio evaluated in DeMiguel, Garlappi, and Uppal (2007), and also other strategies proposed in the literature, including factor portfolios and the parametric portfolios in Brandt, Santa-Clara, and Valkanov (2005); however, the variance of the norm-constrained portfolios is similar to that of the portfolios in Ledoit and Wolf (2003, 2004). In terms of out-of-sample Sharpe ratio, the portfolios we propose attain a Sharpe ratio that is higher than the shortsale constrained minimum-variance portfolio, the $1/N$ portfolio, and the portfolios in Ledoit and Wolf (2003, 2004), although the higher Sharpe ratio is accompanied by higher turnover. Finally, the Sharpe ratio and
turnover of the proposed portfolios is similar to that of Brandt, Santa-Clara, and Valkanov (2005), but without the need to use firm-specific characteristics.\(^5\)

The remainder of this paper is organized as follows. Section 2 reviews the approaches considered in Jagannathan and Ma (2003) and Ledoit and Wolf (2003, 2004), which shrink some or all of the elements of the sample covariance matrix. In Section 3, we propose our general approach, which shrinks the portfolio-weight vector directly. In Section 4, we discuss the performance of the different portfolios on empirical data. Section 5 concludes. Our main results are highlighted in propositions, and proofs for all the propositions are collected in Appendix A. Details of how to compute the partial minimum-variance portfolios are provided in Appendix B in the online supplement.

### 2 Existing Approaches: Shrinking the Sample Covariance Matrix

In the absence of shortsale constraints, the minimum-variance portfolio is the solution to the following optimization problem:

\[
\begin{align*}
\min_{\mathbf{w}} & \quad \mathbf{w}^\top \hat{\Sigma} \mathbf{w}, \\
\text{s.t.} & \quad \mathbf{w}^\top \mathbf{e} = 1,
\end{align*}
\]

in which \(\mathbf{w} \in \mathbb{R}^N\) is the vector of portfolio weights, \(\hat{\Sigma} \in \mathbb{R}^{N \times N}\) is the estimated covariance matrix, \(\mathbf{w}^\top \hat{\Sigma} \mathbf{w}\) is the variance of the portfolio return, \(\mathbf{e} \in \mathbb{R}^N\) is the vector of ones, and the constraint \(\mathbf{w}^\top \mathbf{e} = 1\) ensures that the portfolio weights sum up to one. We denote the solution to this shortsale-unconstrained minimum-variance problem by \(\mathbf{w}_{MNU}\).

Jagannathan and Ma (2003) study the shortsale-constrained minimum-variance portfolio, \(\mathbf{w}_{MNC}\), which is the solution to problem (1)-(2) with the additional constraint that the portfolio weights be non-negative (\(\mathbf{w} \geq 0\)). They show that the solution to the shortsale-constrained problem coincides with the solution to the \textit{unconstrained} problem in (1)-(2) if the sample covariance matrix in (1) is

\(^5\)Because the parametric portfolios in Brandt, Santa-Clara, and Valkanov (2005) rely on firm-specific characteristics, they are not really comparable to the other portfolios we evaluate; however, we decided to include them in our empirical analysis because these portfolios achieve very high Sharpe ratios, and hence, are an important benchmark. Lauprete (2001) also considers the 1- and 2-norm-constrained portfolios; see also Lauprete, Samarov, and Welsch (2002) and Welsch and Zhou (2007). Our additional contribution is to relate these methods to the approaches in Jagannathan and Ma (2003) and Ledoit and Wolf (2003, 2004), propose the partial minimum-variance portfolios, and provide comprehensive empirical results.
replaced by the matrix

$$\hat{\Sigma}_{JM} = \hat{\Sigma} - \lambda e^\top - e\lambda^\top,$$

in which $\lambda \in \mathbb{R}^N$ is the vector of Lagrange multipliers for the shortsale constraint. Because $\lambda \geq 0$, the matrix $\hat{\Sigma}_{JM}$ may be interpreted as the sample covariance matrix after shrinkage, because if the shortsale constraint corresponding to the $i$th asset is binding ($w_i = 0$), then the sample covariance of this asset with any other asset is reduced by $\lambda_i$, the magnitude of the Lagrange multiplier associated with its shortsale constraint.

Ledoit and Wolf (2003, 2004) propose replacing the sample covariance matrix with a weighted average of the sample covariance matrix and a low-variance target estimator, $\hat{\Sigma}_{\text{target}}$. Concretely, they propose solving problem (1)–(2), in which the matrix $\hat{\Sigma}$ is replaced by

$$\hat{\Sigma}_{LW} = \frac{1}{1+\nu} \hat{\Sigma} + \frac{\nu}{1+\nu} \hat{\Sigma}_{\text{target}},$$

and in which $\nu$ is a positive constant. Ledoit and Wolf (2003, 2004) also show how one can estimate the value of $\nu$ that minimizes the expected Frobenius norm of the difference between the matrix $\hat{\Sigma}_{LW}$ and the true covariance matrix. They show that this method can be interpreted as shrinking the sample covariance matrix toward the estimator $\hat{\Sigma}_{\text{target}}$. They consider several candidates for $\hat{\Sigma}_{\text{target}}$, such as the identity matrix and the covariance matrix obtained from estimating a 1-factor model with the market as the factor.\(^6\)

3 A Generalized Approach: Constraining the Portfolio Norms

In this section, we propose a class of portfolios that can be viewed as resulting from shrinking the portfolio-weight vector instead of shrinking the moments of asset returns. Specifically, we define the norm-constrained minimum-variance portfolio as the one that solves the traditional minimum-variance problem (1)–(2) subject to the additional constraint that the norm of the portfolio-weight

\(^6\) Note that Ledoit and Wolf (2004) actually consider a positive multiple of the identity matrix as their shrinkage target. Specifically, they consider the identity matrix multiplied by the average of the diagonal elements of the sample covariance matrix. This makes sense in the context of their work because their objective is to find the estimator that minimizes the Frobenius norm of the difference with the true covariance matrix. In the context of our manuscript, where the objective is to compute minimum-variance portfolios, it does not matter whether one uses the identity matrix as the shrinkage target, or a positive multiple of the identity matrix. The reason for this is that given a shrinkage target $\hat{\Sigma}_{\text{target}} = \alpha I$, with $\alpha > 0$, and a shrinkage weight $\nu_1 > 0$, the resulting minimum-variance portfolio coincides with that obtained using the identity as a shrinkage target and a shrinkage weight $\nu_2 = \alpha \nu_1 > 0$. 

7
vector is smaller than a certain threshold \( \delta \); that is, \( \| w \| \leq \delta \), in which \( \| w \| \) is the norm of the portfolio-weight vector. In particular, we consider the 1-norm, \( \| w \|_1 = \sum_{i=1}^{N} |w_i| \), and the \( A \)-norm, \( \| w \|_A = (w^\top A w)^{1/2} \), in which \( A \in \mathbb{R}^{N \times N} \) is a positive-definite matrix.

Note that the traditional shortsale-unconstrained minimum-variance portfolio, \( w_{MINU} \), is the solution to the norm-constrained problem with \( \delta = \infty \). Consequently, if \( \delta < \| w_{MINU} \| \), then the norm of the portfolio that solves the norm-constrained minimum-variance portfolio problem must be strictly smaller than that of the unconstrained minimum-variance portfolio, \( w_{MINU} \).\(^7\) Hence, imposing a constraint on the norm of the portfolio-weight vector can be interpreted as shrinking the shortsale-unconstrained sample-based minimum-variance portfolio toward a target portfolio. The target portfolio is the one that minimizes the norm of the weight vector subject to the condition that the weights add up to one.\(^8\) Shrinkage estimators have been a popular method for reducing estimation error ever since their introduction by James and Stein (1961). The idea behind shrinkage estimators is that shrinking an unbiased estimator toward a lower-variance target has a negative and a positive effect. The negative effect is that the shrinkage introduces bias into the resulting estimator, while the positive effect is that it reduces the variance of the estimator.

3.1 First Special Case: The 1-Norm-Constrained Portfolios

The 1-norm-constrained portfolio, \( w_{NC1} \), is the solution to the traditional minimum-variance portfolio problem (1)–(2) subject to the additional constraint that the 1-norm of the portfolio-weight vector be smaller than or equal to a certain threshold \( \delta \); that is, \( \| w \|_1 = \sum_{i=1}^{N} |w_i| \leq \delta \).

The following proposition shows that for the case \( \delta = 1 \), the solution to the 1-norm-constrained minimum-variance problem is the same as for the shortsale-constrained minimum-variance portfolio analyzed in Jagannathan and Ma (2003).

**Proposition 1** The solution to the 1-norm-constrained minimum-variance portfolio problem with \( \delta = 1 \) coincides with the solution to the shortsale-constrained problem.

\(^7\)Note that the norm-constrained portfolios are not obtained by shrinking every weight of the minimum-variance portfolio; instead, the norm-constrained portfolios are obtained by shrinking the total norm of the minimum-variance portfolio. This interpretation of shrinkage is the same as in Tibshirani (1996) in the context of lasso regression.

\(^8\)Let \( w_M \) be the portfolio that minimizes the norm. To see that this is the target portfolio in our approach, note that if we impose the constraint that the portfolio norm is smaller than \( \| w_M \| \), then the only portfolio that is feasible with respect to this constraint is precisely \( w_M \).
In contrast to the case where $\delta = 1$, for threshold values of $\delta > 1$ this approach generates a class of portfolios that generalize the shortsale-constrained minimum-variance portfolio. Specifically, some algebraic manipulation can be used to show that the 1-norm constraint $\|w\|_1 \leq \delta$ can be rewritten as

$$- \sum_{i \in \mathcal{N}(w)} w_i \leq \delta - \frac{1}{2},$$

in which $\mathcal{N}(w)$ is the set of asset indexes for which the corresponding portfolio weight is negative, $\mathcal{N}(w) = \{i : w_i < 0\}$. The left-hand side of (5) is the total proportion of wealth that is sold short and $(\delta - 1)/2$ can be interpreted as the shortsale budget. This shortsale budget can then be freely distributed among all of the assets.

### 3.2 Second Special Case: The $A$-Norm-Constrained Portfolios

The $A$-norm-constrained minimum-variance portfolio is the solution to the traditional minimum-variance problem (1)–(2) subject to the additional constraint that the $A$-norm of the portfolio-weight vector is smaller than a particular threshold, $\hat{\delta}$. Because the squared $A$-norm is easier to analyze than the $A$-norm, we instead impose the equivalent constraint $w^\top A w \leq \delta$, in which $\delta = \hat{\delta}^2$.

The following proposition shows the relation between the $A$-norm-constrained portfolios and the shrinkage portfolios proposed by Ledoit and Wolf.

**Proposition 2** Provided $\hat{\Sigma}$ is nonsingular, for each $\nu \geq 0$ there exists a $\delta$ such that the solution to the minimum-variance problem in (1)–(2) with the sample covariance matrix, $\hat{\Sigma}$, replaced by $\hat{\Sigma}_{LW} = \frac{1}{1+\nu} \hat{\Sigma} + \frac{\nu}{1+\nu} A$ coincides with the solution to the $A$-norm-constrained minimum-variance portfolio problem, which is the traditional minimum-variance problem subject to the additional constraint:

$$w^\top A w \leq \delta.$$  

In particular, if we choose the matrix $A$ equal to the identity matrix, $I$, then there is a one-to-one correspondence between the $A$-norm constrained portfolios and the shrinkage portfolio proposed in Ledoit and Wolf (2004). On the other hand, if $A$ equals the 1-factor covariance matrix, $\hat{\Sigma}_F$, then there is a one-to-one correspondence with the shrinkage portfolio in Ledoit and Wolf (2003).
Note that for the special case where $A = I$, the $A$-norm is simply the 2-norm, $\|w\|_2$, and therefore, we refer to these as the 2-norm-constrained minimum-variance portfolios. To gain intuition about these portfolios, note that the 2-norm constraint $\sum_{i=1}^N w_i^2 \leq \delta$ can be reformulated equivalently as follows:

$$\sum_{i=1}^N \left(w_i - \frac{1}{N}\right)^2 \leq \left(\delta - \frac{1}{N}\right).$$

The reformulated constraint (7) demonstrates that imposing the 2-norm constraint on the portfolio weights is equivalent to imposing a constraint that the 2-norm of the difference between this portfolio and the $1/N$ portfolio is bounded by $\delta - 1/N$. Note also that the $1/N$ portfolio is a special case of the 2-norm-constrained portfolio with $\delta = 1/N$.

For the special case where $A = \hat{\Sigma}_F$, imposing a constraint on the $\hat{\Sigma}_F$-norm of the portfolio is equivalent to imposing a constraint on the portfolio variance under the covariance estimator obtained from a 1-factor model; that is, imposing the constraint that $w^\top \hat{\Sigma}_F w$ is smaller than a certain threshold.

We now compare the 1-norm, 2-norm, and $\hat{\Sigma}_F$-norm constrained portfolios. Above, we have shown that the 1-norm constrained portfolios are a generalization of the shortsale-constrained portfolios in which the total amount of shortselling on all assets must remain below a shortsale budget of $(\delta - 1)/2$. Therefore, we may expect the 1-norm-constrained portfolios to have the well-known property of shortsale-constrained minimum-variance portfolios, which tend to assign a weight different from zero to only a few of the assets. In contrast, we have shown that the 2-norm-constrained portfolio is the portfolio that minimizes the sample variance subject to the constraint that the square of the 2-norm of the difference with the $1/N$ portfolio is bounded by $\delta - 1/N$. Consequently, we would expect that the 2-norm constrained portfolios will, in general, remain relatively close to the $1/N$ portfolio, and thus, will assign a positive weight to all assets. Similarly, we should expect the $\hat{\Sigma}_F$-norm constrained portfolios to remain close to the portfolio that minimizes the 1-factor model variance. Thus, investors who believe the total amount of shortselling

---

9To understand why these two formulations are equivalent, observe that $\sum_{i=1}^N (w_i - 1/N)^2 = \sum_{i=1}^N w_i^2 + \sum_{i=1}^N 1/N^2 - \sum_{i=1}^N 2w_i/N = \sum_{i=1}^N w_i^2 - 1/N$, where the last result follows from the fact that $\sum_{i=1}^N 1/N^2 = 1/N$ and $\sum_{i=1}^N 2w_i/N = 2/N$.

10The reason for this is that the 1-norm function has a kink at portfolios where one or more of the weights are zero, while this is not the case for the 2-norm function.
of the optimal portfolio should not exceed a given budget would want to use a 1-norm constraint. On the other hand, investors who believe that the optimal portfolio is close to the well-diversified $1/N$ portfolio would want to use a 2-norm constraint when solving the minimum-variance problem to reduce estimation error. Finally, investors who believe a 1-factor structure holds in the asset-return distribution would use a $\hat{\Sigma}_F$-norm constraint.

### 3.3 Third Special Case: The Partial Minimum-Variance Portfolios

The third class of portfolios we propose are obtained by applying the classical conjugate-gradient method (Nocedal and Wright (1999, Chapter 5)) to solve the minimum-variance problem. The conjugate gradient method takes as a starting portfolio some initial guess (in our implementation we use the $1/N$ portfolio) and then generates a sequence of $N - 1$ portfolios, where the terminal portfolio is the shortsale-unconstrained minimum-variance portfolio. We term each of these $N - 1$ intermediate portfolios a partial minimum-variance portfolio. In Appendix B in the online supplement, we provide a detailed description of how to compute these portfolios.

Even though the partial minimum-variance portfolios are not obtained by explicitly imposing a constraint on the norm of the minimum-variance portfolio, the following proposition shows that the 2-norm of the partial minimum-variance portfolios is smaller than the 2-norm of the shortsale-unconstrained minimum-variance portfolios.

**Proposition 3** The 2-norm of the $k^{\text{th}}$ partial minimum-variance portfolio is smaller than or equal to the 2-norm of the shortsale-unconstrained minimum-variance portfolio for $k \leq N - 1$.

Moreover, Proposition B3 in Appendix B shows that the partial minimum-variance portfolios can be viewed as a discrete first-order approximation to the 2-norm-constrained portfolios.

### 3.4 A Bayesian Interpretation of the Norm-Constrained Portfolios

Tibshirani (1996, Section 5) gives a Bayesian interpretation for the regression-analysis techniques of “least absolute shrinkage and selection operator” (lasso) and “ridge” regressions; see also Hastie, Tibshirani, and Friedman (2001, Chapter 3) and Park and Casella (2005). Here we adapt his
argument to give a Bayesian interpretation of the 1- and \( A \)-norm-constrained minimum-variance portfolios.

The following proposition shows that the 1-norm-constrained portfolio is the mode of the posterior distribution of portfolio weights for an investor whose prior belief is that the portfolio weights are independently and identically distributed as a Double Exponential distribution.

**Proposition 4** Assume that asset returns are normally distributed. Moreover, assume that the investor believes a priori that each of the shortsale-unconstrained minimum-variance portfolio weights, \( w_i \), is independently and identically distributed as a Double Exponential distribution with probability density function:

\[
\pi(w_i) = \frac{\nu}{2} e^{-\nu |w_i|}.
\]  

Furthermore, assume that the investor believes a priori that the variance of the minimum-variance portfolio return, denoted by \( \sigma^2 \), has an independent prior distribution \( \pi(\sigma^2) \). Then there exists a threshold parameter \( \delta \) such that the weights of the 1-norm-constrained minimum-variance portfolio are the mode of the posterior distribution of the minimum-variance portfolio weights.

The next proposition shows that the \( A \)-norm-constrained portfolio is the mode of the posterior distribution of portfolio weights for an investor whose prior belief is that the portfolio weights \( w_i \) have a multivariate Normal distribution with covariance matrix \( A \).

**Proposition 5** Assume that asset returns are normally distributed. Moreover, assume that the investor believes a priori that the shortsale-unconstrained minimum-variance portfolio weights are distributed as a multivariate Normal distribution with probability density function:

\[
\pi(w) = (2\pi)^{-n/2} |A|^{1/2} e^{-\frac{1}{2}w^T A w}.
\]

Furthermore, assume that the investor believes a priori that the variance of the minimum-variance portfolio return, denoted by \( \sigma^2 \), has an independent prior distribution \( \pi(\sigma^2) \). Then there exists a threshold parameter \( \delta \) such that the weights of the \( A \)-norm-constrained minimum-variance portfolio are the mode of the posterior distribution of the minimum-variance portfolio weights.
Note that choosing the portfolio that maximizes the posterior distribution of the minimum-variance portfolio weights guarantees that the investor is choosing the portfolio with the highest probability of being the minimum-variance portfolio given the investor’s prior distribution on portfolio weights and the observed asset-return data. This interpretation is different from that in the traditional Bayesian portfolio choice literature (for instance, Jorion (1986)) because in our framework the investor has a prior belief on the portfolio weights rather than on the asset-return distribution. Consequently, while the Bayesian investor in the traditional setting chooses the portfolio that maximizes expected utility with respect to the posterior distribution of asset returns, in our setting the investor chooses the portfolio that maximizes the posterior distribution of portfolio weights.

3.5 A Moment-Shrinkage Interpretation of the Norm-Constrained Portfolios

The 1-norm and $A$-norm-constrained minimum-variance portfolios can also be interpreted as portfolios that result from shrinking some of the elements of the sample covariance matrix.

**Proposition 6** Let the solution to the 1-norm-constrained minimum-variance problem be such that $(w_{NC1})_i \neq 0$ for $i = 1, \ldots, N$. Then $w_{NC1}$ is also the solution to the shortsale-unconstrained minimum-variance problem (1)–(2) if the sample covariance matrix, $\hat{\Sigma}$, is replaced by the matrix

$$\hat{\Sigma}_{NC1} = \hat{\Sigma} - \nu ne^\top - \nu en^\top,$$

(10)

in which $\nu \in \mathbb{R}$ is the Lagrange multiplier for the 1-norm constraint at the solution to the 1-norm-constrained minimum-variance problem, and $n \in \mathbb{R}^N$ is a vector whose $i^{th}$ component is one if the weight assigned by the 1-norm-constrained portfolio to the $i^{th}$ asset is negative and zero otherwise.

Proposition 6 shows that the 1-norm-constrained portfolios can also be interpreted as those obtained by shrinking some of the elements of the sample covariance matrix. Concretely, equation (10) shows that the 1-norm-constrained portfolios can be seen as the result of shrinking by the constant amount $\nu$ the covariances of those assets that are being sold short with all the other assets. Note that the amount of shrinkage $\nu$ is the same for all assets that are being sold short. This is the main difference between the 1-norm-constrained and the shortsale-constrained portfolios. From
Table 1: List of Datasets Considered

This table lists the various datasets of monthly asset returns analyzed, the abbreviation used to refer to each dataset, the number of risky assets $N$ in each dataset, the time period spanned by the dataset, and the source of the data. The dataset of CRSP returns (500CRSP) is constructed in a way that is similar to Jagannathan and Ma (2003), with monthly rebalancing: in April of each year we randomly select 500 assets among all assets in the CRSP dataset for which there is return data for the previous 120 months as well as for the next 12 months. We then consider these randomly selected 500 assets as our asset universe for the next 12 months.

<table>
<thead>
<tr>
<th>#</th>
<th>Dataset</th>
<th>Abbreviation</th>
<th>$N$</th>
<th>Time Period</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ten industry portfolios representing the U.S. stock market</td>
<td>10Ind</td>
<td>10</td>
<td>07/1963–12/2004</td>
<td>K. French</td>
</tr>
<tr>
<td>2</td>
<td>Forty-eight industry portfolios representing the U.S. stock market</td>
<td>48Ind</td>
<td>48</td>
<td>07/1963–12/2004</td>
<td>K. French</td>
</tr>
<tr>
<td>3</td>
<td>Six Fama and French portfolios of firms sorted by size and book to market</td>
<td>6FF</td>
<td>6</td>
<td>07/1963–12/2004</td>
<td>K. French</td>
</tr>
<tr>
<td>5</td>
<td>500 randomized stocks from CRSP balanced monthly</td>
<td>500CRSP</td>
<td>500</td>
<td>04/1968–04/2005</td>
<td>CRSP</td>
</tr>
</tbody>
</table>

Equation (3) it can be observed that for the shortsale-constrained portfolios the amount of shrinkage applied to the covariances of each of the assets is equal to the Lagrange multiplier corresponding to its shortsale constraint $\lambda_i$, and these Lagrange multipliers may take different values for different assets.

Similarly, observe from Proposition 2 that the $A$-norm-constrained portfolios can be obtained also by shrinking all elements of the sample covariance matrix toward the elements of matrix $A$.

4 Out-of-Sample Evaluation of the Proposed Portfolios

In this section, we compare across five different datasets (listed in Table 1) the out-of-sample empirical performance of the norm-constrained portfolios to ten portfolios from the existing literature using three performance metrics: the out-of-sample portfolio variance, the out-of-sample Sharpe ratio, and turnover.
Table 2: List of Portfolios Considered

This table lists the various portfolio strategies we consider. Panel A of the table lists the norm-constrained portfolios developed in this paper, while Panel B lists the portfolios from the literature. Note that $\delta$ is the threshold parameter that limits the norm of the portfolio-weight vector, while the order parameter $k$ indicates which of the $N-1$ partial minimum-variance portfolios to use. The last column gives the abbreviation that we use to refer to the strategy.

<table>
<thead>
<tr>
<th>#</th>
<th>Model</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Panel A: Portfolio strategies developed in this paper</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-norm-constrained minimum-variance portfolio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>With $\delta$ calibrated using cross-validation over portfolio variance</td>
<td>NC1V</td>
</tr>
<tr>
<td></td>
<td>With $\delta$ calibrated by maximizing portfolio return in previous period</td>
<td>NC1R</td>
</tr>
<tr>
<td></td>
<td>2-norm-constrained minimum-variance portfolio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>With $\delta$ calibrated using cross-validation over portfolio variance</td>
<td>NC2V</td>
</tr>
<tr>
<td></td>
<td>With $\delta$ calibrated by maximizing portfolio return in previous period</td>
<td>NC2R</td>
</tr>
<tr>
<td></td>
<td>$\hat{\Sigma}_F$-norm-constrained minimum-variance portfolio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>With $\delta$ calibrated using cross-validation over portfolio variance</td>
<td>NCFV</td>
</tr>
<tr>
<td></td>
<td>With $\delta$ calibrated by maximizing portfolio return in previous period</td>
<td>NCFR</td>
</tr>
<tr>
<td></td>
<td>Partial minimum-variance portfolios</td>
<td></td>
</tr>
<tr>
<td></td>
<td>With $k$ calibrated using cross-validation over portfolio variance</td>
<td>PARV</td>
</tr>
<tr>
<td></td>
<td>With $k$ calibrated by maximizing portfolio return in previous period</td>
<td>PARR</td>
</tr>
<tr>
<td></td>
<td><strong>Panel B: Portfolio strategies from the existing literature used for comparison</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple benchmarks</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Equally-weighted (1/$N$) portfolio</td>
<td>1/N</td>
</tr>
<tr>
<td>2</td>
<td>Value-weighted (market) portfolio</td>
<td>VW</td>
</tr>
<tr>
<td></td>
<td>Portfolios that use mean returns with shortsales unconstrained</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Mean-variance portfolio with risk aversion parameter $\gamma = 5$</td>
<td>MEAN</td>
</tr>
<tr>
<td>4</td>
<td>Bayesian mean-variance portfolio as in Jorion (1985, 1986) with risk aversion parameter $\gamma = 5$</td>
<td>BAYE</td>
</tr>
<tr>
<td></td>
<td>Minimum-variance portfolios</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Minimum-variance portfolio with shortsales unconstrained</td>
<td>MINU</td>
</tr>
<tr>
<td>6</td>
<td>Minimum-variance portfolio with shortsales constrained (Jagannathan and Ma, 2003)</td>
<td>MINC</td>
</tr>
<tr>
<td>7</td>
<td>Minimum-variance portfolio with the market as the single factor</td>
<td>FAC1</td>
</tr>
<tr>
<td></td>
<td>Minimum-variance portfolios where covariance matrix is average of two estimators</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Weighted average of sample covariance and identity matrix (Ledoit and Wolf, 2004)</td>
<td>LWID</td>
</tr>
<tr>
<td>9</td>
<td>Weighted average of sample covariance and 1-factor matrix (Ledoit and Wolf, 2003)</td>
<td>LW1F</td>
</tr>
<tr>
<td></td>
<td>Parametric portfolios</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Parametric portfolio as in Brandt, Santa-Clara, and Valkanov (2005) with a risk-aversion parameter of $\gamma = 5$ using the factors Size, Book-to-Market, and Momentum</td>
<td>BSV3</td>
</tr>
</tbody>
</table>

4.1 Description of the Portfolios Evaluated

The norm-constrained minimum-variance portfolios and the partial minimum-variance portfolio that we consider are listed in Panel A of Table 2. Note that to use the 1-, 2-, and $\hat{\Sigma}_F$-norm-
constrained minimum-variance portfolios, one needs to choose the value of the threshold parameter \( \delta \), which bounds the maximum value that the portfolio norm may take. Similarly, for the partial minimum-variance portfolios, one needs to choose the order parameter \( k \) that indicates which of the \( N - 1 \) partial minimum-variance portfolios to use. The parameters \( \delta \) and \( k \) could be specified exogenously. But in our framework these can also be calibrated to achieve a particular objective and/or to exploit a particular feature of the returns data. We use two different criteria to calibrate the norm-constrained portfolios: (i) minimizing the portfolio variance, and (ii) maximizing the last period portfolio return in order to exploit positive autocorrelation in portfolio returns, as opposed to autocorrelation in the return of individual securities.\(^{11}\)

If the objective is to minimize the “out-of-sample” portfolio variance, then to choose \( \delta \) and \( k \) we use the nonparametric technique known as cross validation—see Efron and Gong (1983) and Campbell, Lo, and MacKinlay (1997, Section 12.3.2).\(^{12}\) On the other hand, if the objective is to maximize the portfolio return over the last period, we choose \( \theta^* \) so that \( \theta^* = \arg\max_{\theta} w_{\theta}\top r_{\tau}, \) in which \( r_{\tau} \) is the asset-return vector for the last period within the estimation window, and \( w_{\theta} \) is the norm-constrained or partial minimum-variance portfolio computed using all the data over the estimation window with the parameter \( \theta \). That is, we choose the parameter \( \theta = \{\delta, k\} \) to maximize the return in the last period within the estimation window.

Panel B of Table 2 lists the portfolios to which we compare the performance of the norm-constrained portfolios. The first two comparison portfolios are simple benchmarks that require neither estimation nor optimization: the \( 1/N \) portfolio and the value-weighted market portfolio, which we compute as the portfolio that assigns a weight to each asset equal to the market capitalization of

---

\(^{11}\)Our motivation for the portfolio autocorrelation criterion is the work by Campbell, Lo, and MacKinlay (1997), who report: “Despite the fact that individual security returns are weakly negatively autocorrelated, portfolio returns—which are essentially averages of individual security returns—are strongly autocorrelated. This somewhat paradoxical result can mean only one thing: large positive cross-autocorrelations across individual securities across time” (p. 74). We consider the return in only the last period because this is where the autocorrelation is highest. We have also considered the average return in the last 2–6 months but the out-of-sample performance is worse than when using only the return in the previous month.

\(^{12}\)Cross validation works as follows. Given an estimation window composed of \( \tau \) sample returns, for each \( t \) ranging from 1 to \( \tau \) perform the following four steps. First, delete the \( t^{th} \) sample return from the estimation window and compute the sample covariance matrix corresponding to the dataset without the \( t^{th} \) sample return, \( \hat{\Sigma}_{(t)} \). Second, compute the corresponding portfolio \( (w_{\theta})_{(t)} \), where \( \theta = \delta \) for the case of the norm-constrained portfolios and \( \theta = k \) (with \( 1 \leq k \leq N - 1 \)) for the case of the partial minimum-variance portfolios. Third, compute the “out-of-sample” return attained by this portfolio on the \( t^{th} \) sample asset return \( (r_{\theta})_{(t)} = \left((w_{\theta})_{(t)}\right)^\top r_t \). Then the variance of the “out-of-sample” portfolio return is given by the sample variance of the \( \tau \) out-of-sample returns, \( (r_{\theta})_{(t)} \); that is, \( \hat{\sigma}_\theta^2 = \frac{1}{\tau - 1} \sum_{t=1}^{\tau} \left((r_{\theta})_{(t)} - (r_{\theta})_{(t)}\right)^2/\tau - 1, \) in which \( (r_{\theta})_{(t)} = \sum_{t=1}^{\tau} (r_{\theta})_{(t)}/\tau \). Finally, choose the parameter \( \theta^* \) that minimizes this “out-of-sample” return variance; that is, \( \theta^* = \arg\min_{\theta} \hat{\sigma}_\theta^2 \).
that asset divided by the market capitalization of all the assets in the dataset. We also consider two portfolios that rely on estimates of mean returns: the traditional short-sale unconstrained mean-variance portfolio and the Bayesian mean-variance portfolio, which is selected using the approach in Jorion (1985, 1986); both of these portfolios are computed assuming a risk aversion of $\gamma = 5$.\textsuperscript{13} The next three portfolios we consider are the shortsale-unconstrained minimum-variance portfolio, the shortsale-constrained minimum-variance portfolio analyzed in Jagannathan and Ma (2003), and the minimum-variance portfolio under the assumption that returns are described by a 1-factor model with the market as the only factor. We also consider minimum-variance portfolios that are formed using a covariance matrix that is a weighted average of two estimators: in one version of this, we use the approach in Ledoit and Wolf (2004), where the two estimators are the sample covariance matrix and the identity matrix, and in the second version we use the approach in Ledoit and Wolf (2003), where the two estimators are the sample covariance matrix and the 1-factor covariance matrix with the market as the factor. Finally, we consider the parametric portfolios of Brandt, Santa-Clara, and Valkanov (2005) that rely on firm-specific characteristics; again, we assume that the investor has a risk aversion equal to five.\textsuperscript{14}

4.2 Description of the Methodology Used to Evaluate Performance

We compare the performance of the norm-constrained portfolios to the portfolios in the literature using three criteria: (i) out-of-sample portfolio variance; (ii) out-of-sample portfolio Sharpe ratio; and (iii) portfolio turnover (trading volume). We use the following “rolling-horizon” procedure for the comparison. First, we choose a window over which to perform the estimation. We denote the length of the estimation window by $\tau < T,$ where $T$ is the total number of returns in the dataset. For our experiments, we use an estimation window of $\tau = 120$ data points, which for monthly data

\textsuperscript{13}For the risk-aversion parameter we have considered also values of 1, 2, and 10, but because the insights are similar we do not report these results.

\textsuperscript{14}We do not consider several other portfolios. For instance, we do not consider estimators of the covariance matrix based on daily returns because Jagannathan and Ma (2003, Section III) find that their performance is similar to that of the shortsale-constrained minimum-variance portfolio with monthly returns. Also, we do not consider portfolios based on the constant correlation model of Elton and Gruber (1973) because these are outperformed by portfolios proposed in Ledoit and Wolf (2003). We report results only for the single market-factor model because Chan, Karceski, and Lakonishok (1999) show that several factor models with 1, 3, 4, 8, and 10 factors (based on financial market variables and firm-specific characteristics) are not better than the 1-factor market model. Finally, we consider shortsale constraints only for the minimum-variance portfolio because Jagannathan and Ma (2003) find: “for the factor models and shrinkage estimators, imposing such constraints is likely to hurt” (pp. 1653–4).
corresponds to ten years.\footnote{We have tried other estimation window lengths, such as \(\tau=60\) and 240, but the results are similar, and thus, we report the results only for the case \(\tau=120\).} Two, using the return data over the estimation window, \(\tau\), we compute the various portfolios. Three, we repeat this “rolling-window” procedure for the next month, by including the data for the next month and dropping the data for the earliest month. We continue doing this until the end of the dataset is reached. At the end of this process, we have generated \(T-\tau\) portfolio-weight vectors for each strategy; that is, \(w_t^i\) for \(t=\tau, \ldots, T-1\) and for each strategy \(i\). Holding the portfolio \(w_t^i\) for one month gives the out-of-sample return at time \(t+1\): 
\[
r_{t+1}^i = w_t^i r_{t+1},
\]
where \(r_{t+1}\) denotes the asset returns. We use the time series of returns and weights for each strategy to compute the out-of-sample variance, Sharpe ratio, and turnover:

\[
\hat{\sigma}^2_i = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} \left( w_t^i r_{t+1} - \hat{\mu}^i \right)^2, \quad \text{with} \quad \hat{\mu}^i = \frac{1}{T - \tau} \sum_{t=\tau}^{T-1} w_t^i r_{t+1}, \tag{11}
\]

\[
\hat{\text{SR}}^i = \frac{\hat{\mu}^i}{\hat{\sigma}_i}, \tag{12}
\]

\[
\text{Turnover} = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} \sum_{j=1}^{N} \left( |w_{j,t}^i - w_{j,t+1}^i| \right), \tag{13}
\]

where in the definition of turnover, \(w_{j,t}^i\) denotes the portfolio weight in asset \(j\) chosen at time \(t\) under strategy \(i\), \(w_{j,t+1}^i\) the portfolio weight before rebalancing but at \(t+1\), and \(w_{j,t+1}^i\) the desired portfolio weight at time \(t+1\) (after rebalancing), implying that turnover is equal to the sum of the absolute value of the rebalancing trades across the \(N\) available assets and over the \(T - \tau - 1\) trading dates, normalized by the total number of trading dates.

To measure the statistical significance of the difference between the variances and Sharpe ratios of the returns for two given portfolios, we use bootstrapping methods, which are suitable when portfolio returns are not independently and identically distributed as a multivariate Normal. In particular, to compute the P-values for the Sharpe ratios we use a bootstrapping methodology proposed in Wolf (2007) that is designed for the case where portfolio returns have fat tails and are of a time series nature (for instance, returns are serially correlated or exhibit volatility clustering). Specifically, to test the hypothesis that the Sharpe ratio of the return of portfolio \(i\) is equal to that of portfolio \(n\), that is, \(H_0 : \mu_i/\sigma_i - \mu_n/\sigma_n = 0\), we compute a two-sided P-value using the studentized circular block method.
bootstrap proposed in Wolf (2007) with $B = 1000$ bootstrap resamples and a block size equal to $b = 5$. We do this using the code available at http://www.iew.uzh.ch/chairs/wolf.html. To test the hypothesis that the variance of the returns of two portfolios is equal, that is, $H_0: \sigma^2_i - \sigma^2_n = 0$, we use the (non-studentized) stationary bootstrap of Politis and Romano (1994) to construct a two-sided confidence interval for the difference between the variances. We have used $B = 1000$ bootstrap resamples and an expected block size $b = 5$. Then we use the methodology suggested in (Wolf, 2007, Remark 3.2) to generate the resulting bootstrap p-values.

4.3 Discussion of the Out-of-Sample Performance

Table 3 reports the out-of-sample variances for the different portfolios and the corresponding P-value that the portfolio variance for that strategy is different from that for the PARV portfolio. We have also computed all other pairwise P-values, and although we do not report them in the table, we use them when comparing portfolio strategies in our discussion below, and we say that the difference is significant if this P-value is smaller than 5%.

From Panel A in Table 3 we see that the out-of-sample variance for the norm-constrained portfolios calibrated using cross-validation over the return variance (NC1V, NC2V, NCFV, PARV) is similar across the five datasets. Also, not surprisingly, the out-of-sample variance is lower for these portfolios than for those that are calibrated using the criterion of maximizing the return of the portfolio in the previous period (NC1R, NC2R, NCFR, PARR).

Comparing the variances of the norm-constrained portfolios in Panel A to those of the portfolios from the literature listed in Panel B of Table 3, we see that the norm-constrained portfolios typically have lower out-of-sample variances than the portfolios from the literature. For instance, NC1V, NC2V, NCFV, and PARV always achieve out-of-sample variances that are lower than those of the $1/N$, value-weighted (VW), and mean-variance (MEAN) portfolios, and the differences are statistically significant. Similarly, the norm-constrained portfolios have lower variances than the Bayesian portfolio (BAYE) for all datasets except CRSP. The norm-constrained portfolios generally achieve lower out-of-sample variances than the unconstrained minimum-variance portfolio (MINU), and the differences are often statistically significant; for instance, PARV achieves a statistically sig-
Table 3: Portfolio Variances

This table reports the monthly out-of-sample variances and the corresponding P-value that the portfolio variance for a strategy is different from that for the partial minimum-variance portfolio calibrated using cross-validation over portfolio variance (PARV). The P-values are computed using the stationary bootstrap of Politis and Romano (1994) and the method in (Wolf, 2007, Remark 3.2).

<table>
<thead>
<tr>
<th>Strategy</th>
<th>10Ind</th>
<th>48Ind</th>
<th>6FF</th>
<th>25FF</th>
<th>CRSP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel A: Portfolio strategies developed in this paper</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1V</td>
<td>0.00134</td>
<td>0.00126</td>
<td>0.00156</td>
<td>0.00135</td>
<td>0.00074</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.01)</td>
<td>(0.46)</td>
<td>(0.65)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>NC1R</td>
<td>0.00138</td>
<td>0.00135</td>
<td>0.00159</td>
<td>0.00143</td>
<td>0.00080</td>
</tr>
<tr>
<td></td>
<td>(0.98)</td>
<td>(0.26)</td>
<td>(0.97)</td>
<td>(0.10)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>NC2V</td>
<td>0.00134</td>
<td>0.00137</td>
<td>0.00156</td>
<td>0.00130</td>
<td>0.00066</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.21)</td>
<td>(0.13)</td>
<td>(0.43)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>NC2R</td>
<td>0.00149</td>
<td>0.00176</td>
<td>0.00163</td>
<td>0.00152</td>
<td>0.00087</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.00)</td>
<td>(0.63)</td>
<td>(0.02)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>NCFV</td>
<td>0.00135</td>
<td>0.00131</td>
<td>0.00162</td>
<td>0.00134</td>
<td>0.00052</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.03)</td>
<td>(0.47)</td>
<td>(0.82)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>NCFR</td>
<td>0.00144</td>
<td>0.00166</td>
<td>0.00171</td>
<td>0.00170</td>
<td>0.00068</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.01)</td>
<td>(0.07)</td>
<td>(0.00)</td>
<td>(0.53)</td>
</tr>
<tr>
<td>PARV</td>
<td>0.00138</td>
<td>0.00141</td>
<td>0.00159</td>
<td>0.00133</td>
<td>0.00065</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
</tr>
<tr>
<td>PARR</td>
<td>0.00153</td>
<td>0.00163</td>
<td>0.00161</td>
<td>0.00146</td>
<td>0.00085</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.77)</td>
<td>(0.12)</td>
<td>(0.00)</td>
</tr>
<tr>
<td><strong>Panel B: Portfolio strategies from existing literature</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/N</td>
<td>0.00179</td>
<td>0.00221</td>
<td>0.00230</td>
<td>0.00249</td>
<td>0.00169</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>VW</td>
<td>0.00158</td>
<td>0.00190</td>
<td>0.00191</td>
<td>0.00186</td>
<td>0.00157</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.01090</td>
<td>0.38107</td>
<td>0.00353</td>
<td>0.00942</td>
<td>0.00626</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>BAYE</td>
<td>0.00264</td>
<td>0.06793</td>
<td>0.00221</td>
<td>0.00400</td>
<td>0.00666</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>MINU</td>
<td>0.00138</td>
<td>0.00186</td>
<td>0.00156</td>
<td>0.00143</td>
<td>0.00104</td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td>(0.00)</td>
<td>(0.25)</td>
<td>(0.09)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>MINC</td>
<td>0.00134</td>
<td>0.00133</td>
<td>0.00186</td>
<td>0.00176</td>
<td>0.00087</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.33)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>FAC1</td>
<td>0.00145</td>
<td>0.00159</td>
<td>0.00202</td>
<td>0.00241</td>
<td>0.00075</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.10)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>LWID</td>
<td>0.00131</td>
<td>0.00143</td>
<td>0.00155</td>
<td>0.00126</td>
<td>0.00065</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.75)</td>
<td>(0.28)</td>
<td>(0.06)</td>
<td>(0.75)</td>
</tr>
<tr>
<td>LW1F</td>
<td>0.00135</td>
<td>0.00140</td>
<td>0.00158</td>
<td>0.00134</td>
<td>0.00052</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.79)</td>
<td>(0.86)</td>
<td>(0.91)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>BSV3</td>
<td>0.00601</td>
<td>0.00392</td>
<td>0.00306</td>
<td>0.00344</td>
<td>0.00574</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>
nificant lower variance than MINU for the 48Ind and CRSP datasets. Also, the 1-norm-constrained portfolios (NC1V) generally have lower variance than the shortsale-constrained minimum-variance portfolios (MINC) that NC1V nests, and the differences are statistically significant for the 6FF, 25FF, and CRSP datasets. We also observe that the norm-constrained portfolios NC1V, NC2V, NCFV, and PARV all attain lower out-of-sample variances than the minimum-variance portfolio based on a 1-factor market model (FAC1), and the differences are often statistically significant; for example, the difference between the variances of PARV and FAC1 is statistically significant for the 6FF and 25FF datasets.

However, the out-of-sample variances of the 2- and \( \Sigma_F \)-norm-constrained portfolios (NC2V, NCFV) are not always lower than those of the Ledoit and Wolf (2003, 2004) portfolios (LWID, LW1F). For instance, the LWID portfolio attains lower variances than the NC2V portfolio for the 10Ind and 25FF datasets, with the difference being significant for the 10Ind dataset. On the other hand, the LW1F portfolio has a higher variance than the NCFV portfolio for the 48Ind dataset, and the difference is statistically significant.

Table 4 reports the out-of-sample Sharpe ratios for the different portfolios and the corresponding P-value that the Sharpe ratio for each of these strategies is different from that for the PARR portfolio. We have also computed all other pairwise P-values and when comparing portfolio strategies in our discussion below we say that the difference is significant if this P-value is smaller than 5%.

From Panel A of Table 4 we see that the partial minimum-variance portfolio calibrated by maximizing the portfolio return in the last period (PARR) almost always attains higher Sharpe ratios than the 1-, 2-, and \( \Sigma_F \)-norm-constrained portfolios (NC1R, NC2R, NCFR), although the differences are significant only between PARR and NC1R, but not between PARR and NC2R or NCFR.

Comparing the Sharpe ratios of the norm-constrained portfolios in Panel A to those of the portfolios from the literature listed in Panel B, we see that the norm-constrained portfolios have higher Sharpe ratios than both the equally-weighted \((1/N)\) and the value-weighted \((VW)\) portfolios for all datasets, and the difference is substantial and significant in most cases; in fact, PARR attains
Table 4: Portfolio Sharpe Ratios

This table reports the monthly out-of-sample Sharpe ratio and the corresponding P-value that the Sharpe ratio for each of these strategies is different from that for the partial minimum-variance portfolio calibrated by maximizing the portfolio return in the previous period (PARR). The P-values are computed using the studentized circular block bootstrapping methodology in Wolf (2007).

<table>
<thead>
<tr>
<th>Strategy</th>
<th>10Ind</th>
<th>48Ind</th>
<th>6FF</th>
<th>25FF</th>
<th>CRSP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Portfolio strategies developed in this paper</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1V</td>
<td>0.2854</td>
<td>0.2886</td>
<td>0.3385</td>
<td>0.3649</td>
<td>0.4013</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.32)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>NC1R</td>
<td>0.2890</td>
<td>0.2831</td>
<td>0.3374</td>
<td>0.3553</td>
<td>0.3706</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.19)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>NC2V</td>
<td>0.2919</td>
<td>0.2855</td>
<td>0.3527</td>
<td>0.4089</td>
<td>0.3994</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.22)</td>
<td>(0.10)</td>
<td>(0.22)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>NC2R</td>
<td>0.3193</td>
<td>0.2891</td>
<td>0.3922</td>
<td>0.4278</td>
<td>0.4672</td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
<td>(0.05)</td>
<td>(0.93)</td>
<td>(0.36)</td>
<td>(0.55)</td>
</tr>
<tr>
<td>NCFV</td>
<td>0.2927</td>
<td>0.2808</td>
<td>0.3479</td>
<td>0.3728</td>
<td>0.4463</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.27)</td>
<td>(0.08)</td>
<td>(0.03)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>NCFR</td>
<td>0.3114</td>
<td>0.2723</td>
<td>0.3186</td>
<td>0.3815</td>
<td>0.4243</td>
</tr>
<tr>
<td></td>
<td>(0.56)</td>
<td>(0.22)</td>
<td>(0.01)</td>
<td>(0.11)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>PARV</td>
<td>0.2841</td>
<td>0.2823</td>
<td>0.3478</td>
<td>0.4077</td>
<td>0.3937</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.29)</td>
<td>(0.10)</td>
<td>(0.25)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>PARR</td>
<td>0.3293</td>
<td>0.3166</td>
<td>0.3912</td>
<td>0.4403</td>
<td>0.4768</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
</tr>
<tr>
<td><strong>Panel B: Portfolio strategies from existing literature</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/N</td>
<td>0.2541</td>
<td>0.2508</td>
<td>0.2563</td>
<td>0.2565</td>
<td>0.3326</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.10)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>VW</td>
<td>0.2619</td>
<td>0.2698</td>
<td>0.2437</td>
<td>0.2558</td>
<td>0.2748</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.23)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.0499</td>
<td>-0.0334</td>
<td>0.3214</td>
<td>0.2253</td>
<td>0.0723</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.22)</td>
<td>(0.02)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>BAYE</td>
<td>0.1685</td>
<td>-0.0121</td>
<td>0.3666</td>
<td>0.3151</td>
<td>0.4018</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.57)</td>
<td>(0.15)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>MINU</td>
<td>0.2865</td>
<td>0.2222</td>
<td>0.3640</td>
<td>0.4199</td>
<td>0.3820</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.01)</td>
<td>(0.30)</td>
<td>(0.50)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>MINC</td>
<td>0.2852</td>
<td>0.2914</td>
<td>0.2629</td>
<td>0.2720</td>
<td>0.3985</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.44)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>FAC1</td>
<td>0.3060</td>
<td>0.2674</td>
<td>0.2485</td>
<td>0.2486</td>
<td>0.4166</td>
</tr>
<tr>
<td></td>
<td>(0.54)</td>
<td>(0.26)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.37)</td>
</tr>
<tr>
<td>LWID</td>
<td>0.2962</td>
<td>0.2620</td>
<td>0.3226</td>
<td>0.3974</td>
<td>0.4086</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.05)</td>
<td>(0.00)</td>
<td>(0.10)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>LW1F</td>
<td>0.2902</td>
<td>0.2544</td>
<td>0.3296</td>
<td>0.3927</td>
<td>0.4500</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.09)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>BSV3</td>
<td>0.1157</td>
<td>0.3314</td>
<td>0.3907</td>
<td>0.4047</td>
<td>0.2674</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.82)</td>
<td>(0.99)</td>
<td>(0.62)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>
a significantly higher Sharpe ratio for all datasets except 48Ind. The difference in performance is even more striking when the norm-constrained portfolios are compared to the traditional mean-variance (MEAN) and the Bayesian-mean-variance strategies (BAYE).

The PARR portfolio attains a higher Sharpe ratio than MINU, MINC, and FAC1 for all the datasets, and the difference is often significant. Moreover, the 1-norm-constrained portfolio calibrated by maximizing the return of the portfolio in the previous period (NC1R) typically outperforms the shortsale-constrained minimum-variance portfolio (MINC) that it nests, and the differences are significant for the 6FF and 25FF datasets.

The PARR portfolio attains a higher Sharpe ratio than LWID and LW1F for all the datasets, and the difference is statistically significant for the 48Ind and 6FF datasets, although as we will see below, the higher Sharpe ratio of PARR is accompanied by higher turnover. Also, the 2-norm-constrained portfolio (NC2R) always attains a higher out-of-sample Sharpe ratio than the LWID portfolio, although the difference is significant only for the 6FF dataset. However, the $\hat{\Sigma}_F$-norm-constrained portfolio calibrated by maximizing the return of the portfolio in the previous period (NCFR) fails to achieve higher Sharpe ratios than the corresponding LW1F portfolio.

Finally, even though the PARR portfolio does not use firm-specific characteristics, it achieves Sharpe ratios that are at least as good as those for the parametric portfolios (BSV) developed in Brandt, Santa-Clara, and Valkanov (2005), and the differences are significant for the 10Ind and CRSP datasets.

From Panel A of Table 5 we see that the turnover of the norm-constrained portfolios calibrated using cross-validation over the portfolio variance is much lower than that of the portfolios calibrated by maximizing the portfolio return. Panel B of this table shows, not surprisingly, that the best portfolios in terms of turnover are the $1/N$ and value-weighted portfolios.\footnote{We compute the value-weighted portfolio for each dataset as the portfolio that assigns a weight to each asset equal to the market capitalization of that asset divided by the market capitalization of all the assets in the dataset. Note that because the composition of the “market portfolio” may be changing over time, the strategy of holding the value-weighted portfolio may have a turnover that is different from zero.} The turnover of these portfolios is followed by that of the shortsale-constrained minimum-variance portfolio (MINC). The turnovers of the 1-, 2-, and $\hat{\Sigma}_F$-norm-constrained portfolios and the partial minimum-variance portfolio calibrated with cross-validation over variance, and the LWID and LW1F portfolios pro-
Table 5: Portfolio Turnovers

This table reports the monthly turnover of the various portfolios. Turnover is the average percentage of wealth traded in each period and, as defined in Equation (13), is equal to the sum of the absolute value of the rebalancing trades across the \( N \) available assets and over the \( T - \tau - 1 \) trading dates, normalized by the total number of trading dates.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>10Ind</th>
<th>48Ind</th>
<th>6FF</th>
<th>25FF</th>
<th>CRSP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Portfolio strategies developed in this paper</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1V</td>
<td>0.1494</td>
<td>0.2680</td>
<td>0.1729</td>
<td>0.2407</td>
<td>0.6141</td>
</tr>
<tr>
<td>NC1R</td>
<td>0.6013</td>
<td>0.8232</td>
<td>1.0064</td>
<td>0.9767</td>
<td>0.9753</td>
</tr>
<tr>
<td>NC2V</td>
<td>0.1448</td>
<td>0.3266</td>
<td>0.1946</td>
<td>0.4570</td>
<td>0.5808</td>
</tr>
<tr>
<td>NC2R</td>
<td>1.0177</td>
<td>2.7556</td>
<td>1.6594</td>
<td>3.6275</td>
<td>1.0443</td>
</tr>
<tr>
<td>NCFV</td>
<td>0.1052</td>
<td>0.2469</td>
<td>0.2790</td>
<td>0.4134</td>
<td>0.5729</td>
</tr>
<tr>
<td>NCFR</td>
<td>0.6944</td>
<td>2.3117</td>
<td>1.9952</td>
<td>3.3560</td>
<td>1.0584</td>
</tr>
<tr>
<td>PARV</td>
<td>0.1689</td>
<td>0.3838</td>
<td>0.2600</td>
<td>0.4628</td>
<td>0.5743</td>
</tr>
<tr>
<td>PARR</td>
<td>1.0414</td>
<td>2.4846</td>
<td>1.6407</td>
<td>3.5657</td>
<td>1.0984</td>
</tr>
<tr>
<td><strong>Panel B: Portfolio strategies from existing literature</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/N</td>
<td>0.0232</td>
<td>0.0311</td>
<td>0.0155</td>
<td>0.0174</td>
<td>0.0595</td>
</tr>
<tr>
<td>VW</td>
<td>0.0382</td>
<td>0.0540</td>
<td>0.0213</td>
<td>0.0310</td>
<td>0.2316</td>
</tr>
<tr>
<td>MEAN</td>
<td>1.0135</td>
<td>105.6126</td>
<td>0.7987</td>
<td>4.2495</td>
<td>3.0014</td>
</tr>
<tr>
<td>BAYE</td>
<td>0.3565</td>
<td>6.6314</td>
<td>0.5388</td>
<td>2.1264</td>
<td>0.6191</td>
</tr>
<tr>
<td>MINU</td>
<td>0.1656</td>
<td>0.8286</td>
<td>0.2223</td>
<td>0.7953</td>
<td>0.7769</td>
</tr>
<tr>
<td>MINC</td>
<td>0.0552</td>
<td>0.0741</td>
<td>0.0461</td>
<td>0.0841</td>
<td>0.4222</td>
</tr>
<tr>
<td>FAC1</td>
<td>0.0935</td>
<td>0.2047</td>
<td>0.1152</td>
<td>0.2398</td>
<td>0.3650</td>
</tr>
<tr>
<td>LWID</td>
<td>0.1132</td>
<td>0.4029</td>
<td>0.0905</td>
<td>0.3144</td>
<td>0.5594</td>
</tr>
<tr>
<td>LW1F</td>
<td>0.1428</td>
<td>0.4034</td>
<td>0.1455</td>
<td>0.4893</td>
<td>0.5463</td>
</tr>
<tr>
<td>BSV3</td>
<td>0.4683</td>
<td>0.9066</td>
<td>0.5384</td>
<td>0.5564</td>
<td>2.1926</td>
</tr>
</tbody>
</table>

The short-sale-unconstrained minimum-variance portfolio and the portfolios based on factor models have higher turnover than MINC and the norm-constrained strategies calibrated to minimize portfolio variance. The partial minimum-variance portfolio calibrated by maximizing the portfolio return in the last month (PARR) and the parametric portfolios based on the work by Brandt, Santa-Clara, and Valkanov (2005) have similar turnovers, which are much higher than those of the rest of the portfolios.

5 Conclusion

We provide a general unifying framework for determining portfolios in the presence of estimation error. This framework is based on shrinking directly the portfolio-weight vector rather than some or
all of the elements of the sample covariance matrix. This is accomplished by solving the traditional minimum-variance problem (based on the sample covariance matrix) but subject to the additional constraint that the norm of the portfolio-weight vector be smaller than a given threshold. We show that our general framework nests as special cases the shrinkage approaches of Jagannathan and Ma (2003), Ledoit and Wolf (2003, 2004), and the $1/N$ portfolio studied in DeMiguel, Garlappi, and Uppal (2007). We also compare empirically the out-of-sample performance of the new portfolios we have proposed to ten strategies in the literature across five datasets. We find that the norm-constrained portfolios often have a higher Sharpe ratio than the portfolio strategies in Jagannathan and Ma (2003), Ledoit and Wolf (2003, 2004), the $1/N$ portfolio, and other strategies in the literature, such as factor portfolios.

References


A Appendix: Proofs for the Propositions

Proof for Proposition 1

Note that the 1-norm-constrained minimum-variance portfolio problem with δ = 1 can be obtained from the shortsale-constrained problem by replacing the shortsale constraint w ≥ 0 with the 1-norm-constraint \( \|w\|_1 \leq 1 \). Straightforward algebra shows that, in the presence of the additional constraint that the portfolio weights add up to one, these two constraints are indeed equivalent. ■

Proof for Proposition 2

Note that the portfolio that minimizes the A-norm subject to the constraint that the weights add up to one can be written as \( w_F = A^{-1}e/(e^TA^{-1}e) \). Hence, if \( \delta \geq 1/(e^TA^{-1}e) \), then there exists a feasible point for the A-norm-constrained minimum-variance portfolio problem. If, in addition, \( \hat{\Sigma} \) is nonsingular, then there is a unique global minimizer to the problem, and this unique minimizer is characterized by the first-order optimality conditions, which imply that there exist Lagrange multipliers \( \psi \) and \( \nu \) such that

\[
2\hat{\Sigma}w + 2\nu Aw - \psi e = 0, \quad (A1)
\]

\[
w^T e = 1, \quad (A2)
\]

\[
w^T Aw \leq \delta, \quad (A3)
\]

\[
\nu \geq 0. \quad (A4)
\]

These conditions imply that the first-order optimality conditions hold for the minimum-variance problem with \( \hat{\Sigma}_{LW} = \frac{1}{1+\nu} \hat{\Sigma} + \frac{\nu}{1+\nu} A \). Moreover, because \( \hat{\Sigma} \) is nonsingular and positive definite, \( A \) is positive definite, and \( \nu \geq 0 \), we know that the matrix \( \hat{\Sigma}_{LW} \) is positive definite, and the first-order optimality conditions for the minimum-variance problem with \( \hat{\Sigma}_{LW} \) are then also sufficient for optimality. ■

Proof for Proposition 3

Note that the 1/N portfolio is the portfolio with minimum 2-norm out of all portfolios whose weights sum up to one. Moreover, we use the 1/N portfolio as the starting point for the conjugate
gradient method when we apply this method to solve the minimum-variance portfolio problem. Then, it trivially follows from the analysis in Phatak and De Hoog (2003) and Steihaug (1983) that the 2-norm of the iterates generated by the conjugate-gradient method is nondecreasing. ■

Proof for Proposition 4

Because the prior distributions of the portfolio weights are independent, the prior distribution for a portfolio \( w \) is

\[
\pi(w) = \prod_{i=1}^{N} \frac{\nu}{2} e^{-\nu |w_i|} = \frac{\nu^N}{2^N} e^{-\nu \|w\|_1}.
\]  

(A5)

Then Bayes theorem implies that the posterior distribution of the portfolio weights conditional on the observed sample returns is

\[
\pi(w, \sigma^2 | \{r_t\}_{t=1}^T) = \frac{\pi(w) \pi(\sigma^2) L(w, \sigma^2 | \{r_t\}_{t=1}^T)}{\int_w \pi(w) \pi(\sigma^2) L(w, \sigma^2 | \{r_t\}_{t=1}^T) dw d\sigma^2},
\]  

(A6)

where

\[
L(w, \sigma^2 | \{r_t\}_{t=1}^T) = \frac{1}{\sigma \sqrt{2^N \pi^N}} \exp \left( -\frac{\sum_{t=1}^{T} (w^\top r_t - w^\top \hat{\mu})^2}{2\sigma^2} \right)
\]

is the likelihood function of \( w \) and \( \sigma^2 \) given the observed sample returns \( \{r_t\}_{t=1}^T \). Hence, we have that

\[
\pi(w, \sigma^2 | \{r_t\}_{t=1}^T) \propto \pi(w) \pi(\sigma^2) L(w, \sigma^2 | \{r_t\}_{t=1}^T),
\]  

(A7)

where the symbol \( \propto \) means that the distribution is proportional to the right-hand-side term. Simple algebra then shows that

\[
\pi(w, \sigma^2 | \{r_t\}_{t=1}^T) \propto \frac{\sigma^2}{\sigma^N} \exp \left( -\frac{\sum_{t=1}^{T} (w^\top r_t - w^\top \hat{\mu})^2}{2\sigma^2} - \nu \|w\|_1 \right).
\]  

(A8)

Hence, for any value of \( \sigma \) the portfolio that maximizes the posterior distribution of the minimum-variance portfolio weights, subject to the condition that the portfolio weights add up to one, is the

---

\( \text{Note that } L(w, \sigma^2 | \{r_t\}_{t=1}^T) \text{ coincides with the joint density function for the distribution of the sequence of independent portfolio returns } \{w^\top r_t\}_{t=1}^T \text{ conditional on the values of } \sigma \text{ and } w; \text{ that is, } \pi(\{w^\top r_t\}_{t=1}^T | \sigma, w). \text{ This conditional distribution is customarily termed in Bayesian Statistics as the likelihood function.} \n\)
solution to the following optimization problem:
\[
\min_w \ w^\top \hat{\Sigma} w + \rho \|w\|_1,
\]
\[\text{s.t. } \ w^\top e = 1,\]  
where \(w^\top \hat{\Sigma} w = \sum_{t=1}^{T}(w^\top r_t - w^\top \hat{\mu})^2/(T - 1)\) and \(\rho = 2\sigma^2\nu/(T - 1)\). Moreover, it follows from optimization theory (Nocedal and Wright, 1999, Chapter 17) that there exists a threshold parameter \(\delta\) such that the solution to problem (A9)–(A10) coincides with the 1-norm-constrained minimum-variance portfolio. Finally, because for any fixed value of \(\sigma\) the portfolio that maximizes the posterior distribution is a 1-norm-constrained minimum-variance portfolio, we know that the portfolio that maximizes the posterior distribution must also be a 1-norm-constrained minimum-variance portfolio.

\[\blacksquare\]

**Proof for Proposition 5**

Following the same line of reasoning as in the proof of Proposition 4, we have by Bayes theorem and simple algebra that
\[
\pi(w, \sigma^2 | \{r_t\}_{t=1}^T) \propto \frac{\pi(\sigma^2)}{\sigma^N} \exp \left( -\frac{\sum_{t=1}^{T}(w^\top r_t - w^\top \hat{\mu})^2}{2\sigma^2} - \frac{1}{2} w^\top A w \right).
\]
Hence, for any value of \(\sigma\) the portfolio that maximizes the posterior distribution of the minimum-variance portfolio weights subject to the condition that the portfolio weights add up to one is the solution to the following optimization problem:
\[
\min_w \ w^\top \hat{\Sigma} w + \rho w^\top Aw, \quad \text{(A12)}
\]
\[\text{s.t. } \ w^\top e = 1, \quad \text{(A13)}\]
where \(w^\top \hat{\Sigma} w = \sum_{t=1}^{T}(w^\top r_t - w^\top \hat{\mu})^2/(T - 1)\) and \(\rho = \sigma^2/(T - 1)\). Moreover, it follows from optimization theory (Nocedal and Wright, 1999, Chapter 17) that there exists a threshold parameter \(\delta\) such that the solution to problem (A12)–(A13) coincides with the \(A\)-norm-constrained minimum-variance portfolio for the threshold \(\delta\). Finally, because for any fixed value of \(\sigma\) the portfolio that maximizes the posterior distribution is a \(A\)-norm-constrained minimum-variance portfolio, we know that the portfolio that maximizes the posterior distribution must also be a \(A\)-norm-constrained minimum-variance portfolio.

\[\blacksquare\]
Proof for Proposition 6

Straightforward algebra shows that the 1-norm constraint can be written as

$$\|w\|_1 = 1 - 2 \sum_{i \in \mathcal{N}(w)} w_i \leq \delta, \quad (A14)$$

in which $\mathcal{N}(w)$ is the set of asset indexes for which the corresponding portfolio weight is negative, $\mathcal{N}(w) = \{i : w_i < 0\}$. Hence, the 1-norm-constrained minimum-variance problem is

$$\min_w w^\top \hat{\Sigma} w, \quad \text{s.t.} \quad w^\top e = 1, \quad 1 - 2 \sum_{i \in \mathcal{N}(w)} w_i \leq \delta.$$ 

At the solution to this problem, and assuming none of the elements of $w$ are equal to zero, there exists a Lagrange multiplier $\nu \geq 0$ such that the solution to this problem coincides with the solution to

$$\min_w w^\top (\hat{\Sigma} - 2 \nu \sum_{i \in \mathcal{N}(w)} w_i), \quad \text{s.t.} \quad w^\top e = 1.$$ 

Moreover, because at the solution $w^\top e = 1$, the problem can be equivalently rewritten as

$$\min_w w^\top (\hat{\Sigma} - \nu ne^\top - \nu n^\top e) w, \quad \text{s.t.} \quad w^\top e = 1,$$

which completes the proof for the proposition.  

\[31\]
Material for on-line appendix

B Appendix: Details of the Partial Minimum-Variance Portfolios

In this appendix we explain the method used to compute the partial minimum-variance portfolios. Because the conjugate-gradient method can be used to solve only unconstrained optimization problems, we first show in Section B.1 how to eliminate the constraint that the portfolio weights sum to one. Section B.2 gives the method statement.

B.1 Expressing Minimum-Variance Problem Without Adding-Up Constraint

To eliminate the constraint $w^\top e = 1$, we will use the following proposition showing that any portfolio whose weights sum to one can be rewritten as the $1/N$ portfolio plus a portfolio whose weights sum to zero, herein a zero-cost portfolio.

Proposition B1 Let $w \in \mathbb{R}^N$ and $w^\top e = 1$, then the following holds:

1. The portfolio $w$ may be expressed as

$$w = \frac{e}{N} + w_0,$$

in which $w_0$ is a zero-cost portfolio; that is, $w_0^\top e = 0$.

2. There exists a matrix $Z \in \mathbb{R}^{N \times (N-1)}$ whose columns form an orthonormal basis for the subspace of zero-cost portfolios (that is, an orthonormal basis for the subspace of all portfolios satisfying $w^\top e = 0$), and this matrix satisfies $e^\top Z = 0$ and $Z^\top Z = I$.

3. There exists a vector $w_Z \in \mathbb{R}^{N-1}$ such that

$$w = \frac{e}{N} + Z w_Z.$$

Proof: Part 1: Let $w_0$ be the portfolio defined as $w_0 = w - e/N$. Then note that if the weights of the portfolio $w$ sum up to one, $w^\top e = 1$, then the weights of the portfolio $w_0$ must sum up to zero because $e^\top w_0 = e^\top w - e^\top e/N = 0$. 

32
Part 2: The existence of an orthonormal basis for the subspace of zero-cost portfolios follows from general vector space theory (Nocedal and Wright (1999)). Let \( Z \in \mathbb{R}^{N \times (N-1)} \) be the matrix whose columns are the vectors composing the orthonormal basis for the zero-cost portfolio subspace, then, because all zero-cost portfolios are orthogonal to \( e \), we must have \( e^\top Z = 0 \). Finally, because the columns of \( Z \) are orthogonal to each other and their 2-norm is equal to one, we have that \( Z^\top Z = I \).

Part 3: The result follows from Part 1 because \( w_0 \) is a zero-cost portfolio and \( Z \) is a basis of the zero-cost portfolio subspace. ■

We define portfolio space to be the \( \mathbb{R}^N \) space, zero-cost portfolio subspace the subspace in \( \mathbb{R}^N \) that is formed by all linear combinations of the columns of \( Z \), and reduced space the \( N-1 \) dimensional space \( \mathbb{R}^{N-1} \), which \( w_Z \) inhabits. The following proposition shows that expression (B2) may be used to eliminate the constraint from the minimum-variance problem.

**Proposition B2** The sample minimum-variance portfolio can be written as \( w_{MINU} = \frac{e}{N} + Z w_Z \), where \( w_{MINU} \) is the solution to the unconstrained minimum-variance problem:

\[
\min_{w_Z} \left( \frac{e}{N} + Z w_Z \right)^\top \hat{\Sigma} \left( \frac{e}{N} + Z w_Z \right). \tag{B3}
\]

**Proof:** Substituting (B1) into (1)–(2) yields the result. ■

**B.2 Applying the Conjugate-Gradient Method**

To compute the partial minimum-variance portfolios, we apply the conjugate-gradient method to solve the first-order optimality conditions for the unconstrained minimum-variance problem:

\[
Z^\top \hat{\Sigma} Z w_Z = -Z^\top \hat{\Sigma} \frac{e}{N}. 
\]

In our implementation, we use the 1/N portfolio as the starting point for the conjugate-gradient method. The method commences by finding a zero-cost portfolio \( w_{CG1} \) that induces the maximum marginal rate of decrease in sample variance when combined with the 1/N portfolio. In other words, \( w_{CG1} \) is a zero-cost portfolio such that when an infinitesimally small multiple of this zero-cost portfolio is added to the 1/N portfolio, it induces the maximum rate of decrease in the overall
portfolio sample variance. Mathematically, this implies that the first conjugate portfolio must be proportional to the negative of the projection of the gradient of the portfolio sample variance onto the zero-cost portfolio subspace:

$$w_{CG1} \propto -ZZ^\top \nabla_w \left( w^\top \hat{\Sigma} w \right) \bigg|_{e/N},$$  \hspace{1cm} (B4)

where $ZZ^\top$ is the projection matrix that projects any portfolio onto the zero-cost subspace and $\nabla_w \left( w^\top \hat{\Sigma} w \right) \bigg|_{e/N}$ is the gradient of the portfolio sample variance evaluated at the $1/N$ portfolio. Evaluating this gradient we have that the first conjugate portfolio can be written as:  

$$w_{CG1} = -ZZ^\top \hat{\Sigma} \frac{e}{N}.$$ \hspace{1cm} (B5)

Then the conjugate-gradient method finds the combination of the $1/N$ portfolio and the first conjugate-gradient portfolio that minimizes the sample portfolio variance. We term the resulting portfolio the first partial minimum-variance portfolio ($w_{PAR1}$), where the term “partial” refers to the fact that the portfolio minimizes the sample variance within the subset of portfolios formed by combinations of the $1/N$ portfolio and the first conjugate portfolio.

One may feel tempted to follow the same strategy in the successive iterations, and hence find the zero-cost portfolio that would induce the maximum marginal rate of decrease in sample variance when combined with $w_{PAR1}$, but it can be shown (Nocedal and Wright (1999)) that it is more efficient to find the zero-cost portfolio $w_{CG2}$ that induces the maximum marginal rate decrease in sample variance, subject to the condition that it is conjugate with respect to $w_{CG1}$ (i.e., $w_{CG2} \hat{\Sigma} w_{CG1} = 0$). Then the method finds the optimal combination of $w_{PAR1}$ and $w_{CG2}$, which is termed the second partial minimum-variance portfolio. If this process is iterated $N - 1$ times, we recover the minimum-variance portfolio. If instead we perform this operation only $K < N - 1$ times, we obtain a portfolio that lies between the $1/N$ portfolio and the minimum-variance portfolio.

The following proposition describes the relation between the partial minimum-variance portfolios and the 2-norm-constrained portfolios.

---

18 For notational convenience we have divided the gradient of the sample variance by a factor of 2. Note that this is correct because in (B4) we only state that $w_{CG1}$ is proportional to the expression in the right-hand side.
Proposition B3: Provided the matrix $\hat{\Sigma}$ is nonsingular, the 2-norm-constrained portfolios form a continuously differentiable curve for $\delta$ ranging from $1/N$ to $\|w_{MINU}\|_2^2$, and the difference between the first partial minimum-variance portfolio and the $1/N$ portfolio is tangential to this curve at $\delta = 1/N$.

Proof: Using the unconstrained minimum-variance problem (B3), we can rewrite the 2-norm-constrained minimum-variance problem as

$$
\min_{w_Z} \quad \left( \frac{e}{N} + Zw_Z \right)^\top \hat{\Sigma} \left( \frac{e}{N} + Zw_Z \right)
$$

s.t. $\|e/N\|_2^2 + \|w_Z\|_2^2 \leq \delta$.

Moreover, because $\|e/N\|_2^2 = 1/N$ we can then rewrite this problem as

$$
\min_{w_Z} \quad \left( \frac{e}{N} + Zw_Z \right)^\top \hat{\Sigma} \left( \frac{e}{N} + Zw_Z \right)
$$

s.t. $\|w_Z\|_2^2 \leq \tilde{\delta},$

in which $\tilde{\delta} = \delta - 1/N$. Also, provided the matrix $\hat{\Sigma}$ is nonsingular, then we also have that $Z^\top \hat{\Sigma} Z$ is nonsingular; hence, the solution to problem (B8)–(B9) is unique and is the unique solution to the first-order optimality conditions for problem (B8)–(B9). Then, following an argument similar to that in the proof of Proposition 2, we have that, provided the matrix $\hat{\Sigma}$ is nonsingular, for each $\delta$ there exists a $\nu \geq 0$ such that the solution to the unique 2-norm-constrained problem coincides with the unique solution to the following problem:

$$
\min_{w_Z} \quad \left( \frac{e}{N} + Zw_Z \right)^\top \hat{\Sigma} \left( \frac{e}{N} + Zw_Z \right) + \nu w_Z^\top w_Z,
$$

in which $w_Z^\top w_Z = \|w_Z\|_2^2$. Moreover, the uniqueness of the minimizer to this problem ensures that the curve of 2-norm-constrained portfolios for different values of $\delta$ or $\nu$ is continuously differentiable (see Fiacco and McCormick (1968, Theorem 14)). The first-order conditions give the unique solution to problem (B10):

$$
w_Z = -(Z^\top \hat{\Sigma} Z + \nu I)^{-1} Z^\top \hat{\Sigma} \frac{e}{N}.
$$

We are interested in the tangent to the curve of solutions to the 2-norm-constrained problem. This is given by the differential of expression (B11) with respect to $\nu$. We are interested in this
differential at the $1/N$ portfolio, which is obtained for very large values of the penalty parameter $\nu$. For very large values of $\nu$, we have that

$$ (Z^\top \hat{\Sigma} Z + \nu I) \approx \nu I, \quad \text{(B12)} $$

and hence,

$$ \frac{d}{d\nu} (Z^\top \hat{\Sigma} Z + \nu I)^{-1} \approx \frac{d}{d\nu} (\nu I)^{-1} = \frac{d}{d\nu} (\frac{1}{\nu} I) = -\frac{1}{\nu^2} I. \quad \text{(B13)} $$

Therefore,

$$ \frac{dw}{d\nu} = \frac{1}{\nu^2} Z^\top \hat{\Sigma} e \frac{1}{N}, \quad \text{(B14)} $$

and thus,

$$ \frac{dw}{d\nu} = \frac{1}{\nu^2} ZZ^\top \hat{\Sigma} e \frac{1}{N}. \quad \text{(B15)} $$

By equation (B5), we know that the right-hand side in (B15) is proportional to the first conjugate portfolio. Moreover, by definition, the first conjugate portfolio is proportional to the difference between the first partial minimum-variance portfolio and the $1/N$ portfolio.

We now give a brief mathematical description of the algorithm followed to compute the partial minimum-variance portfolios; for a rigorous treatment of the conjugate-gradient method see Nocedal and Wright (1999, Chapter 5). Let $w_{PARk}$ be the $k$th partial minimum-variance portfolio and let $\epsilon^k = -Z^\top \hat{\Sigma} w_{PARk}$ denote the residual at the $k$th partial minimum-variance portfolio. Select the number of conjugate portfolios $K \leq N$, and let $w_{PAR0} = e/N$ and $k = 0$. The $(k + 1)$th partial minimum-variance portfolio is defined as $w_{PARk+1} = w_{PARk} + \alpha_k w_{CG(k+1)}$, in which $w_{CG(k+1)} = Ze^k + \beta^k w_{CGk}$ is the $(k + 1)$th conjugate portfolio; for $k = 0$ we set $\beta^k = 0$ and for $k > 0$ we set

$$ \beta^k = \frac{\epsilon^k \epsilon^k}{(\epsilon^{k-1}) \epsilon^{k-1}}, \quad \text{and the step-size is } \alpha_k = \frac{(\epsilon_k)^\top \epsilon_k}{(w_{CG(k+1)})^\top \hat{\Sigma} w_{CG(k+1)}}. $$

---

19 This follows from basic penalty parameter theory (see Nocedal and Wright (1999, Chapter 17)).