An averaging framework for minimum-variance portfolios: Optimal rules for combining portfolio weights*

Roland Füss†, Thorsten Glück‡, Christian Koeppel§, Felix Miebs¶

September 20, 2024

Abstract

We propose an averaging framework for combining minimum-variance strategies to either minimize the expected out-of-sample variance or maximize the expected out-of-sample Sharpe ratio. Our framework overcomes the problem of selecting the "best" strategy ex-ante by optimally averaging over portfolio weights. This averaging procedure has an intuitive economic interpretation because it resembles a fund-of-fund approach, where each minimum-variance strategy represents a single fund. In a range of simulations, for a set of well-established strategies, we show that optimally averaging over portfolio weights improves the out-of-sample variance and Sharpe ratio. We confirm the finding of our simulation study on empirical data.

JEL Classifications: G11

Keywords: Averaging; diversification; estimation error; portfolio optimization; shrinkage.

*We would like to thank Gopal Basak, Victor DeMiguel, and Gianluca De Nard as well as the participants of the New Zealand Finance Meeting 2021, the 27th Annual Meeting of the German Finance Association, the 40th International Conference of the French Finance Association, the 31st Annual Global Finance Conference, and the ULME research seminar at University Ulm for valuable comments and suggestions. A previous version of the paper circulated under the title "Diversifying estimation errors: An efficient averaging rule for portfolio optimization."

†Swiss Institute of Banking and Finance (s/bf), University of St. Gallen, Unterer Graben 21, 9000 St.Gallen, Switzerland; Center for Real Estate and Environmental Economics, NTNU Business School, Trondheim, Norway; Swiss Finance Institute (SFI), Geneva, Switzerland, Phone: +41 (0)71 224-7055; E-mail: roland.fuess@unisg.ch.

‡Wiesbaden Business School, Hochschule RheinMain, 65183 Wiesbaden, Germany; Tel.: +49 (0)611 94953145, E-mail address: thorsten.glueck@hs-rm.de.

§University of St.Gallen, Unterer Graben 21, 9000 St. Gallen, Switzerland; Phone: +41 (0)76 38 30 757, E-mail: christian.koeppel@live.com.

¶University of Applied Sciences Cologne, Claudiusstr. 1, 50678 Cologne, Germany; Phone: +49 (0)221 82753792; E-mail: felix.miebs@th-koeln.de.

An averaging framework for minimum-variance portfolios: Optimal rules for combining portfolio weights

Abstract

We propose an averaging framework for combining minimum-variance strategies to either minimize the expected out-of-sample variance or maximize the expected out-of-sample Sharpe ratio. Our framework overcomes the problem of selecting the "best" strategy ex-ante by optimally averaging over portfolio weights. This averaging procedure has an intuitive economic interpretation because it resembles a fund-of-fund approach, where each minimum-variance strategy represents a single fund. In a range of simulations, for a set of well-established strategies, we show that optimally averaging over portfolio weights improves the out-of-sample variance and Sharpe ratio. We confirm the finding of our simulation study on empirical data.

JEL Classifications: G11

Keywords: Averaging; diversification; estimation error; portfolio optimization; shrinkage.

1 Introduction

The literature on portfolio selection offer a broad range of approaches to alleviate the impact of estimation errors on out-of-sample portfolio performance. Nevertheless, DeMiguel et al. (2009) find that no single portfolio strategy consistently outperforms the 1/N rule. Stimulated by this observation, several authors suggested improved estimates of the input parameters on the covariance (see, e.g., Fan et al., 2013; Ledoit and Wolf, 2015, 2017, 2020), inverse covariance (see, e.g., Kourtis et al., 2012; DeMiguel et al., 2013; Ledoit and Wolf, 2018; Shi et al., 2020), and portfolio weight level (see, e.g., Kan and Zhou, 2007; Tu and Zhou, 2011; DeMiguel et al., 2013). Further contributions propose extensions of the optimization problem via norm constraints to regularize portfolio weights (see, e.g., Brodie et al., 2009; DeMiguel et al., 2009; Xing et al., 2014; Li, 2015; Yen, 2016). Unfortunately, the plethora of evolving strategies challenges practitioners and academics alike to select one of the proposed strategies. To overcome this problem, we suggest not selecting a single "best" strategy but instead simultaneously considering a set of strategies. This can be achieved by linearly combining the strategies' portfolio weights. Economically, averaging resembles a fund-of-fund structure, with each strategy representing a fund.

Following Jagannathan and Ma (2003) and DeMiguel et al. (2009), we also focus on minimum-variance portfolios with different estimators of the covariance matrix. In particular, we propose a framework that can exploit the correlation structure between the different minimum-variance portfolios such that their linear combination achieves an improved out-of-sample performance. Within our framework, an averaging rule, that is, the weights used to linearly combine the strategies, is obtained by utilizing adjusted versions of the Basak et al. (2009b) jackknife-type estimator of tracking error variance. These estimators serve as optimization criteria, which are either minimized or maximized to determine the averaging weights. In this paper, we derive three averaging rules. The first rule minimizes the expected out-of-sample portfolio variance. The corresponding estimator of the expected out-of-sample portfolio variance is similar to that in Basak et al. (2009b). The second and third averaging rules maximize the expected out-of-sample

ex-ante and ex-post Sharpe ratios, respectively. The former is defined as the portfolio's expected out-of-sample return divided by the its expected out-of-sample standard deviation, whereas the latter is given by the expectation of out-of-sample returns divided by its out-of-sample standard deviation.

To illustrate our framework, we focus on minimum-variance strategies. Our setup follows the idea of linear shrinkage, where an asymptotically unbiased estimator is blended with a structured estimator. In this spirit, we combine the minimum-variance portfolio based on either (i) the sample covariance or (ii) the nonlinear shrinkage estimator of Ledoit and Wolf (2015), respectively, with a multitude of established structured covariance estimators.

In extensive simulations, we generate asset returns, which allows us to construct portfolio returns of the different minimum-variance strategies. The order of magnitude of correlations between these portfolio returns is similar to that of empirical asset returns, which, in turn, suggests that combining portfolio strategies offers diversification benefits. These diversification benefits are persistent between the considered strategies and estimation windows, ranging from six months up to 200 years. The observed correlation structure of portfolio returns is driven by estimation errors in portfolio weights. This suggests that the benefit of averaging increases the more distinct strategies are taken into account.

Evaluating our averaging rules on simulated data shows that, for all estimation windows, our variance-minimizing rule achieves a lower or similar standard deviation than the respective baseline strategies, that is, the minimum-variance portfolios based on the sample and nonlinear shrinkage estimators. For our Sharpe ratio-maximizing averaging rules, we observe consistently higher Sharpe ratios than the respective baseline strategies. Comparing the results for averaging with the sample and nonlinear shrinkage estimator, we find that averaging with the more sophisticated nonlinear shrinkage estimator results in more favorable results with respect to both variance minimization and Sharpe ratio maximization. In relation to competing approaches from the portfolio optimization literature, we find that averaging with the nonlinear shrinkage estimator delivers lower

(higher) standard deviations (Sharpe ratios) than any considered approach.

The evaluation of our averaging rule on empirical data sets broadly confirms the findings from the simulation study. Across all data sets, our variance-minimizing averaging rule achieves a statistically significantly lower standard deviation than the respective baseline strategies. The two Sharpe ratio-maximizing averaging rules compare favorably to the respective baseline strategies, with the ex-post Sharpe ratio rule yielding stronger results for both baseline strategies. Averaging with the ex-post Sharpe ratio rule improves the annualized out-of-sample Sharpe ratio of the sample (nonlinear shrinkage) strategy on each data set by at least 20% (8%) or 0.095 (0.049) in absolute terms. Accounting for transaction costs improves the Sharpe ratio of our averaging rules because of a lower turnover. The increase of the Sharpe ratio on each data set amounts to at least 36% (22%) or 0.134 (0.096) in absolute terms for the averaging rule, including the sample (nonlinear shrinkage) estimator. We find that the improvements are statistically significant, except for the nonlinear shrinkage estimator on smaller data sets.

Our paper relates to prior contributions proposing combinations of portfolio weights or input parameters to reduce estimation errors. Tu and Zhou (2011) show that combining portfolio weights of established mean-variance strategies with the 1/N rule improves the out-of-sample performance of the respective strategies. Kourtis et al. (2012) find that a linear combination of the sample and two structured estimators on the inverse covariance level improves the out-of-sample performance of minimum-variance portfolios. Lancewicki and Aladjem (2014) propose a combination of multiple targets for the shrinkage estimation of covariance matrices. Bonaccolto and Paterlini (2020) provide optimal combining portfolio weights of five portfolio strategies by comparing different choices of optimized utility functions. Based on a machine learning approach, Kazak and Pohlmeier (2023) derive portfolio weights from a set of standalone strategies guided by pseudo out-of-sample portfolio pretesting. Their empirical results confirm that their approach not only yields better portfolio performance but also reduces outliers, and thus, average turnover costs. Overall, the results of the aforementioned articles suggest that estimation errors may further be reduced if additional resources for reducing estimation

errors are included. We follow this path by exploring whether estimation errors can be diversified more comprehensively if multiple resources for the diversification are taken into account. Contrary to these prior contributions, our paper is the first to provide an averaging framework for minimum-variance portfolios that allows the estimation of asymptotically consistent averaging weights.

We contribute to the literature in two ways: First, we provide a new framework for combining strategies that (*i*) builds on established methods for estimation and optimization, (*ii*) is easy to implement, and (*iii*) can be extended to further—possibly novel—strategies not selected in this paper. Second, we provide simulation-based and empirical evidence that our approach improves the performance; this is done by using a set of well-established minimum-variance strategies. In particular, our approach compares favorably to various selected benchmarks in terms of out-of-sample standard deviation and Sharpe ratio.

This paper is organized as follows: Section 2 introduces our framework and outlines its implementation in this paper. Section 3 describes the simulation procedure and explores the behavior of the averaging rule using simulated data. Section 4 evaluates the out-of-sample performance of our rule on five empirical data sets. Section 5 concludes.

2 Model diversification: An averaging approach

2.1 Averaging Framework

Throughout the paper, we focus on minimum-variance strategies resulting from the well-known optimization problem $\min_\omega \omega'\Sigma\omega$ subject to $\omega' 1_N = 1$, where Σ is the true but unknown covariance matrix, 1_N is a N-dimensional vector of ones, and N is the number of investable risky assets, respectively. Allocation to these N assets within the minimum-variance portfolio results from solving this optimization problem with optimal portfolio weights given by

$$\omega = \Sigma^{-1} 1_N (1_N' \Sigma^{-1} 1_N)^{-1}. \tag{1}$$

Because ω depends on the true but unknown covariance matrix, Σ has to be replaced by a suitable estimator: $\hat{\Sigma}$. A natural estimator for the covariance matrix is the sample covariance matrix. Nevertheless, the estimation error associated with the sample covariance matrix can be substantial (see, e.g., Ledoit and Wolf, 2003; DeMiguel et al., 2009; Ledoit and Wolf, 2020). Any estimation errors in $\hat{\Sigma}$ translate into suboptimal allocations that have an adverse effect on portfolio performance. The literature proposes different approaches to mitigate the impact of estimation errors on minimum-variance portfolio weights and out-of-sample performance. The first branch of the literature suggests structured estimators of the covariance matrix. Examples include the constant correlation and (approximate) factor models. The second branch proposes shrinkage estimators of either the covariance matrix or its inverse and portfolio weights. Finally, the third branch regularizes portfolio weights through the imposition of weight constraints such as short-sale constraints or norm constraints.[1]

All approaches aim to determine the single best strategy to alleviate the impact of estimation errors. The wide variety of approaches leaves investors with the question of which suggested strategy to follow. Without any ex-ante knowledge about the "best" approach, investors face the problem of picking a single strategy. In contrast to choosing one single strategy, we advocate that investors should diversify over various strategies to mitigate estimation errors. Therefore, let $W = [w^{(1)}, w^{(2)}, ..., w^{(M)}]$ be the $N \times M$ matrix comprising the portfolio weights of M distinct strategies, that is, $w^{(m)}$ is the $N \times 1$ vector with minimum-variance weights as defined by Equation (1), where Σ is replaced by the m-th strategy-specific estimator $\hat{\Sigma}^{(m)}$. We then define an averaging rule as a $M \times 1$ vector β satisfying $\beta \geq 0$ and $1'_M \beta = 1$. The averaging portfolio weights implied by this rule are $\overline{w} = \beta' W$. Given an objective function $f(\beta)$ that shall be minimized, the averaging

[1]The constant correlation model is provided in Elton and Gruber (1973). For factor models see Chan et al. (1999) and Fan et al. (2013). Shrinkage-based estimators are developed in, e.g., Ledoit and Wolf (2003, 2004a,b, 2015, 2017, 2018, 2020), Kourtis et al. (2012), DeMiguel et al. (2013), Shi et al. (2020), Kan and Zhou (2007), and Tu and Zhou (2011). Weight constraints can be found in Frost and Savarino (1986), Jagannathan and Ma (2003), Brodie et al. (2009), DeMiguel et al. (2009), Xing et al. (2014), Li (2015), and Yen (2016).

rule can be determined by[2]

$$\begin{aligned} \beta^* &= \arg\min_{\beta} f(\beta) && (2) \\ \text{s.t.} \quad & 1_M'\beta = 1 && (3) \\ & \beta \geq 0, && (4) \end{aligned}$$

In the remainder of this section, we suggest three objective functions to determine the averaging weights β. Because each objective function is based on an adjusted version of the Basak et al. (2009b) jackknife-type estimator of (tracking error) variance, we follow the authors and consider throughout the paper an out-of-sample period of one month, that is, the period from $T+1$ until $T+L$, with T being the number of observations within the in-sample period and L denoting the number of out-of-sample observations per month, that is, $L = 21$ on a daily data frequency. To simplify the notation, we assume that the in-sample period can be split into $J = T/L$ equally sized blocks, such that block $J+1$ refers to the out-of-sample period.

2.1.1 Out-of-sample variance

For our first averaging rule, we define an estimator for the expectation of the out-of-sample variance as a function of the averaging rule β. To this end, we first note that the block $J+1$ variance of a minimum-variance strategy with portfolio weights w is given by

$$s_{J+1}(w) = w' S_{J+1} w, \tag{5}$$

where S_{J+1} denotes the covariance matrix of the out-of-sample period. Let $H_T = \mathbb{E}_T[S_{J+1}]$ denote the expectation of S_{J+1} conditional on the information set up to T. The corresponding conditional expectation of the out-of-sample variance induced by strategy

[2]The matrix W is omitted intentionally in the objective function. The reason will become clear in the following two sections 2.1.1 and 2.1.2, where we define such functions for the expected out-of-sample portfolio variance minimization and expected out-of-sample ex-ante and ex-post Sharpe ratio maximization.

w is then given by:

$$\mathbb{E}_T\left[s_{J+1}(w)\right] = w' H_T w. \tag{6}$$

Basak et al. (2009a) suggest a jackknife-type estimator for Equation (6). The first step of this jackknife procedure requires dropping the excess returns falling into the j-th month from the in-sample period. The second step, consists of estimating the minimum-variance portfolio weights based on the remaining $J-1$ blocks with a total of $T-L$ returns. We denote this estimate by w_{-j}. The third step comprises the computation of the jackknife excess returns for the j-th month, k_j. The latter are the portfolio returns that would have been achieved from holding w_{-j} in the hold-out month j, that is, $k_j = [r_{(j-1)L+1}w_{-j}, ..., r_{jL}w_{-j}]$, with $r_{(j-1)L+1}, ..., r_{jL}$ being the $1 \times N$ vectors of excess asset returns in hold-out month j. We denote the variance of k_j by

$$q_j = w'_{-j} S_j w_{-j}, \tag{7}$$

with covariance matrix

$$S_j = \frac{1}{L-1} \sum_{i=1}^{L} \left[r_{(j-1)L+i} - \bar{r}_j\right]' \left[r_{(j-1)L+i} - \bar{r}_j\right], \tag{8}$$

and average excess returns $\bar{r}_j = 1/L \sum_{i=1}^{L} r_{(j-1)L+i}$. Repeating these steps across all J months within the in-sample period yields the sequence $(q_1, q_2, ..., q_J)$. The idea behind the jackknife procedure is that each k_j has approximately the same distribution as the out-of-sample portfolio returns induced by w. Hence, an average over the sequence $(q_1, q_2, ..., q_J)$ serves as an estimate for the conditional expectation of the out-of-sample variance in Equation (6):

$$q_{J+1} = \sum_{j=1}^{J} \alpha_j q_j, \tag{9}$$

with α_j being a weighting factor, that is, $\alpha_j \geq 0$ and $\sum_{j=1}^{J} a_j = 1$. In case of independently and identically distributed (i.i.d) returns over time, Basak et al. (2009a) set $\alpha_j = 1/J$ for all j, such that the formula for q_{J+1} collapses to the arithmetic mean. If returns encounter time-varying second moments, Basak et al. (2009a) suggest that α_j follows an

exponential weighting scheme, assigning a higher weight to more recent observations.

Basak et al. (2009a) proof the consistency of q_{J+1} for minimum-variance portfolios based on the sample covariance matrix. In section A.2 of our online appendix, we show that their proof also holds if single strategy weights w are replaced by some averaging weights $\overline{w}$; hence, q_{J+1} consistently estimates $\mathbb{E}_T\left[s_{J+1}(\overline{w})\right]$. Obviously, in this case, Equation (9) can be written as a function of the averaging rule β, that is,

$$q_{J+1}(\beta) = \sum_{j=1}^{J} \alpha_j q_j(\beta) \tag{10}$$

where $q_j(\beta) = \left[\beta' W_{-j}\right]' S_j \left[\beta' W_{-j}\right]$ is the variance of portfolio returns k_j induced by averaging weights $\beta W_{-j} = \overline{w}_j$. These weights are obtained by removing block j from the sample. For a reasonably large sample, we can determine the optimal averaging rule β^*_{VAR} by setting $f(\beta) = q_{J+1}(\beta)$ in Equation (2).

We note that Equation (10) also has an appealing economic interpretation. To show this, we rewrite $q_{J+1}(\beta) = \sum_{j=1}^{J} \beta' \Omega_j \beta$ with $M \times M$ matrix $\Omega_j = \alpha_j W'_{-j} S_j W_{-j}$ by representing the (exponentially) weighted covariance matrix of jackknife returns for the M minimum-variance strategies in the respective hold-out month j. Thus, the averaging rule β^*_{VAR} minimizes the average variance in the J hold-out months. Given the constraints in Equations (3) and (4), our rule resembles a minimum-variance portfolio optimization problem with short-sale constraints. Hence, we can interpret our averaging rule as an *out-of-sample* variance-minimizing fund-of-fund strategy that utilizes the single minimum-variance strategies as underlying portfolios.

2.1.2 Out-of-sample Sharpe ratio

The objective when averaging over portfolio strategies may not only be confined to variance minimization but may also include expected returns. Thus, in the following two objective functions, we suggest determining the averaging rules that aim at maximizing the risk-return trade-off as measured by the Sharpe ratio.[3]

[3]Both metrics have recently been used in the portfolio optimization literature. For example, Demiguel et al. (2014) use (11) as a calibration criterion to compute optimal shrinkage and intensities whereas

The first objective function is an estimator for the ex-ante Sharpe ratio SRA. For a minimum-variance strategy with portfolio weights w, this ratio is defined as the expected excess return over the expected standard deviation of excess returns in the next period, that is,

$$SRA = \mathbb{E}_T\left[\overline{r}_{J+1}(w)\right] / \sqrt{\mathbb{E}_T\left[s_{J+1}(w)\right]}, \tag{11}$$

where $\mathbb{E}_t[\overline{r}_{J+1}(w)]$ denotes the portfolio's expected return in the out-of-sample period. The components of SRA are, in the case of averaging weights, $\mathbb{E}_T\left[\overline{r}_{J+1}(\overline{w})\right]$ and $\mathbb{E}_T\left[s_{J+1}(\overline{w})\right]$. The estimator for the latter is defined in Equation (10). For the former, we propose following jackknife-type estimator as a function of β:

$$k_{J+1}(\beta) = \sum_{j=1}^{J} \alpha_j k_j(\beta), \tag{12}$$

where $k_j(\beta) = \bar{r}_j'\left(\beta' W_j\right)$ is the average excess return that would have been achieved from holding $\bar{w}_{-j}$ in the hold-out month j. Again, the proof for the consistency of k_{J+1} is relegated to section A.2 of our online appendix. Substituting q_{J+1} and k_{J+1} for $\mathbb{E}_T\left[\overline{r}_{J+1}(\overline{w})\right]$ and $\mathbb{E}_T\left[s_{J+1}(\overline{w})\right]$ in Equation (11) yields our estimator for the ex-ante Sharpe ratio:[4]

$$\widehat{SRA}(\beta) = k_{J+1}(\beta) / \sqrt{q_{J+1}(\beta)} \tag{13}$$

As a result, we set $f(\beta) = -\widehat{SRA}(\beta)$ in Equation (2) to determine the ex-ante Sharpe ratio-maximizing averaging rule β^*_{SRA}. The second objective function is an estimator for the expected ex-post Sharpe ratio SRP, which is given by the expected value of the realized Sharpe ratio in the out-of-sample period, that is,

$$SRP = \mathbb{E}_T\left[\overline{r}_{J+1}(w) / \sqrt{s_{J+1}(w)}\right]. \tag{14}$$

An analytically tractable estimator for the expected ex-post Sharpe ratio is, to the best of our knowledge, not available.[5] However, our simulation results in section A.3 of the

Kircher and Rösch (2021) use (14) to determine bootstrap-based portfolio weights.

[4] Note that consistency of this estimator follows from the continuous mapping theorem.

[5] DeMiguel et al. (2013) come to the same conclusion regarding tractability when determining the shrinkage intensities for different calibration criteria in the context of portfolio optimization.

Internet Appendix indicate that following the estimator of the ex-post Sharpe ratio is reasonable:

$$\widehat{SRP}(\beta) = \sum_{j=1}^{J} \alpha_j \left[k_j(\beta) / \sqrt{q_j(\beta)} \right]. \tag{15}$$

Again, we set $f(\beta) = -\widehat{SRP}(\beta)$ in Equation (2) to obtain the ex-ante Sharpe ratio-maximizing averaging rule β^*_{SRP}.

2.2 Implementation

We demonstrate our framework by focusing on minimum-variance portfolios, following common practice in the literature (see, e.g., Jagannathan and Ma, 2003; DeMiguel et al., 2009). The setup for the implementation of our averaging rules follows the idea of linear shrinkage, where an asymptotically unbiased estimator is blended with a structured estimator. In this spirit, we combine the minimum-variance portfolios based on sample and nonlinear shrinkage estimators, respectively, with portfolios based on a multitude of structured estimators that follow our three rules. The weights of each minimum-variance portfolio come from Equation (1) and differ only by the estimator of the covariance matrix, $\hat{\Sigma}$. In our averaging procedure, we consider those minimum-variance portfolios resulting from the following covariance matrix estimators:[6]

1. The (unbiased) sample covariance matrix, $\hat{\Sigma}_S$.

2. A nonlinear shrinkage estimator based on eigenvalues determined via large-dimensional asymptotics, that is, $N/T \to c$ as $T \to \infty$, following Ledoit and Wolf (2015), $\hat{\Sigma}_{LWNLS}$.

3. The covariance matrix implied by the single-factor model of Sharpe (1963), $\hat{\Sigma}_{1F}$.

4. The covariance matrix recovered from a constant correlation matrix with the average of all pairwise correlations off the main diagonal and assets' individual standard deviations on the main diagonal, $\hat{\Sigma}_{CC}$ (Elton and Gruber, 1973).

[6]We provide a comprehensive review of the estimators in Section A.1 of the Internet Appendix.

5. A diagonal matrix with the average variance of all assets on the main diagonal, $\hat{\Sigma}_{ID}$. The estimator was suggested by Ledoit and Wolf (2004a) as a shrinkage target. Note that the resulting minimum-variance portfolio corresponds to the 1/N portfolio, which represents the benchmark in terms of risk-adjusted performance (DeMiguel et al., 2009).

6. The covariance matrix corresponding to the short-sale constrained minimum-variance portfolio from Jagannathan and Ma (2003), hence implying non-negative portfolio weights, $\hat{\Sigma}_{SC}$.

Our choice of portfolio strategies considered in the averaging process is motivated as follows: first, we include only portfolio strategies that are, by construction, unlikely to be perfectly correlated. A perfect correlation of portfolio strategies offers, on the one hand, no diversification potential, making averaging redundant. On the other hand, a perfect correlation induces numerical problems in the averaging procedure because of multicollinearity issues. As a result, we only consider estimators that have distinct information about the dependence of asset returns. Second, we include minimum-variance portfolios based on estimators 3. to 6. because these strategies are well established in the portfolio optimization literature. Additionally, the aforementioned covariance matrix estimators play a vital role in the linear shrinkage literature. Third, we aim to demonstrate the flexibility of our averaging framework and include with the short-sale constrained minimum-variance portfolio an analytically intractable portfolio strategy. Possibly because of its intractability, the short-sale constrained minimum-variance portfolio has so far not been considered for optimal combinations with other portfolio strategies.[7]

The implementation of our averaging framework gives rise to six averaging portfolios for each jackknife weighting scheme, that is, equally and exponential weighting. We abbreviate the expected out-of-sample variance-minimizing combination of the minimum-variance portfolios based on $\hat{\Sigma}_S$ and the aforementioned structured estimators based

[7] DeMiguel et al. (2009) consider a combination of the short-sale constrained minimum-variance portfolio and the $1/N$ portfolio. However, the authors set the combination intensity arbitrarily, not following an optimality criterion.

on the equally weighted jackknife by *AV-Sample-Var*. The ex-ante and ex-post Sharpe ratio-maximizing averaging portfolios are abbreviated by *AV-Sample-SR ex-ante* and *AV-Sample-SR ex-post*, respectively. In case of the *exponentially* weighted jackknife, we refer to the aforementioned averaging portfolios by *AV-Sample-Var-E*, *AV-Sample-SR ex-ante-E*, and *AV-Sample-SR ex-post-E*. The abbreviations for the averaging portfolios combining the minimum-variance portfolios corresponding to LWNLS and the structured estimators follows analogously.

Table 1 provides an overview of our averaging portfolios in Panel A and the minimum-variance portfolios considered in the averaging procedure, to which we refer in the following as strategies, in Panel B. Panel C lists minimum-variance portfolios from the literature, to which we benchmark our results.[8]

[TABLE 1 ABOUT HERE]

3 Simulation study

3.1 Simulation procedure

The data-generating process (DGP) of our simulation study follows Tu and Zhou (2011). We consider $N = 25$ assets with the simulated daily excess return of security i at time t coming from a Fama-French three-factor model with mispricing, that is,

$$r_{i,t} = b_{i,0} + b_{i,1} r_{MKT,t} + b_{i,2} r_{SMB,t} + b_{i,3} r_{HML,t} + u_{i,t}, \tag{16}$$

where $b_{i,0}$ is the mispricing factor, $b_{i,1}$, $b_{i,2}$ and $b_{i,3}$ are the factor loadings and $u_{i,t}$ is the residual at time t.

We calibrate the data-generating process in (16) on the daily returns of the 25 Fama-French portfolios formed on size and book-to-market and the three Fama-French

[8] We provide a comprehensive review of the benchmark strategies in Section A.1 of the Internet Appendix.

factors over the period from June 1973 until June 2019.[9] The assets' factor loadings are randomly paired and evenly spread between the respective lower and upper empirical parameter estimates, that is, in each simulation run $b_{i,1} \in [0.85, 1.15]$, $b_{i,2} \in [-0.30, 1.10]$, and $b_{i,3} \in [-0.40, 1.10]$. The annualized mispricing factor $b_{i,0} \in [-5\%, 5\%]$. The premia of the market $r_{MKT,t}$, size $r_{SMB,t}$, and value factor $r_{HML,t}$ at t follow a multivariate normal distribution with the annualized means and standard deviations of the factors, as well as the correlation matrix between the factors reported in Table 2.

[TABLE 2 ABOUT HERE]

The error term of each asset at t, $u_{i,t}$ comes from a multivariate normal distribution with $N \sim (0, \Sigma_u)$. Again, we follow Tu and Zhou (2011) and assume that the covariance matrix Σ_u is diagonal. The residuals' annualized volatility is drawn from a uniform distribution with a lower bound of 3.50% and upper bound of 8.50%, which again matches the empirical values over the calibration horizon.

We run 1,000 simulations. In each run k, we draw $\tau = \{0.5, 1, 5, 10, 25, 50, 100, 200\}$ years of daily observations from the outlined data-generating process, with each year comprising 252 returns. We then estimate the portfolio weights of the m-th portfolio strategy in the k-th run, $\mathrm{w}_k^{(m)}$, which is based on the $T = \tau \cdot 252$ simulated excess returns.

3.2 Simulation results

We begin our analysis motivating the diversification potential of portfolio strategies that we aim to leverage with our averaging rules. To this end, we investigate the correlation between the returns of the strategies considered in the averaging rules. We compute the returns of the m-th portfolio strategy for a given estimation window in the k-th simulation run as $r_k^{(k)} = \left(\mathrm{w}_k^{(m)}\right)' r_k$, where r_k denotes the $T \times N$ matrix of the simulated returns in run k. We gather the aforementioned returns for all K simulation runs in $R^{(m)}$ by stacking returns, such that $R^{(m)} = \left[r_1^{(m)}; r_2^{(m)}; \ldots, r_K^{(m)}\right]$.

[9] Data comes from Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[TABLE 3 ABOUT HERE]

Panel A of Table 3 reports the correlation between $R^{(m)}$ for all strategies considered in our averaging rules based on an estimation window of one year. The correlation between Sample (LWNLS) and minimum-variance portfolios based on the structured estimators 1F, CC, ID, and SC ranges between 0.618 (0.675) for ID and 0.828 (0.860) for 1F. Overall, we observe that those portfolio strategies with a somewhat similar structure exhibit higher correlations and vice versa. For example, we find that the strategies without short-sales, that is, ID and SC, have a correlation of 0.975. Contrary to this, we observe that the correlation between ID and strategies, allowing for a more data-driven structure (1F and CC) or the Sample strategy, is lower with values between 0.618 and 0.749. To put the reported correlations into perspective, we investigate the correlation between the returns of the 25 Fama-French portfolios formed on size and book-to-market over our calibration period. We find that the correlation between the portfolios varies between 0.611 and 0.955, with the average correlation amounting to 0.826. These correlations are in the same order of magnitude as the ones reported in Panel A of Table 3. This suggests—at least for the setup of the simulation—that the diversification potential between the considered strategies is similar to that on the asset level.

Investigating the source behind the observed return correlations, we turn to the portfolio weights of the considered strategies. We inspect the difference between the portfolio weights of any two strategies by focusing on the correlation of estimation errors defined for the k-th simulation run and m-th strategy as

$$\Delta w_k^{(m)} = w_k^{(m)} - \omega_k. \tag{17}$$

and then stack these vectors for each strategy to obtain $\Delta W^{(m)} = \left[\Delta w_1^{(m)}, \Delta w_2^{(m)}, \ldots, \Delta w_K^{(m)}\right]$. The correlation between $\Delta W^{(m)}$ and $\Delta W^{(n)}$ is the metric to which we refer in the following as estimation error correlation. Panel B of Table 3 shows the respective results. The correlations range between 0.003 (Sample and ID) and 0.950 (SC and ID). Overall, we observe that the average correlation between estimation errors amounts to 0.527, suggest-

ing that combining portfolio strategies may alleviate the estimation errors in portfolio weights.

[Figure 1 ABOUT HERE]

Because all portfolio strategies are evaluated on the identical set of simulated returns, any imperfect correlation between strategy returns are mirrored in the respective portfolio weights. This link between the correlations of strategy returns (Panel A) and correlations of estimation errors (Panel B) is illustrated in Figure 1, which plots the former against the latter. The pattern shows that the correlation between portfolio returns is lower (higher) when the correlation between the respective estimation errors is lower (higher).

In further results, we investigate the persistence of the correlation between our considered strategies. To this end, we inspect the correlation between returns, as well as estimation errors, based on estimation windows of $\tau = \{0.5, 1, 5, 10, 25, 50, 100, 200\}$ years, as reported in Tables A.3.1 and A.3.2 in the Internet Appendix. We observe that all correlations remain remarkably stable. For example, we find that the correlation between Sample (LWNLS) and 1F increases gradually from 0.828 (0.860) for an estimation window of one year to 0.876 (0.876) for an estimation window of 200 years. The changes in correlation between the remaining strategy pairs is roughly of the same order.[10]

We compare the out-of-sample standard deviations of our averaging rules against the optimum and competing strategies in Table 4. Optimum (first row) is based on averaging weight vector β as determined by the optimization problem in Equations (2) to (4) using the true parameters of the DGP. In case of variance minimization, the objective $f(\beta)$ is the out-of-sample variance. If the parameters are known, Optimum is in this case always the true covariance matrix, which is represented by the sample strategy. For estimation windows of up to five years, our averaging rule AV-Sample-VAR achieves smaller standard deviations than the Sample strategy. The standard deviations of AV-

[10]We note that the correlation between Sample and LWNLS increases from 0.986 for a one-year estimation window to 1.000. Note that the precise value does not imply a perfect correlation as reported but rather stems from rounding to three decimal places.

Sample-VAR and Sample are identical for larger estimation windows. Turning to AV-LWNLS-VAR compared with LWNLS, we find smaller standard deviations for LWNLS in case of estimation windows of up to one year. For larger estimation windows, we again observe identical standard deviations. Compared with the competing benchmark strategies from the literature (LW1F, LWID, LWCC, and BBCVC), we find that AV-LWNLS-VAR achieves lower standard deviations for estimation windows of one year and less.

Overall, we find that the differences between the standard deviations of our averaging rules and baseline strategies Sample and LWNLS, as well as the competing benchmark strategies from the literature, are small, particularly for larger estimation windows. In contrast to the simulation results, we observe more pronounced differences in favor of our averaging rules on the empirical data sets, which we discuss in detail in section 4.2.

[TABLE 4 ABOUT HERE]

Table 5 reports the risk-adjusted performance of our rules compared with the established strategies. Again, Optimum refers to the optimal strategy combination. Here, we combine the strategies based on the known parameters of the DGP such that the Sharpe ratio is maximized.[11] Most importantly, we observe that both the ex-ante and the ex-post Sharpe ratio combinations using Sample and LWNLS as a baseline strategy approach the Sharpe ratio of Optimum for larger estimation windows. Beginning with the smallest estimation windows of 0.5 years, we observe that our averaging rules achieve Sharpe ratios of 0.54 (AV-Sample-SRA, AV-Sample-SRP, and AV-LWNLS-SRP) and 0.55 (AV-LWNLS-SRA), which is well below the 0.71 of Optimum. Additionally, the respective Sharpe ratio of the data agnostic ID strategy in Panel B is 0.57, and thus, larger than that of our averaging rules. For a five-year estimation window, our rules achieve Sharpe ratios in the range of 0.64 and 0.65, which lies between the Sharpe ratios of ID and Optimum. The difference to the Sharpe ratio of Optimum (ID) narrows (widens) for estimation windows of ten years and more. In contrast to the remaining single strategies

[11] There is only one optimum because Ex-ante and Ex-post Sharpe ratios are identical if the parameters are known.

in Panel B, we observe that the ex-post and ex-ante Sharpe ratio rules with Sample as well as LWNLS achieve consistently higher Sharpe ratios than any other strategy. With respect to the benchmarks in Panel C, we find that BBCVC obtains a higher Sharpe ratio than our averaging rules for an estimation window of 0.5 years. However, for estimation windows of one year and more, our Sharpe ratio-maximizing rules outperform all benchmarks.

[TABLE 5 ABOUT HERE]

Our simulation study has three key results. First, the returns and estimation errors of portfolio strategies considered in this paper are not perfectly correlated. This suggests that appropriately combining these strategies can leverage diversification benefits. Second, our variance-minimizing rules are at least as good as the best single strategy, with only one exception. Third, our Sharpe ratio-maximizing rules outperforms any other single strategy and benchmark for larger estimation windows.

4 Empirical study

4.1 Empirical setup

We evaluate the out-of-sample performance of our averaging rule in comparison to the six strategies and four benchmarks on different U.S. single stock investment universes. Starting in 1975, we identify in June of each year the largest $N = \{30; 50; 100; 150; 200\}$ stocks by market cap from all NYSE, AMEX, and NASDAQ stocks with a share code of 10 or 11 in the Center for Research in Security Prices (CRSP) database and that fulfill the following criteria: We filter out stocks with a price of less than USD 5, as well as stocks with missing excess returns in the 252 trading days preceding or the 252 days subsequent to the selection date.[12] The selected stocks constitute the investment universe for one year, which is then revised based on the outlined criteria. For the stocks

[12]We follow the convention that a calendar month comprises 21 trading days.

in the constructed investment universes, we obtain daily data from the CRSP database and compute the respective excess returns over the 30-day T-bill rate.

We obtain the out-of-sample excess returns for each portfolio from Table 1, that is, the averaging rules, strategies, and benchmarks, by using the common rolling-sample procedure, setting the estimation window to $\tau = 252$ observations. Starting in June 1976, we estimate the portfolio weights of each portfolio using only the information in the estimation window comprising the most recent τ excess returns. We hold the estimated portfolio weights constant for 21 trading days, representing a one-month holding period, and save the corresponding out-of-sample excess returns. We then move the estimation window forward by 21 trading days and repeat the aforementioned procedure over the entire sample period.

We compare our averaging rules with each competing portfolio in terms of standard deviation (s) and Sharpe ratio (SR) for out-of-sample excess returns. The respective metrics for the p-th portfolio are given by:

$$s_p^2 = \frac{1}{T-\tau-1} \sum_{t=\tau}^{T-1} \left[w_{p,t}' r_{t+1} - \bar{r}_p \right]^2, \tag{18}$$

$$SR_p = \frac{\bar{r}_p}{s_p}, \tag{19}$$

where $\bar{r}_p = 1/(T-\tau) \sum_{t=\tau}^{T-1} w_{p,t}' r_{t+1}$ is the average out-of-sample excess return,[13] $w_{p,t}$ are the weights of portfolio p at time t, and T is the total number of observations in our sample period from June 1975 to June 2019. In addition to the outlined performance analysis, we evaluate the viability of our averaging rule compared with other portfolios by accounting for transaction costs. To this end, we compute the excess returns net of transaction costs for the p-th portfolio at $t+1$ as follows:

$$\underline{r}_{p,t+1} = w_{p,t}' r_{t+1} - \kappa \left|\left| w_{p,t+1} - w_{p,t+} \right|\right|_1, \tag{20}$$

where w_{p,t^+} are the weights of strategy i at the end of t before rebalancing, $w_{p,t+1}$ are the

[13] We use the 30-day T-bill rate from Kenneth French's website to proxy the risk-free rate.

congruent weights after rebalancing at time $t+1$, $||\cdot||_1$ is the one-norm, that is, the sum of absolute values, and κ quantifies the one-way transaction costs. We then reevaluate the risk-return characteristic of each portfolio strategy by computing the Sharpe ratio according to Equation (19), with $\bar{r}_p$ and s_p being substituted by the arithmetic mean and standard deviation of the excess returns net of transaction costs for portfolio p.

Following Ledoit and Wolf (2017), we measure the statistical significance of the differences between the out-of-sample standard deviation and Sharpe ratio of our averaging rules and the competing strategies and benchmarks using the HAC inference proposed by Ledoit and Wolf (2008, 2011). Specifically, we test whether our averaging rules achieve a statistically significantly lower (higher) out-of-sample standard deviation (Sharpe ratio) than the respective baseline strategies of the averaging rules, that is, Sample and LWNLS.

We report the results for the averaging rules with equally weighted jackknife returns. The results for the exponentially weighted jackknife returns are reported in Tables A.4.5 to A.4.7 in the Internet Appendix. We find that the performance of the two weighting schemes is qualitatively similar.

4.2 Out-of-sample results

Table 6 reports the annualized out-of-sample standard deviations of our proposed rules, strategies, and benchmarks. The standard deviations of AV-Sample-VAR and AV-LWNLS-VAR in Panel A compare favorably to the baseline strategies of the respective averaging rules. On all data sets, both averaging rules achieve a statistically significantly lower standard deviation than Sample and LWNLS, respectively. Comparing AV-Sample-VAR and AV-LWNLS-VAR, we find that, across all data sets, the latter achieves the lowest standard deviation. Hence, blending the most sophisticated minimum-variance strategy, LWNLS, with additional strategies from the literature further improves the objective of variance minimization.

The standard deviations of our Sharpe ratio rules are substantially higher vis-a-vis the metrics of the respective baseline strategies for most data sets. Interestingly, we find

that AV-Sample-SRA and AV-Sample-SRP achieve for $N = 150$ and $N = 200$ significantly lower standard deviations than Sample. We observe that the standard deviation of Sample increases on the two aforementioned data sets because of the increasing impact of estimation errors. The results for our averaging rules suggest that the impact of estimation errors is alleviated when combining Sample with other strategies, which is consistent with the shrinkage literature. Compared with the strategies (Panel B) and benchmarks (Panel C), we observe that AV-LWNLS-VAR achieves consistently lower standard deviations across all data sets. The improvements are in magnitude higher as those by LWNLS.

Table 7 reports the annualized out-of-sample Sharpe ratios. The Sharpe ratios of AV-Sample-SRA (AV-Sample-SRP) are between 0.059 (0.096) and 0.336 (0.367) higher than the respective metrics for Sample. The Sharpe ratio increases of AV-LNWLS-SRA (AV-LWNLS-SRP) range between 0.008 (0.052) and 0.047 (0.108). Similar to variance-minimizing averaging rules, we observe that averaging with the more sophisticated baseline strategy LWNLS yields more favorable results. Even though the discussed averaging rules achieve consistently higher Sharpe ratios than the respective baseline strategies, we observe only for AV-Sample-SRP and AV-Sample-SRA statistically significant improvements on the investment universes, with $N = \{100, 150\}$ and $N = 100$, respectively. Our subsequent results including transaction costs show that the Sharpe ratio improvements of our averaging rules compared with the respective baseline strategies increase and are statistically significant.

The Sharpe ratios of our variance-minimizing averaging rules are similar or above the Sharpe ratios of the respective baseline strategies. We observe for AV-Sample-VAR (AV-LWNLS-Var) Sharpe ratio increases between 0.042 (0.000) and 0.296 (0.032). The results show that a reduction of the out-of-sample standard deviation does not come at the cost of a reduced risk-adjusted return. Comparing our Sharpe ratio-maximizing averaging rules with the strategies (Panel B) and benchmarks (Panel C), we find that, on all investment universes, AV-LWNLS-SRP achieves higher Sharpe ratios than any competing strategy or benchmark. The improvements to the best competing strategy or

benchmark on each investment universe range between 0.021 and 0.070.

Table 8 reports the annualized out-of-sample Sharpe ratios after transaction costs of 15 bps. We observe that the Sharpe ratio improvements of all averaging rules compared with the respective baseline strategies increase substantially when taking transaction costs into account. This is driven by a lower turnover of our rules compared with the baseline strategies.[14] The Sharpe ratio increases of AV-Sample-SRA are statistically significant, except for the $N = 30$ investment universe, while the respective increases of AV-Sample-SRP and AV-Sample-VAR for all investment universes are statistically significant.

For larger investment universes, our averaging rules using LWNLS as a baseline strategy achieve statistically significant Sharpe ratio improvements compared with LWNLS. Specifically, we observe that the increases of AV-LWNLS-SRP are significant for investment universes of $N = 100$ or more. The respective increases of AV-LWNLS-SRA are only significant in the $N = 100$ investment universe. Notably, we find that AV-LWNLS-VAR achieves significantly higher Sharpe ratios than LWNLS, except for the $N = 30$ investment universe. However, the respective increases are smaller than those of AV-LWNLS-SRA and AV-LWNLS-SRP. AV-LWNLS-VAR achieves a higher Sharpe ratio before transaction costs as well as a lower standard deviation and turnover than LWNLS. Thus, subtracting transaction costs widens the return difference between AV-LWNLS-VAR and LWNLS, while the standard deviation of the respective portfolios remains roughly constant. Hence, the Sharpe ratio difference between AV-LWNLS-VAR and LWNLS comes with a smaller standard error compared with the respective differences between the Sharpe ratio-maximizing averaging rules and LWNLS.

5 Conclusion

In this paper, we propose a novel framework for portfolio optimization that leverages the benefits of averaging multiple portfolio strategies to alleviate the impact of estimation errors on out-of-sample portfolio performance. Our approach builds on the idea of a

[14] We report the average monthly turnover of the averaging rules, strategies, and benchmarks in Table A.4.3 in the Internet Appendix.

fund-of-fund strategy and consists of combining the Sample with a multitude of structured estimators to both variance-minimizing and Sharpe ratio-maximizing rules. The motivation behind this idea is to bring the principle of diversification from the portfolio to the estimation level. If errors in the individual estimators are not perfectly correlated, the combination of the latter offers diversification benefits. Consequently, the resulting estimates of the minimum-variance portfolio weights from our averaging procedure should be less error prone and, on average, deliver improved out-of-sample performance. We implement and test our framework on simulated data as well as in an empirical setting of U.S. stocks, comparing the performance of our averaging rules against baseline strategies and benchmarks.

Our simulation study yields three key insights. First, the returns and estimation errors of portfolio strategies under consideration are not perfectly correlated. This suggests that the diversification potential of appropriately combining these strategies could be harnessed. Second, our variance-minimizing rules demonstrate, with only one exception, comparable or superior performance to the best single strategy across various scenarios. Third, our Sharpe ratio-maximizing rules consistently outperform both single strategies and benchmarks, particularly for larger estimation windows. We attribute the performance gains to the leveraged benefits of estimation error diversification between the considered estimators of our averaging framework.

Our empirical results on U.S. single stocks are in line with the findings from the simulation study showing the efficacy of both variance-minimizing and Sharpe ratio-maximizing averaging rules in improving risk-adjusted returns across diverse investment universes. The reduction in out-of-sample standard deviation achieved by variance-minimizing rules consistently demonstrate lower standard deviations compared with the baseline strategies, confirming their success in minimizing variance without sacrificing risk-adjusted returns. The Sharpe ratio analysis underscores the practical significance of these improvements, with our averaging strategies consistently achieving higher Sharpe ratios, hence emphasizing their superior risk-adjusted performance, especially after accounting for transaction costs. Notably, our nonlinear averaging rules on the ex-post

Sharpe ratio consistently and significantly outperform other strategies and benchmarks, emphasizing the efficacy and robustness of our proposed framework.

In conclusion, our study introduces a robust and flexible portfolio optimization framework demonstrating promising results across simulations and empirical evaluations. By combining multiple strategies, we enhance the risk-return profile of portfolios, showcasing the potential for practical implementation in real-world investment scenarios. Whether focused on minimizing variance or maximizing Sharpe ratios, these rules offer compelling solutions for investors seeking enhanced risk-adjusted returns across different investment environments, particularly in a fund-of-fund context when facing the choice between different investment strategies. Further research can explore additional strategies, refine weighting schemes, and extend the application of our framework to diverse asset classes.

References

Basak, G. K., R. Jagannathan, and T. Ma (2009a). Electronic Companion: "Jackknife Estimator for Tracking Errors".

Basak, G. K., R. Jagannathan, and T. Ma (2009b). Jackknife estimator for tracking error variance of optimal portfolios. Management Science 55(6), 990–1002.

Bonaccolto, G. and S. Paterlini (2020). Developing new portfolio strategies by aggregation. Annals of Operations Research 292(2), 933–971.

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

Chan, L. K., J. Karceski, and J. Lakonishok (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. Review of Financial Studies 12(5), 937–974.

De Nard, G. (2022). Oops! I Shrunk the Sample Covariance Matrix Again: Blockbuster Meets Shrinkage. Journal of Financial Econometrics 20(4), 569–611.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science 55(5), 798–812.

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

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

Demiguel, V., F. J. Nogales, and R. Uppal (2014). Stock return serial dependence and out-of-sample portfolio performance. Review of Financial Studies 27(4), 1031–1073.

Elton, E. J. and M. J. Gruber (1973). Estimating the dependence structure of share prices - implications for portfolio selection. Journal of Finance 28(5), 1203–1232.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society. Series B: Statistical Methodology 75(4), 603–680.

Frost, P. A. and J. E. Savarino (1986). An empirical Bayes approach to efficient portfolio selection. Journal of Financial and Quantitative Analysis 21(3), 293–305.

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

Kan, R. and G. Zhou (2007). Optimal portfolio choice with parameter uncertainty. Journal of Financial and Quantitative Analysis 42(3), 621–656.

Kazak, E. and W. Pohlmeier (2023). Bagged pretested portfolio selection. Journal of Business & Economic Statistics 41(4), 1116–1131.

Kircher, F. and D. Rösch (2021). A shrinkage approach for sharpe ratio optimal portfolios with estimation risks. Journal of Banking and Finance 133, 106281.

Kourtis, A., G. Dotsis, and R. N. Markellos (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. Journal of Banking and Finance 36(9), 2522–2531.

Lancewicki, T. and M. Aladjem (2014). Multi-target shrinkage estimation for covariance matrices. IEEE Transactions on Signal Processing 62(24), 6380–6390.

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

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

Ledoit, O. and M. Wolf (2004b). Honey, I shrunk the sample covariance matrix. Journal of Portfolio Management 30(4), 110–119.

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

Ledoit, O. and M. Wolf (2011). Robust performances hypothesis testing with the variance. Wilmott 2011(55), 86–89.

Ledoit, O. and M. Wolf (2015). Spectrum estimation: A unified framework for covariance matrix estimation and pca in large dimensions. Journal of Multivariate Analysis 139, 360–384.

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

Ledoit, O. and M. Wolf (2018). Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Bernoulli 24(4B), 3791–3832.

Ledoit, O. and M. Wolf (2020). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. Journal of Financial Econometrics 20(1), 187–218.

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. Journal of Business and Economic Statistics 33(3), 381–392.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. Management Science 9(2), 277–293.

Shi, F., L. Shu, A. Yang, and F. He (2020). Improving minimum-variance portfolios by alleviating overdispersion of eigenvalues. Journal of Financial and Quantitative Analysis 55(8), 2700–2731.

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

Xing, X., J. Hu, and Y. Yang (2014). Robust minimum variance portfolio with L-infinity constraints. Journal of Banking and Finance 46, 107–117.

Yen, Y. M. (2016). Sparse weighted-norm minimum variance portfolios. Review of Finance 20(3), 1259–1287.

Tables and figures

Table 1: List of averaging rules, strategies, and benchmarks

This table lists the proposed averaging rules in Panel A, the strategies in Panel B, and the benchmarks, representing sophisticated estimates of the minimum-variance portfolio, in Panel C.

#	Description	Abbreviation
Panel A: Averaging rules		
1	Averaging with variance optimization on $i.i.d$ jackknife returns	AV-VAR
2	Averaging with ex-ante Sharpe ratio optimization on $i.i.d$ jackknife returns	AV-SRA
3	Averaging with ex-post Sharpe ratio optimization on $i.i.d$ jackknife returns	AV-SRP
Panel B: Strategies		
4	Minimum-variance portfolio based on the sample covariance matrix	Sample
5	Minimum-variance portfolio with the non-linear shrinkage estimator of the covariance matrix (Ledoit and Wolf, 2015)	LWNLS
6	Minimum-variance portfolio with a single market factor (Sharpe, 1963)	1F
7	Minimum-variance portfolio with the constant correlation model implied covariance matrix (Elton and Gruber, 1973)	CC
8	Minimum-variance portfolio with a scalar multiple of the identity matrix, representing the 1/N strategy of DeMiguel et al. (2009)	ID
9	Minimum-variance portfolio with short-sale constraints (Jagannathan and Ma, 2003)	SC
Panel C: Benchmarks		
10	Minimum-variance portfolio with the covariance matrix as weighted average between the sample covariance and the single factor covariance matrix (Ledoit and Wolf, 2003)	LW1F
11	Minimum-variance portfolio with the covariance matrix as weighted average between the sample covariance and a scalar multiple of the identity matrix (Ledoit and Wolf, 2004a)	LWID
12	Minimum-variance portfolio with the covariance matrix as weighted average between the sample covariance and a constant correlation covariance matrix (Ledoit and Wolf, 2004b)	LWCC
13	Minimum-variance portfolio with the covariance matrix as weighted average between the sample covariance and a blockbuster clustered constant-variance-covariance matrix (De Nard, 2022)	BBCVC
14	Minimum-variance portfolio based on a linear combination of the inverse covariance matrix of the sample, identity matrix, and single factor matrix (Kourtis et al., 2012)	KDM

Table 2: Summary statistics of simulation parameters

This table reports the annualized daily means, standard deviations (Std.), and cross-correlations between the three Fama-French factors market (MKT), size (SMB), and value (HML) in the simulation. The reported values correspond to the empirical values over the period from June 1973 until June 2019 based on daily returns.

	MKT	SMB	HML
Descriptive statistics (in percent)			
Mean	7.21	1.51	3.78
Std.	16.53	8.81	8.50
Cross-correlations			
MKT	1.000	-0.135	-0.179
SMB		1.000	-0.069
HML			1.000

Table 3: Correlation of portfolio returns and estimation errors for one year estimation window

This table shows the correlation between the simulated returns and between estimation errors of the single strategies in Panel A and Panel B, respectively. All values are based on an estimation window of one year over all 1,000 simulation runs

	Sample	LWNLS	1F	CC	ID	SC
Panel A: Return correlation						
Sample	1.000	0.986	0.828	0.788	0.618	0.703
LWNLS		1.000	0.860	0.831	0.675	0.756
1F			1.000	0.953	0.670	0.760
CC				1.000	0.749	0.830
ID					1.000	0.975
SC						1.000
Panel B: Estimation error correlation						
Sample	1.000	0.824	0.248	0.155	0.003	0.064
LWNLS		1.000	0.580	0.530	0.371	0.432
1F			1.000	0.840	0.654	0.690
CC				1.000	0.758	0.808
ID					1.000	0.950
SC						1.000

Table 4: Out-of-sample standard deviation in simulations

This table reports the average annualized out-of-sample standard deviations (in percent) over 1,000 simulation runs. *Optimum* refers to the standard deviation of the minimum-variance portfolio based on the covariance matrix with known parameters of the data-generating process. Panel A reports our proposed averaging rules with *variance* (AV-Var), *ex-ante Sharpe ratio* (AV-SRA), and *ex-post Sharpe ratio* optimization (AV-SRP) based on *i.i.d* jackknife returns using either Sample or LWNLS as a baseline strategy. The respective standard deviations of the single strategies and benchmarks are reported in Panel B and C, respectively. The standard deviations are reported for estimation windows of $\tau = \{0.5, 1, 5, 10, 25, 50, 100, 200\}$ years of data, with each year comprising 252 simulated daily returns. The abbreviations for the averaging rules, strategies, and benchmarks are explained in Table 1. We mark the portfolio with the lowest standard deviation for each estimation window bold.

τ (in Years)	0.5	1	5	10	25	50	100	200
Optimum	10.54	10.54	10.54	10.54	10.54	10.54	10.54	10.54
Panel A: Averaging rules								
AV-Sample-VAR	11.57	11.04	**10.64**	**10.59**	**10.56**	**10.55**	**10.55**	**10.55**
AV-Sample-SRA	13.80	13.65	13.50	13.55	13.51	13.55	13.55	13.57
AV-Sample-SRP	13.80	13.66	13.48	13.53	13.50	13.55	13.55	13.57
AV-LWNLS-VAR	11.40	11.00	**10.64**	**10.59**	**10.56**	**10.55**	**10.55**	**10.55**
AV-LWNLS-SRA	13.87	13.73	13.52	13.56	13.52	13.55	13.55	13.57
AV-LWNLS-SRP	13.89	13.73	13.49	13.54	13.50	13.55	13.55	13.57
Panel B: Single strategies								
Sample	11.72	11.08	10.65	**10.59**	**10.56**	**10.55**	**10.55**	**10.55**
LWNLS	**11.37**	**10.99**	**10.64**	**10.59**	**10.56**	**10.55**	**10.55**	**10.55**
1F	13.21	12.67	12.15	12.08	12.04	12.02	12.01	12.01
CC	13.75	13.26	12.80	12.74	12.71	12.69	12.68	12.68
ID	16.17	16.17	16.17	16.17	16.17	16.17	16.17	16.17
SC	14.62	14.51	14.40	14.39	14.38	14.38	14.38	14.38
Panel C: Benchmarks								
LW1F	11.45	11.02	**10.64**	**10.59**	**10.56**	**10.55**	**10.55**	**10.55**
LWID	11.42	11.01	**10.64**	**10.59**	**10.56**	**10.55**	**10.55**	**10.55**
LWCC	11.68	11.08	10.65	**10.59**	**10.56**	**10.55**	**10.55**	**10.55**
BBCVC	14.74	12.07	10.70	10.61	10.57	**10.55**	**10.55**	**10.55**
KDM	13.21	12.67	12.15	12.08	12.04	12.02	12.01	12.01

Table 5: Out-of-sample Sharpe ratios in Simulations

This table reports the average annualized out-of-sample Sharpe ratio (in percent) over 1,000 simulation runs. *Optimum* refers to the Sharpe ratio maximizing combination of the Sample, 1F, CC, ID and SC strategies based on the known parameters of the data-generating process. Panel A reports the proposed averaging rules with *variance* (AV-Var), *ex-ante Sharpe ratio* (AV-SRA), and *ex-post Sharpe ratio* optimization (AV-SRP) based on *i.i.d* jackknife returns using either Sample or LWNLS as a baseline strategy. The respective Sharpe ratios of the single strategies and benchmarks are reported in Panel B and C, respectively. The Sharpe ratios are reported for estimation windows of $\tau = \{0.5, 1, 5, 10, 25, 50, 100, 200\}$ years of data, with each year comprising 252 simulated daily returns. We mark the portfolio with the highest Sharpe ratio for each estimation window bold.

T (in Years)	0.5	1	5	10	25	50	100	200
Optimum	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
Panel A: Averaging rules								
AV-Sample-VAR	0.41	0.39	0.37	0.36	0.37	0.36	0.36	0.36
AV-Sample-SRA	0.54	**0.57**	**0.65**	**0.68**	**0.69**	**0.70**	**0.70**	**0.70**
AV-Sample-SRP	0.54	**0.57**	0.64	0.67	**0.69**	**0.70**	**0.70**	**0.70**
AV-LWNLS-VAR	0.41	0.39	0.37	0.37	0.37	0.36	0.36	0.36
AV-LWNLS-SRA	0.55	**0.57**	0.64	0.67	**0.69**	**0.70**	**0.70**	**0.70**
AV-LWNLS-SRP	0.54	**0.57**	0.64	0.67	**0.69**	**0.70**	**0.70**	**0.70**
Panel B: Single strategies								
Sample	0.33	0.35	0.36	0.36	0.36	0.36	0.36	0.36
LWNLS	0.40	0.39	0.37	0.36	0.37	0.36	0.36	0.36
1F	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.43
CC	0.46	0.47	0.48	0.48	0.48	0.48	0.48	0.48
ID	**0.57**	**0.57**	0.57	0.57	0.57	0.57	0.57	0.57
SC	0.55	0.55	0.57	0.56	0.56	0.56	0.56	0.56
Panel C: Benchmarks								
LW1F	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
LWID	0.39	0.38	0.36	0.36	0.36	0.36	0.36	0.36
LWCC	0.44	0.42	0.37	0.37	0.37	0.36	0.36	0.36
BBCVC	0.56	0.51	0.41	0.39	0.37	0.37	0.37	0.36
KDM	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.43

Table 6: Empirical out-of-sample standard deviations

This table reports the annualized out-of-sample standard deviations of our averaging rules with *variance* (AV-VAR), *ex-ante Sharpe ratio* (AV-SRA), and *ex-post Sharpe ratio* optimization (AV-SRP) using Sample and LWNLS as a baseline strategy in Panel A. The respective standard deviations of the single strategies and benchmarks are reported in Panel B and C, respectively. The out-of-sample period for all five data sets, comprising the largest single stocks $N = \{30, 50, 100, 150, 200\}$ in the U.S. by June each year, ranges from June 1976 until June 2019. We report statistical significance for the null hypothesis that the log out-of-sample variance of each averaging rule including the Sample (LWNLS) strategy is equal to or higher than the log out-of-sample variance of the Sample (LWNLS) strategy. All p-values are based on the HAC inference of Ledoit and Wolf (2011). We mark the portfolio with the lowest standard deviation for each estimation window bold.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR	13.51	12.74	11.20	10.70	10.39
	0.0000	*0.0000*	*0.0000*	*0.0000*	*0.0000*
AV-Sample-SRA	14.96	14.68	13.65	13.48	13.24
	0.9999	*1.0000*	*1.0000*	*0.0024*	*0.0000*
AV-Sample-SRP	15.01	14.64	13.78	13.68	13.49
	1.0000	*1.0000*	*1.0000*	*0.0116*	*0.0000*
AV-LWNLS-VAR	**13.50**	**12.73**	**11.13**	**10.47**	**10.01**
	0.0000	*0.0000*	*0.0000*	*0.0000*	*0.0000*
AV-LWNLS-SRA	14.99	14.75	13.79	13.47	13.03
	1.0000	*1.0000*	*1.0000*	*1.0000*	*1.0000*
AV-LWNLS-SRP	14.95	14.64	13.77	13.50	13.07
	1.0000	*1.0000*	*1.0000*	*1.0000*	*1.0000*
Panel B: Strategies					
Sample	14.00	13.48	12.87	14.18	17.70
LWNLS	13.73	12.96	11.31	10.66	10.20
1F	14.89	14.36	12.89	12.37	12.10
CC	14.95	15.07	14.46	14.57	14.37
ID	17.40	17.41	17.52	17.28	17.08
SC	13.78	13.21	12.30	11.76	11.34
Panel C: Benchmarks					
LW1F	13.74	12.94	11.20	10.60	10.10
LWID	13.75	13.08	11.78	11.52	11.23
LWCC	13.68	12.98	11.33	10.89	10.47
BBCVC	13.73	13.04	11.67	11.13	10.76
KDM	14.89	14.36	12.89	12.37	12.10

Table 7: Empirical out-of-sample Sharpe ratios

This table reports the annualized out-of-sample Sharpe ratio of our averaging rules with *variance* (AV-VAR), *ex-ante Sharpe ratio* (AV-SRA), and *ex-post Sharpe ratio* optimization (AV-SRP) based on exponentially smoothed jackknife returns using either Sample or LWNLS as a baseline strategy in Panel A. The respective Sharpe ratios of the single strategies and benchmarks are reported in Panel B and C, respectively. The out-of-sample period for all five data sets, comprising the largest single stocks $N = \{30, 50, 100, 150, 200\}$ in the U.S. by June each year, ranges from June 1976 until June 2019. We report statistical significance for the null hypothesis that the Sharpe ratio of each averaging rule including the Sample (LWNLS) strategy is equal to or less than the Sharpe ratio of the Sample (LWNLS) strategy. All p-values are based on the HAC inference of Ledoit and Wolf (2008). We mark the portfolio with the highest Sharpe ratio for each estimation window bold.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR	0.528	0.563	0.511	0.562	0.591
	0.1393	*0.0409*	*0.0920*	*0.0020*	*0.0618*
AV-Sample-SRA	0.545	0.566	0.540	0.602	0.522
	0.2567	*0.2011*	*0.1500*	*0.0100*	*0.1705*
AV-Sample-SRP	0.582	0.590	**0.579**	0.633	0.525
	0.1474	*0.1394*	*0.0857*	*0.0049*	*0.1656*
AV-LWNLS-VAR	0.525	0.554	0.502	0.570	0.603
	0.4943	*0.1697*	*0.4039*	*0.1906*	*0.3612*
AV-LWNLS-SRA	0.538	0.546	0.531	0.585	0.597
	0.4380	*0.4142*	*0.3710*	*0.3429*	*0.4740*
AV-LWNLS-SRP	**0.589**	**0.594**	0.570	**0.646**	**0.641**
	0.2367	*0.2390*	*0.2423*	*0.1713*	*0.3242*
Panel B: Strategies					
Sample	0.486	0.479	0.410	0.266	0.372
LWNLS	0.525	0.525	0.494	0.538	0.589
1F	0.455	0.491	0.485	0.508	0.499
CC	0.464	0.523	0.456	0.440	0.475
ID	0.500	0.494	0.493	0.515	0.524
SC	0.513	0.543	0.501	0.576	0.567
Panel C: Benchmarks					
LW1F	0.501	0.521	0.516	0.549	0.620
LWID	0.505	0.505	0.455	0.437	0.546
LWCC	0.514	0.519	0.507	0.494	0.578
BBCVC	0.556	0.537	0.467	0.534	0.583
KDM	0.455	0.491	0.485	0.508	0.499

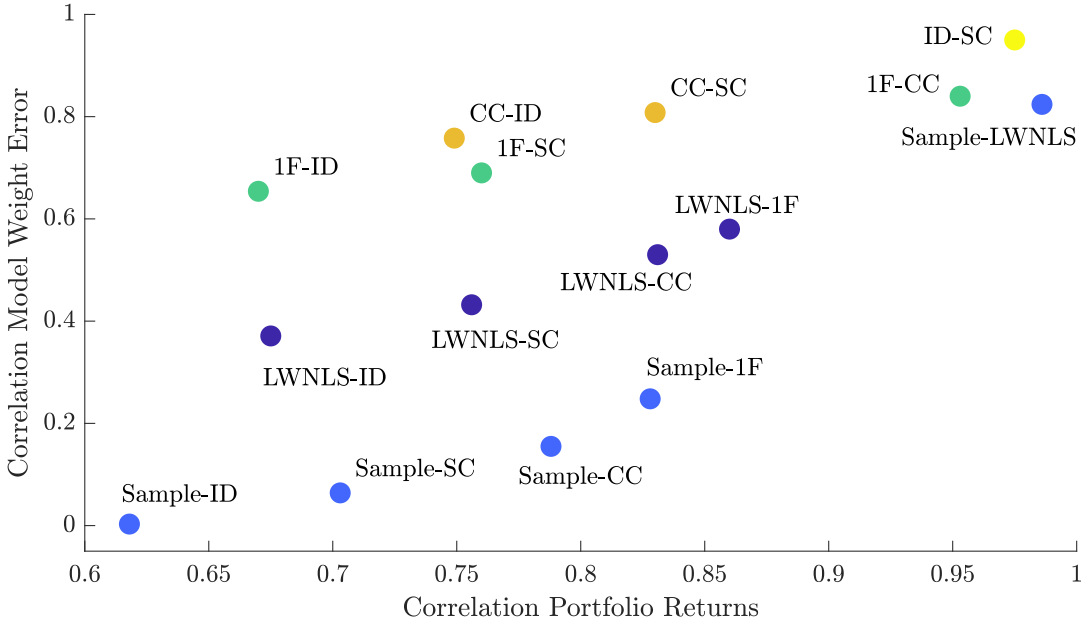
Table 8: Empirical out-of-sample Sharpe ratios adjusted for transaction costs of 15bps

This table reports the annualized out-of-sample Sharpe ratio after transaction costs of 15 basis points for our averaging rules with *variance* (AV-VAR), *ex-ante Sharpe ratio* (AV-SRA), and *ex-post Sharpe ratio* optimization (AV-SRP) using either Sample or LWNLS as a baseline strategy in Panel A. The respective Sharpe ratios of the single strategies and benchmarks are reported in Panel B and C, respectively. The out-of-sample period for all five data sets, comprising the largest single stocks $N = \{30, 50, 100, 150, 200\}$ in the U.S. by June each year, ranges from June 1976 until June 2019. We report statistical significance for the null hypothesis that the Sharpe ratio of each averaging rule including the Sample (LWNLS) strategy is equal to or less than the Sharpe ratio of the Sample (LWNLS) strategy. All p-values are based on the HAC inference of Ledoit and Wolf (2008). We mark the portfolio with the highest Sharpe ratio for each estimation window bold.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR	0.453	0.458	0.352	0.370	0.397
	0.0185	*0.0004*	*0.0001*	*0.0000*	*0.0000*
AV-Sample-SRA	0.464	0.470	0.404	0.444	0.277
	0.1514	*0.0476*	*0.0026*	*0.0000*	*0.0000*
AV-Sample-SRP	0.507	0.499	0.455	0.489	0.328
	0.0691	*0.0236*	*0.0006*	*0.0000*	*0.0000*
AV-LWNLS-VAR	0.451	0.454	0.357	0.396	0.410
	0.2231	*0.0183*	*0.0508*	*0.0096*	*0.0327*
AV-LWNLS-SRA	0.463	0.458	0.422	0.476	0.473
	0.3438	*0.2439*	*0.1404*	*0.0782*	*0.1272*
AV-LWNLS-SRP	**0.518**	**0.512**	0.469	**0.544**	**0.527**
	0.1585	*0.1048*	*0.0619*	*0.0195*	*0.0471*
Panel B: Strategies					
Sample	0.370	0.295	0.052	-0.304	-0.543
LWNLS	0.428	0.391	0.300	0.309	0.336
1F	0.379	0.398	0.372	0.388	0.376
CC	0.395	0.436	0.347	0.326	0.357
ID	0.479	0.472	**0.470**	0.491	0.499
SC	0.467	0.489	0.435	0.504	0.488
Panel C: Benchmarks					
LW1F	0.403	0.385	0.309	0.292	0.324
LWID	0.406	0.353	0.186	0.072	0.100
LWCC	0.424	0.389	0.291	0.214	0.242
BBCVC	0.479	0.424	0.273	0.271	0.269
KDM	0.379	0.398	0.372	0.388	0.376

Figure 1: Correlations of portfolio returns vs. estimation errors

This figure plots the correlations of portfolio returns provided in Table 3 in Panel A vs. the correlations of estimation errors in Panel B for all bivariate strategy combinations based on the simulated returns.



Internet Appendix for

An averaging framework for minimum-variance portfolios: Optimal rules for combining portfolio weights

Roland Füss, Thorsten Glück, Christian Koeppel, Felix Miebs

This Version: September 20, 2024

Abstract

We propose an averaging framework for combining minimum-variance strategies to either minimize the expected out-of-sample variance or maximize the expected out-of-sample Sharpe ratio. Our framework overcomes the problem of selecting the "best" strategy ex-ante by optimally averaging over portfolio weights. This averaging procedure has an intuitive economic interpretation because it resembles a fund-of-fund approach, where each minimum-variance strategy represents a single fund. In a range of simulations, for a set of well-established strategies, we show that optimally averaging over portfolio weights improves the out-of-sample variance and Sharpe ratio. We confirm the finding of our simulation study on empirical data.

JEL Classification: G11

Keywords: Averaging; diversification; estimation error; portfolio optimization; shrinkage.

A Internet Appendix

In this appendix, we present supplementary information and results with the following structure:

Appendix A.1: Description of estimators

Appendix A.2: Consistency of jackknife-type estimators

Appendix A.3: Consistency of ex-post Sharpe ratio-maximizing averaging rule

A.1 Description of estimators

In this section, we describe the considered minimum-variance estimators introduced in Section 2.2.

Sample estimator

The sample estimator is denoted by $\hat{\Sigma}_S$ and is based on the unbiased estimation of the vector of means m of the in-sample asset excess returns in the estimation window, that is, the sample size τ:

$$\hat{\Sigma}_S = \frac{1}{\tau} X \left(I - \frac{1}{\tau} 1_N 1_N^{'} \right) X' \tag{A.1.1}$$

where X is a $N \times \tau$ matrix of τ observations and N assets, 1_N is a conformable vector of ones, and I is a conformable identity matrix.

Single-factor model

The return of stock i at time t, $r_{i,t}$ is in the single-factor model of Sharpe (1963), which is described as:

$$r_{i,t} = \alpha_i + \beta_i r_{MKT,t} + \epsilon_{i,t}, \tag{A.1.2}$$

where α_i is the mispricing of stock i, β_i is the beta factor of stock i, $r_{MKT,t}$ is the excess return of the market over the risk-free rate at time t, and $\epsilon_{i,t}$ is the residual return of stock i at time t. Following Chan et al. (1999), the covariance matrix estimator $\hat{\Sigma}_{1F}$ implied by the single-factor model (1F) is given by:

$$\hat{\Sigma}_{1F} = \hat{\sigma}^2_{MKT} \hat{\beta} \hat{\beta}' + \hat{\Delta}, \tag{A.1.3}$$

where $\hat{\sigma}^2_{MKT}$ is an unbiased estimate of the market variance, $\hat{\beta}$ is a $N \times 1$ vector comprising the estimates of stock betas, and $\hat{\Delta}$ is a diagonal matrix with the residual variances along the main diagonal.

Constant correlation model

Elton and Gruber (1973) suggest a structured estimator of the covariance matrix assuming that each pair of assets has the identical correlation coefficient, while each of the N assets has an individual standard deviation. The resulting constant correlation covariance matrix estimator is given by:

$$\hat{\Sigma}_{CC} = \Lambda C \Lambda', \tag{A.1.4}$$

where Λ denotes a diagonal matrix containing the sample standard deviations $\hat{s}_i$ of the individual assets with $i = \{1, 2, 3, \ldots, N\}$, and C is the constant correlation matrix with ones along the main diagonal and the average of all pairwise correlations off the main diagonal.

Multiple of the identity matrix

This estimator implies that all assets are uncorrelated and exhibit equal variances. Following Ledoit and Wolf (2004a), the elements along the main diagonal are set to the average variance of all assets in the investment universe. The covariance matrix estimator for the scaled identity matrix model ID is given by:

$$\hat{\Sigma}_{ID} = I\mu_{\hat{s}}, \text{ with } \mu_{\hat{s}} = \frac{1}{N}\sum_{i=1}^{N} \hat{s}_i^2, \tag{A.1.5}$$

where $\hat{s}_i^2$ is the sample variance of asset i. The resulting minimum-variance portfolio corresponds to the 1/N portfolio.

Linear shrinkage estimators of the covariance matrix

The linear shrinkage estimators of Ledoit and Wolf (2003, 2004a,b) and DeNard et al.(2020) have the following general form:

$$\hat{\Sigma}_{LW} = \phi\hat{F} + (1-\phi)\hat{\Sigma}_S. \tag{A.1.6}$$

The shrunk covariance matrix $\hat{\Sigma}_{LW}$ is a convex combination of the estimated sample covariance matrix $\hat{\Sigma}_S$ and a shrinkage target $\hat{F}$, with the shrinkage intensity, ϕ, taking values between zero and one. The optimal shrinkage intensity is determined by minimizing the mean squared error of $\hat{\Sigma}_{LW}$. Ledoit and Wolf (2003, 2004a,b) consider three different candidates for $\hat{F}$: the single-factor model covariance matrix $\hat{\Sigma}_{LW1F}$ (Ledoit and Wolf, 2003), a multiple of the identity matrix $\hat{\Sigma}_{LWID}$ (Ledoit and Wolf, 2004a), and the constant correlation model implied covariance matrix $\hat{\Sigma}_{LWCC}$ (Ledoit and Wolf, 2004b).

De Nard (2022) suggests a novel class of purely statistical constant-variance-covariance (CVC) shrinkage targets. Starting from $\hat{\Sigma}_{ID}$ as a target $\hat{F}$, the authors aim at reducing the bias of the target by setting the diagonal elements to φ and the off-diagonal elements to ν:

$$F_{CVC} = \varphi I + \nu Q, \tag{A.1.7}$$

with $Q := 11' - I$, such that the mean squared error of the shrinkage estimator in Equation (A.1.6) is minimized. Generalizing this idea, De Nard (2022) suggests the described structure for (homogeneous) groups of stocks. With Z groups of (homogeneous) stocks, this gives rise to the following generalized constant-variance-covariance (GCVC) shrinkage target:

$$F_{GCVC} = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1Z} \\ F_{21} & F_{22} & \dots & F_{2Z} \\ \vdots & \vdots & \ddots & \vdots \\ F_{Z1} & F_{Z2} & \dots & F_{ZZ} \end{bmatrix}, \tag{A.1.8}$$

where F_{lm} denotes the (l, m)-th cluster of F. Each cluster has the structure in Equation (A.1.7), such that:

$$F_{lm} = \begin{cases} \varphi_z I_{N_z} + \nu_z Q_{N_z}, \text{ for } l = m = z \in \{1, ...Z\} \\ \nu_{lm} 1_{N_l} 1'_{N_m}, \text{ for } l \neq m \in \{1, ...Z\} \end{cases}$$

where N_z denotes number of stocks in group g, I_{N_z} is the $N_z \times N_z$ identity matrix, Q_{N_z} is the respective $N_z \times N_z$ off-diagonal matrix, and 1_{N_z} and 1_{N_l} denote a $N_z \times 1$ and $N_l \times 1$ vector of ones, respectively.

De Nard (2022) suggests different approaches to identify homogeneous groups. We use the authors' blockbuster approach[1], because it does not require any additional data except a specification of the number of homogeneous groups. We set the latter to two, motivated by the illustrative example of fincancial and non-financial firms discussed in De Nard (2022).

Nonlinear shrinkage estimator of the covariance matrix

Ledoit and Wolf (2015) consider the class of rotation-equivariant covariance estimators of type $\hat{\Sigma} = \sum_{n=1}^{N} \hat{d}_n u_n u'_n$ where u_n is the n'th eigenvector of the sample covariance matrix. Their estimator $\hat{d}_n$ is a consistent approximation to the optimal but infeasible estimator derived in Ledoit and Péché (2011). The basic idea is to assume that the ratio

[1]The exact procedure is described in Alogorithm 1 in De Nard (2022). We thank Gianluca De Nard for providing us the code.

$N/T \to c \neq 1$ as $T \to \infty$, such that the eigenvalues of the sample covariance matrix can be used to consistently estimate their population counterparts. This can be achieved by utilizing the relationship between the limiting spectral distributions of the sample and population eigenvalues via the QuEST function introduced in Ledoit and Wolf (2015).[2] The estimator $\hat{d}_n$ is then a function of the sample and estimated population eigenvalues. In case of $N < T$, $\hat{d}_n$ is defined as

$$\hat{d}_n = \frac{\lambda_i}{|1 - \frac{N}{T} - \frac{N}{T}\lambda_i \cdot m_{T,N}(\hat{\tau}, \lambda_i)|^2} \tag{A.1.9}$$

where λ_i is the i'th sample eigenvalue, $\hat{\tau} = \{\tau_1, ..., \tau_N\}$ are the estimated population eigenvalues, and $m_{T,N}(\hat{\tau}, \lambda_i)$ is the unique solution to a complex valued function as defined by Ledoit and Wolf (2015), Equation (2.14). In case of $N > T$, one has to differentiate between zero and nonzero sample eigenvalues. If $\lambda_i > 0$, then $\hat{d}_n$ is defined as in Equation (A.1.9). If $\lambda_i = 0$, then

$$\hat{d}_i = \frac{1}{\left(\frac{N}{T} - 1\right) m_{\underline{F}}(0)} \tag{A.1.10}$$

where $m_{\underline{F}}(0)$ is the unique solution to Equation (3.10) in Ledoit and Wolf (2015). Finally, note that the case $N = T$ is ruled out as $c \neq 1$ by assumption.

Shrinkage estimators of the inverse covariance matrix

Kourtis et al. (2012) propose a shrinkage estimator of the inverse covariance matrix as a convex combination of the estimated inverse sample covariance matrix $\hat{\Sigma}_S^{-1}$, the identity I, the inverse covariance matrix $\hat{\Sigma}_{1F}^{-1}$ implied by the single-factor model of Sharpe (1963). The shrunken inverse covariance matrix $\hat{\Sigma}_{KDM}^{-1}$ is then given by:

$$\hat{\Sigma}_{KDM}^{-1} = \phi_1 \hat{\Sigma}_S^{-1} + \phi_2 I + (1 - \phi_1 - \phi_2)\hat{\Sigma}_{1F}^{-1}, \tag{A.1.11}$$

[2]For an excellent and detailed treatment, we refer the interested reader to Ledoit and Wolf (2017), particularly sections 3 and 4. A Matlab implementation of the QuEST function is provided as a supplemental to Ledoit and Wolf (2017) and can be downloaded from Michael Wolf's website: https://www.econ.uzh.ch/en/people/faculty/wolf/publications.html. Our implementation of the nonlinear shrinkage estimator is based on this code.

where the intensity parameters ϕ_1, ϕ_2, ϕ_3 are determined by minimizing the out-of-sample portfolio variance of jackknife returns.

Short-sale constraints

Jagannathan and Ma (2003) provide a moment shrinkage interpretation of short-sale constraints in the minimum-variance portfolio optimization problem. The short-sale constraint augments the optimization problem $\min_\omega \omega'\Sigma\omega$ s.t. $\omega' 1_N$ by the following inequality:

$$\hat{w}_i \geq 0, \text{ for } i = 1, 2, \ldots, N. \tag{A.1.12}$$

Jagannathan and Ma (2003) show that the asset weights of the short-sale constrained minimum-variance portfolio correspond to those of the global minimum-variance portfolio based on the following covariance estimator:

$$\hat{\Sigma}_{SC} = \hat{\Sigma}_S - (\delta 1' + 1\delta'), \tag{A.1.13}$$

where $\delta = (\delta_1, \delta_2, \ldots, \delta_N)$ is the vector of Lagrange multipliers for the non-negativity constraints in Equation (A.1.12).

A.2 Consistency of jackknife-type estimators

In this section, we proof the consistency of our jackknife-type estimators for the conditional expectations of the out-of-sample variance and out-of-sample mean of portfolio returns as induced by some averaging weights $\overline{w}$. The proofs heavily build on Basak et al. (2009). At several steps, we have to use the Frobenius norm, which is defined for any $n \times m$ Matrix $A \in \mathbb{R}^{n \times m}$ as $||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$, where a_{ij} is the ij'th entry of A. We also need following well-known properties:

$$||cA||_F = |c| \cdot ||A||_F \tag{A.2.1}$$

$$||A + B||_F \leq ||A||_F + ||B||_F \tag{A.2.2}$$

$$||Ax||_F \leq ||A||_F \cdot ||x||_F \tag{A.2.3}$$

$$\lambda_{min}||x||_F^2 \leq ||x'Cx||_F, \tag{A.2.4}$$

where $A, B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times m}$ is positive definite with the smallest eigenvalue $\lambda_{min} > 0$, $x \in \mathbb{R}^{m \times 1}$ and $c \in \mathbb{R}$. For a detailed discussion, however, we refer to section 5.6 in Horn and Johnson (2013).[3] The consistency of the out-of-sample variance is subject to the next section followed by the out-of-sample mean in section A.2.2.

A.2.1 Out-of-sample variance

For the sake of completeness, we first restate the definition of the jackknife-type estimator: Let w_{-j} be portfolio weights constructed from the returns series without block j. The estimator is then defined as

$$q_{J+1} = \sum_{j=1}^{J} \alpha_j w'_{-j} S_j w_{-j} \tag{A.2.5}$$

where $J = T/L$ is the number of blocks, S_j the sample covariance of block j, $\alpha_j \geq 0$ for all j and $\sum_{j=1}^{J} \alpha_j = 1$. We also recall that H_T denotes the conditional expectation of the sample covariance matrix of the first out-of-sample block $J+1$, i.e. $H_T = \mathbb{E}_T[S_{J+1}]$.

[3] Regarding (A.2.4), note that we can decompose C such that $x'Cx = x'ULU'x$ where $L = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with λ_i being the i'th eigenvalue of C and U being orthogonal. Defining $y = U'x$, we can rewrite $x'Cx = y'Ly = \sum_{i=1}^{n} y_i^2 \lambda_i$. Because $y'y = x'UU'x = x'x$ we get $x'Cx \geq \lambda_{\min}(y'y) = \lambda_{\min}||x||_F^2$ and, thus, (A.2.4).

Regarding the proof that q_{J+1} consistently estimates the out-of-sample variance $w'H_Tw$ for some given weights w, we can directly refer to Basak et al. (2009):

Proposition A.1. *Assume that (i) the blocks of L returns follow a stationary process, (ii) returns and portfolio weights w and w_{-j} have finite fourth moment, (iii) the portfolio weights satisfy $||w-w_{-j}||_F = O_p(J^{-1})$, and (iv) $\sum_{j=1}^{J} a_j S_j - H_T$ converges to zero in mean square. Then, for a fixed w, the jackknife-type estimator q_{J+1} consistently estimates the conditional expectation of the out-of-sample variance, that is, $w'H_Tw$.*

Proof. See Basak et al. (2009), Proposition EC.1 together with Remark EC.2. ■

Note that assumptions (i) and (iv) in this proposition must hold, regardless of how the portfolio weights are constructed. As a second step, we only have to deal with assumptions (ii) and (iv). Regarding (ii), we note that the averaging weights $\overline{w}$ and $\overline{w}_{-j}$ have finite fourth moment if all M strategies, that is, $w^{(1)}, ..., w^{(M)}$ and $w_{-j}^{(1)}, ..., w_{-j}^{(M)}$, have finite fourth moment.[4] If we assume this is indeed the case, we only have to show that $||\overline{w}-\overline{w}_{-j}||_F = O_p(J^{-1})$. To verify this, we first provide in Lemma A.1 below, under which conditions $||w^{(m)} - w_{-j}^{(m)}||_F = O_p(J^{-1})$.[5] Then, we show that if $||w^{(m)} - w_{-j}^{(m)}||_F = O_p(J^{-1})$ holds for all $m = 1, ..., M$ models, it follows that $||\bar{w} - \bar{w}_{-j}||_F = O_p(J^{-1})$.

Lemma A.1. *Let $\hat{\Sigma}$ and $\hat{\Sigma}_{-j}$ be estimators for Σ with and without block j, respectively. Assume that (i) $||\hat{\Sigma} - \hat{\Sigma}_{-j}||_F = O_p(J^{-1})$, (ii) $\hat{\Sigma}$ and $\hat{\Sigma}_{-j}$ are positive definite and (iii) $\lambda_{min}^{-1} = O_p(1)$, where λ_{min} is the smallest eigenvalue of $\hat{\Sigma}$. If w and w_{-j} are the minimum-variance weights constructed from $\hat{\Sigma}$ and $\hat{\Sigma}_{-j}$, respectively, it follows that $||w-w_{-j}||_F = O_p(J^{-1})$.*

Proof. We only have to consider the case where $\hat{\Sigma} \neq \hat{\Sigma}_{-j}$ and $w \neq w_{-j}$. Let $\delta_w = w_{-j} - w$. Because $\delta_w \neq 0$ and $\hat{\Sigma}$ is positive definite, we obtain for the quadratic form

[4] This can be verified via the c_r inequality (see White, 2000, Theorem 3.8).

[5] This proof is basically identical to the proof of Lemma EC.1 in Basak et al. (2009), who show that property $||w-w_{-j}||_F = O_p(J^{-1})$ holds if w and w_{-j} are minimum-variance portfolio weights constructed from the sample covariance estimator. They also explicitly note in Remark EC.2 that their proof should hold for a general class of suitable covariance estimators. We provide Lemma A.1 only for the sake of completeness.

$\delta_w'\hat{\Sigma}\delta_w > 0$. As shown in Basak et al. (2009), the optimality criteria of the minimum-variance optimization problem also yield $\delta_w'\hat{\Sigma}\delta_w \leq \delta_w'(\hat{\Sigma} - \hat{\Sigma}_{-j})w_{-j}$. Because both sides of this inequality are positive scalars, it also holds with the Frobenius norm, that is, $||\delta_w'\hat{\Sigma}\delta_w||_F \leq ||\delta_w'(\hat{\Sigma} - \hat{\Sigma}_{-j})w_{-j}||_F$. As a result, property (A.2.3) and property (A.2.4) of the Frobenius norm yield

$$\lambda_{min}||\delta_w||_F^2 \leq ||\delta_w||_F \cdot ||\hat{\Sigma} - \hat{\Sigma}_{-j}||_F \cdot ||w_{-j}||_F. \tag{A.2.6}$$

and thus

$$||\delta_w||_F \leq \lambda_{min}^{-1}||\hat{\Sigma} - \hat{\Sigma}_{-j}||_F \cdot ||w_{-j}||_F. \tag{A.2.7}$$

As noted in Basak et al. (2009), $||w_{-j}||_F = O_p(1)$. By assumption $||\hat{\Sigma} - \hat{\Sigma}_{-j}||_F = O_p(J^{-1})$ and $||\lambda_{min}^{-1}||_F = O_p(1)$. Hence, the right side of (A.2.7) is $O_p(J^{-1})$ implying $||\delta_w||_F = O_p(J^{-1})$. ■

With Lemma A.1 at hand, proving the following proposition, that is, $||\overline{w} - \overline{w}_{-j}||_F = O_p(J^{-1})$, as a final step is straightforward:

Proposition A.2. *For $m = 1, ..., M$, let $\hat{\Sigma}^{(m)}$ and $\hat{\Sigma}_{-j}^{(m)}$ be the model m estimators for Σ and let $w^{(m)}$ and $w_{-j}^{(m)}$ be the minimum-variance weights constructed from $\hat{\Sigma}^{(m)}$ and $\hat{\Sigma}_{-j}^{(m)}$, respectively. For any vector $\beta \geq 0$ with $1_M'\beta = 1$, we define the averaging weights $\overline{w} = \sum_{m=1}^{M}\beta_m w^{(m)}$ and $\overline{w}_{-j} = \sum_{m=1}^{M}\beta_m w_{-j}^{(m)}$. If, for each model m, the estimators fulfill assumptions $(i) - (iii)$ in Lemma A.1, it follows that $||\overline{w} - \overline{w}_{-j}||_F = O_p(J^{-1})$.*

Proof. Because weight vector $\beta \geq 0$, property (A.2.2) of the Frobenius norm yields

$$\begin{aligned} ||\bar{w} - \bar{w}_{-j}||_F &= ||\sum_{m=1}^{M}\beta_m\left(w^{(m)} - w_{-j}^{(m)}\right)||_F \\ &\leq \sum_{m=1}^{M}\beta_m||w^{(m)} - w_{-j}^{(m)}||_F. \end{aligned} \tag{A.2.8}$$

By Lemma A.1, $\beta_m||w^{(m)} - w_{-j}^{(m)}||_F = O_p(J^{-1})$ for each m. Hence, the right-hand-side of inequality (A.2.8) is $O_p(J^{-1})$, which, in turn, yields the required result. ■

Proposition A.2 together with Proposition A.1 yields the consistency of the jackknife-type estimator in case of our averaging rule, that is, given the respective required assumptions are fulfilled, the estimator $q_{J+1} = \sum_{j=1}^{J} \alpha_j \overline{w}'_{-j} S_j \overline{w}_{-j}$ consistently estimates $\overline{w}' H_T \overline{w}$.

A.2.2 Out-of-sample mean

We aim to estimate the conditional expectation of the out-of-sample mean of portfolio returns k induced by portfolio weights w, that is, $k = w'\rho_T$, where $\rho_T = \mathbb{E}_T[\overline{r}_{J+1}]$. The jackknife-type estimator for k is similar to the estimator for the out-of-sample variance and defined as

$$k_{J+1} = \sum_{j=1}^{J} \alpha_j w'_{-j} \overline{r}_j \tag{A.2.9}$$

where, as above, w_{-j} are the portfolio weights constructed from the returns series without block j and $\overline{r}_j$ is the vector of average block j returns. The steps of the consistency proof are again very close those in Basak et al. (2009), that is, we decompose the difference between k and its estimator k_{J+1} and then consider the convergence of the components.

Proposition A.3. *Assume that (i)* $\mathbb{E}[||w - w_{-j}||_F^2] = o(1)$ *and (ii)* $\mathbb{E}[||\overline{\Delta}||_F^2] = o(1)$ *where* $\overline{\Delta} = \sum_{j=1}^{J} \alpha_j \Delta_j$ *and* $\Delta_j = \overline{r}_j - \rho_T$. *Furthermore, assume that (iii)* $||\rho_T||_F^2$ *and (iv)* $||\Delta_j||_F^2$ *are bounded by some constant c. Then, for a fixed* w, *the jackknife-type estimator* k_{J+1} *consistently estimates the conditional expectation of the out-of-sample mean* $w'\rho_T$.

Proof. Let $\delta_{w_{-j}} = w_{-j} - w$. The difference between k_{J+1} and k can then be written as

$$\begin{aligned} k_{J+1} - k &= \sum_{j=1}^{J} \alpha_j w'_{-j} \overline{r}_j - \sum_{j=1}^{J} \alpha_j w' \rho_T \\ &= \sum_{j=1}^{J} \alpha_j \delta'_{w_{-j}} \rho_T + \sum_{j=1}^{J} \alpha_j \delta'_{w_{-j}} \Delta_j + w' \overline{\Delta}. \end{aligned} \tag{A.2.10}$$

Because $a_j \geq 0$, we get by Minkowski's inequality together with property (A.2.2) and

property (A.2.3) of the Frobenius norm:

$$\begin{aligned}\mathbb{E}\,|k_{J+1}-k| &\leq \mathbb{E}\,|\sum_{j=1}^{J}\alpha_j\delta'_{w_{-j}}\rho_T| + \mathbb{E}\,|\sum_{j=1}^{J}\alpha_j\delta'_{w_{-j}}\Delta_j| + \mathbb{E}\,|w'\overline{\Delta}| \\ &\leq \sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\delta'_{w_{-j}}\rho_T||_F\right] + \sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\delta'_{w_{-j}}\Delta_j||_F\right] + \mathbb{E}\left[||w'\overline{\Delta}||_F\right]\end{aligned}$$

If each of the three sums on the right side converge to zero, consistency follows by White (2000, Theorem 2.42). First, consider $\mathbb{E}[||w'\overline{\Delta}||_F]$. By property (A.2.3) of the Frobenius norm and the Cauchy–Schwarz inequality, we obtain

$$\mathbb{E}\left[||w'\overline{\Delta}||_F\right] \leq \left(\mathbb{E}\left[||w||_F^2\right]\right)^{\frac{1}{2}} \cdot \left(\mathbb{E}\left[||\overline{\Delta}||_F^2\right]\right)^{\frac{1}{2}}. \tag{A.2.11}$$

The first term of (A.2.11) is bounded, whereas the second converges to zero by assumption (ii). Next, the second term in (A.2.11). Similar steps as above yield

$$\begin{aligned}\sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\delta'_w\Delta_j||_F\right] &\leq \sum_{j=1}^{J}\alpha_j\left(\mathbb{E}\left[||\delta_w||_F^2\right]\right)^{\frac{1}{2}} \cdot \left(\mathbb{E}\left[||\Delta_j||_F^2\right]\right)^{\frac{1}{2}} \\ &\leq \left(\sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\delta_w||_F^2\right]\right)^{\frac{1}{2}} \cdot \left(\sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\Delta_j||_F^2\right]\right)^{\frac{1}{2}},\end{aligned} \tag{A.2.12}$$

where, as noted by Basak et al. (2009), the last inequality can be obtained by the Cauchy-Schwarz inequality with weights α interpreted as probabilities. In exactly the same way, we obtain for the first term in (A.2.11) the inequality

$$\sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\delta'_w\rho_T||_F\right] \leq \left(\sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\delta_w||_F^2\right]\right)^{\frac{1}{2}} \cdot \left(\sum_{j=1}^{J}\alpha_j\,\mathbb{E}\left[||\rho_T||_F^2\right]\right)^{\frac{1}{2}}. \tag{A.2.13}$$

By assumptions (iii) and (iv), the second terms on the right side of (A.2.12) and (A.2.13) are bounded, whereas the first converge to zero by assumption (i). Hence, all three terms of (A.2.11) converge to zero, implying consistency of k_{J+1}. ■

We note that assumptions (ii), (iii), and (iv) in Proposition A.3 must hold, regardless of how portfolio weights are constructed. Regarding (i), we again refer to Basak et al.

(2009), who argue that returns and portfolio weights with finite fourth moment together with property $||w - w_{-j}||_F = O_p(J^{-1})$ are sufficient.[6] In Proposition A.2, we show that our averaging weights satisfy $||\overline{w} - \overline{w}_{-j}||_F = O_p(J^{-1})$. Given all required moment restrictions on returns and portfolio weights hold, we can thus conclude that $k_{J+1} = \sum_{j=1}^{J} \alpha_j \overline{r}_j' \overline{w}_{-j}$ consistently estimates $\overline{w}' \rho_T$.

A.3 Consistency of ex-post Sharpe ratio-maximizing averaging rule

In this section, we investigate the asymptotic behavior of our ex-post Sharpe ratio-maximizing averaging weights β^*_{SRP}, as defined in section 2.1.2. For this purpose, we simulate $K = 1000$ samples of $T = \tau \cdot 252$ daily excess returns for $N = 25$ assets. Then, we determine $\beta^*_{SRP}(\tau)$ and calculate for its m'th element

$$\widehat{\text{MSE}}\left(\beta^*_{SRP}(\tau, m)\right) = \frac{1}{K} \sum_{k=1}^{K} \left(\beta^*_{SRP,k}(\tau, m) - \beta_{SR,k}(m)\right)^2, \tag{A.3.1}$$

where $\beta_{SR}(m)$ is the Sharpe ratio-maximizing averaging weight based on the true parameters of the data-generating process for the excess returns. We use Equation (A.3.1) as an estimator for the mean squared error of $\beta^*_{SRP}(\tau, m)$, which allows us to investigate the asymptotic behavior of $\beta^*_{SRP}(\tau, m)$. In particular, if $\text{MSE}(\beta^*_{SRP}(\tau, m)) \to 0$ as $\tau \to \infty$ then $\beta^*_{SRP}(\tau, m) \to_p \beta_{SR}(m)$ (see, e.g., White, 2000, Theorem 2.42). Thus, we conclude that our ex-post Sharpe ratio-maximizing rule is consistent if the simulations show that our estimator in Equation (A.3.1) converges to zero with increasing sample size.

Regarding the data-generating process, we resort to the simulation setup outlined in section 3.1 and defined excess returns as in Equation (16). In contrast to section 3.1, however, we consider two scenarios. In the first scenario, mispricing factor and factor loadings are fixed, which isolates the effect of the sample size τ. In the second scenario, the simulation in section 3.1 with stochastic mispricing factor and factor loadings is

[6] To be precise, Basak et al. (2009) write that these conditions imply $\sum_{j=1}^{J} \alpha_j \, \mathbb{E}\left[||w - w_{-j}||_F^4\right] \to 0$ (see Basak et al., 2009, Equation EC(16)).

reproduced. The sample size $T = \tau \cdot 252$, where $\tau = \{0.5, 1, 5, 10, 25, 50, 100, 200\}$. Thus, we consider sample sizes between half a year and 200 years of daily data.

The simulation results are illustrated in Figure A.3.1. For simplicity, we only show $\widehat{\text{MSE}}(\beta^*_{SRP}(\tau))$, which is obtained by summing over all M strategy weight MSEs. This aggregation is feasible because each element of the sum is non-negative such that $\widehat{\text{MSE}}(\beta^*_{SRP}(\tau)) \to 0$ implies $\widehat{\text{MSE}}(\beta^*_{SRP}(\tau, m)) \to 0$ for all m. The left panel shows the behavior of the simulated MSE in case of constant factors. The right panel shows the simulated MSE for stochastic factors. As shown, the estimated MSE is strictly decreasing with sample size τ, indicating consistency of the averaging rule.

Figure A.3.1: Mean Squared Error of ex-post Sharpe ratio maximizing averaging weights

This figure plots estimated MSEs of the ex-post Sharpe ratio maximizing averaging weights β^*_{SRP}. The MSEs are based on simulated asset returns as defined in Equation (16) with either constant or stochastic factors. The blue markers show the MSEs obtained when the set of strategies includes the sample covariance matrix. Replacing the covariance matrix with its nonlinear shrinkage counterpart yields MSEs, represented by red markers.

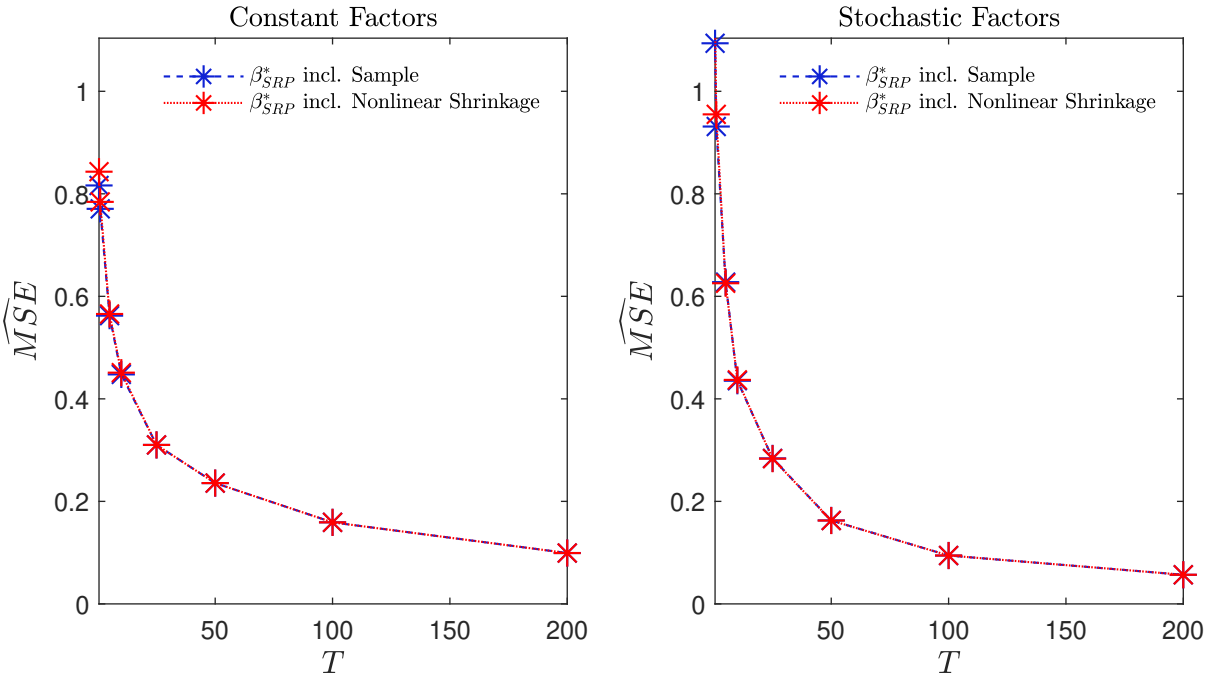


Table A.3.1: Correlation of portfolio returns across estimation windows

This table shows the Bravais-Pearson correlation coefficients between the simulated returns of the single strategies over all 1,000 simulation runs. The return of the m-th portfolio strategy in the k-th simulation run is computed as $r_{k,m}^{Sim} = \hat{w}_{k,m}' r_k^{Sim}$, where r_k^{Sim} denotes the $\tau \times N$ matrix of the τ simulated returns in run k. Computing the returns for all $K = 1,000$ simulation runs, we obtain the $K \cdot \tau \times 1$ vector $R_m^{Sim} = \left[r_{1,m}^{Sim}; r_{2,m}^{Sim}; \ldots, r_{K,m}^{Sim}\right]$. Repeating the aforementioned step for all M strategies, we obtain the $K \cdot \tau \times M$ matrix of returns $R = \left[R_1^{Sim}, R_2^{Sim}, \ldots, R_M^{Sim}\right]$. Panels A to H report the correlation coefficient for estimation windows of $\tau = \{0.5, 1, 5, 10, 25, 50, 100, 200\}$ years, with each year comprising 252 simulated daily returns. The abbreviations for the strategies are explained in Table 1.

	Sample	LWNLS	1F	CC	ID	SC
Panel A: 0.5 years estimation window						
Sample	1.000	0.958	0.781	0.745	0.584	0.669
LWNLS		1.000	0.845	0.826	0.685	0.767
1F			1.000	0.954	0.635	0.735
CC				1.000	0.712	0.806
ID					1.000	0.972
SC						1.000
Panel B: 1 year estimation window						
Sample	1.000	0.986	0.828	0.788	0.618	0.703
LWNLS		1.000	0.860	0.831	0.675	0.756
1F			1.000	0.953	0.670	0.760
CC				1.000	0.749	0.830
ID					1.000	0.975
SC						1.000
Panel C: 5 years estimation window						
Sample	1.000	0.999	0.866	0.822	0.646	0.728
LWNLS		1.000	0.873	0.831	0.659	0.740
1F			1.000	0.952	0.705	0.783
CC				1.000	0.784	0.853
ID					1.000	0.979
SC						1.000
Panel D: 10 years estimation window						
Sample	1.000	1.000	0.871	0.826	0.649	0.731
LWNLS		1.000	0.874	0.831	0.656	0.737
1F			1.000	0.952	0.709	0.786
CC				1.000	0.789	0.855
ID					1.000	0.979
SC						1.000

Table continues on next page

Table A.3.1 continues here

	Sample	LWNLS	1F	CC	ID	SC
Panel E: 25 years estimation window						
Sample	1.000	1.000	0.874	0.829	0.651	0.733
LWNLS		1.000	0.875	0.831	0.654	0.735
1F			1.000	0.952	0.712	0.788
CC				1.000	0.791	0.857
ID					1.000	0.979
SC						1.000
Panel F: 50 years estimation window						
Sample	1.000	1.000	0.875	0.830	0.652	0.733
LWNLS		1.000	0.876	0.831	0.653	0.735
1F			1.000	0.952	0.713	0.789
CC				1.000	0.793	0.858
ID					1.000	0.979
SC						1.000
Panel G: 100 years estimation window						
Sample	1.000	1.000	0.876	0.830	0.652	0.734
LWNLS		1.000	0.876	0.831	0.653	0.734
1F			1.000	0.952	0.714	0.789
CC				1.000	0.793	0.859
ID					1.000	0.979
SC						1.000
Panel H: 200 years estimation window						
Sample	1.000	1.000	0.876	0.831	0.652	0.734
LWNLS		1.000	0.876	0.831	0.653	0.734
1F			1.000	0.952	0.714	0.789
CC				1.000	0.794	0.859
ID					1.000	0.979
SC						1.000

Table A.3.2: Descriptive statistics of estimation errors across estimation windows

This table shows the Bravais-Pearson correlation coefficients of estimation errors in portfolio weights of the single strategies. The estimation error for the m-th portfolio strategy in the k-th simulation run is defined as the difference between the estimated portfolio weights $\hat{w}_{k,m}$ and the true minimum-variance portfolio weights ω_k, i.e. $\Delta w_{k,m} = \hat{w}_{k,m} - \omega_k$. The correlation of estimation errors is computed as the correlation between estimation errors over all 1,000 simulation run. Panels A to H report the respective results for estimation windows of $\tau = \{0.5, 1, 5, 10, 25, 50, 100, 200\}$ years, with each year comprising 252 simulated daily returns. The abbreviations for the strategies are explained in Table 1.

	Sample	LWNLS	1F	CC	ID	SC
Panel A: 0.5 years estimation window						
Sample	1.000	0.711	0.286	0.189	0.005	0.077
LWNLS		1.000	0.695	0.657	0.473	0.544
1F			1.000	0.859	0.642	0.694
CC				1.000	0.747	0.809
ID					1.000	0.946
SC						1.000
Panel B: 1 year estimation window						
Sample	1.000	0.824	0.248	0.155	0.003	0.064
LWNLS		1.000	0.580	0.530	0.371	0.432
1F			1.000	0.840	0.654	0.690
CC				1.000	0.758	0.808
ID					1.000	0.950
SC						1.000
Panel C: 5 years estimation window						
Sample	1.000	0.949	0.135	0.076	-0.001	0.026
LWNLS		1.000	0.307	0.265	0.181	0.209
1F			1.000	0.820	0.677	0.692
CC				1.000	0.774	0.808
ID					1.000	0.956
SC						1.000
Panel D: 10 years estimation window						
Sample	1.000	0.970	0.098	0.056	-0.003	0.020
LWNLS		1.000	0.222	0.192	0.128	0.151
1F			1.000	0.816	0.680	0.693
CC				1.000	0.775	0.809
ID					1.000	0.957
SC						1.000

Table continues on next page

Table A.3.2 continues here

	Sample	LWNLS	1F	CC	ID	SC
Panel E: 25 years estimation window						
Sample	1.000	0.985	0.052	0.025	-0.012	0.001
LWNLS		1.000	0.134	0.115	0.074	0.087
1F			1.000	0.814	0.683	0.693
CC				1.000	0.777	0.808
ID					1.000	0.957
SC						1.000
Panel F: 50 years estimation window						
Sample	1.000	0.991	0.036	0.018	-0.010	-0.003
LWNLS		1.000	0.094	0.080	0.050	0.057
1F			1.000	0.814	0.685	0.694
CC				1.000	0.778	0.809
ID					1.000	0.957
SC						1.000
Panel G: 100 years estimation window						
Sample	1.000	0.995	0.026	0.014	-0.004	0.001
LWNLS		1.000	0.066	0.057	0.038	0.043
1F			1.000	0.813	0.685	0.694
CC				1.000	0.778	0.809
ID					1.000	0.958
SC						1.000
Panel H: 200 years estimation window						
Sample	1.000	0.997	0.023	0.011	-0.002	0.003
LWNLS		1.000	0.052	0.042	0.028	0.032
1F			1.000	0.813	0.685	0.694
CC				1.000	0.778	0.809
ID					1.000	0.957
SC						1.000

A.4 Additional empirical results

Table A.4.3: Average monthly turnover

This table reports the average monthly turnover in percent for our averaging rules with *variance* (AV-VAR), *ex-ante Sharpe ratio* (AV-SRA), and *ex-post Sharpe ratio* optimization (AV-SRP) using Sample and LWNLS as a baseline strategy in Panel A. The respective turnover of the single strategies and benchmarks are reported in Panel B and C, respectively. The average daily turnover of portfolio p is measured as the average percentage of total wealth traded each day, i.e. $TRN_p = \frac{1}{T-\tau-1} \sum_{t=\tau}^{T-1} \|\hat{w}_{p,\tau+1} - \hat{w}_{p,\tau+}\|_1$, where w_{p,τ^+} are the weights at the end of τ before rebalancing, $w_{p,\tau+1}$ are the congruent weights after rebalancing at time $\tau + 1$, and $||\cdot||_1$ is the one-norm, i.e. the sum of absolute values. We compute the average monthly turnover by multiplying TRN_p by 21, following the convention that a month comprises 21 trading days. The turnovers are computed over the out-of-sample period from June 1976 until June 2019.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR	56.93	74.03	98.87	114.45	111.45
AV-Sample-SRA	67.48	78.44	102.32	118.02	178.77
AV-Sample-SRP	62.69	73.80	94.47	109.25	146.92
AV-LWNLS-VAR	55.63	70.70	89.59	101.45	107.03
AV-LWNLS-SRA	62.66	72.02	83.12	81.86	89.02
AV-LWNLS-SRP	58.95	66.10	77.58	76.18	82.28
Panel B: Single strategies					
Sample	89.96	137.74	255.94	450.18	909.22
LWNLS	73.73	96.69	121.36	135.33	142.84
1F	62.97	74.29	80.54	82.51	83.00
CC	57.89	72.57	87.25	92.07	93.88
ID	20.22	21.09	22.34	22.86	23.34
SC	35.32	39.57	44.60	47.45	49.90
Panel C: Benchmarks					
LW1F	74.35	97.71	128.60	150.91	165.37
LWID	75.88	110.44	175.51	233.83	277.92
LWCC	68.18	94.22	135.95	169.45	195.28
BBCVC	58.53	81.41	125.90	162.67	187.31
KDM	62.97	74.29	80.54	82.51	83.00

Table A.4.4: Average daily short interest

This table reports the average monthly short interest in percent for our averaging rules with *variance* (AV-VAR), *ex-ante Sharpe ratio* (AV-SRA), and *ex-post Sharpe ratio* optimization (AV-SRP) using *Sample* and *LWNLS* as a baseline strategy in Panel A. The respective short interest of the single strategies and benchmarks are reported in Panels B and C, respectively. The average daily short interest of portfolio p is measured as the average amount of wealth that is held in short positions, i.e. $SI_p = \frac{1}{T-\tau}\sum_{t=\tau}^{T}(\|\hat{w}_{p,t}\|_1 - 1)/2$, where $w_{p,t}$ are the portfolio weights at t. The short interest is computed over the out-of-sample period from June 1976 until June 2019. The abbreviations for the averaging rules, strategies, and benchmarks are explained in Table 1.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR	21.59	30.47	48.38	49.84	44.46
AV-Sample-SRA	26.01	28.16	62.12	56.07	94.02
AV-Sample-SRP	22.39	26.80	48.10	50.36	77.41
AV-LWNLS-VAR	20.03	29.41	48.37	54.59	56.06
AV-LWNLS-SRA	19.94	24.25	36.67	40.61	46.79
AV-LWNLS-SRP	16.06	23.45	36.73	37.56	39.89
Panel B: Single strategies					
LWNLS	42.03	58.29	78.95	88.98	94.71
Sample	54.45	85.08	159.24	263.36	446.37
1F	44.36	55.10	61.92	64.54	64.48
CC	41.48	57.02	71.99	78.08	80.20
ID	0.00	0.00	0.00	0.00	0.00
SC	0.00	0.00	0.00	0.00	0.00
Panel C: Benchmarks					
LW1F	44.55	61.33	85.04	100.09	110.73
LWID	43.11	64.99	108.40	145.27	168.23
LWCC	40.07	60.61	92.30	115.71	134.65
BBCVC	25.84	43.26	68.03	94.13	106.49
KDM	44.36	55.10	61.92	64.54	64.48

Table A.4.5: Empirical out-of-sample standard deviations for averaging rules with exponentially weighted jackknife returns

This table reports the annualized out-of-sample standard deviations of our averaging rules with *variance* (AV-Var-E), *ex-ante Sharpe ratio* (AV-SRA-E), and *ex-post Sharpe ratio* optimization (AV-SRP-E) based on *exponentially* smoothed jackknife returns using Sample and LWNLS as a baseline strategy in Panel A. The respective standard deviations of the single strategies and benchmarks are reported in Panel B and C, respectively. The out-of-sample period for all five data sets, comprising the largest single stocks $N = \{30, 50, 100, 150, 200\}$ in the U.S. by June each year, ranges from June 1976 until June 2019. We report statistical significance for the null hypothesis that the log out-of-sample variance of each averaging rule including the Sample (LWNLS) strategy is equal to or higher than the log out-of-sample variance of the Sample (LWNLS) strategy. All p-values are based on the HAC inference of Ledoit and Wolf (2011). We mark the portfolio with the lowest standard deviation for each estimation window bold.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR-E	13.51	12.73	11.20	10.70	10.38
	0.0000	*0.0000*	*0.0000*	*0.0000*	*0.0000*
AV-Sample-SRA-E	14.96	14.68	13.65	13.48	13.24
	0.9999	*1.0000*	*1.0000*	*0.0024*	*0.0000*
AV-Sample-SRP-E	14.97	14.63	13.76	13.66	13.46
	1.0000	*1.0000*	*1.0000*	*0.0082*	*0.0000*
AV-LWNLS-VAR-E	**13.50**	**12.72**	**11.12**	**10.47**	**10.01**
	0.0000	*0.0000*	*0.0000*	*0.0000*	*0.0000*
AV-LWNLS-SRA-E	14.99	14.75	13.79	13.47	13.03
	1.0000	*1.0000*	*1.0000*	*1.0000*	*1.0000*
AV-LWNLS-SRP-E	14.95	14.63	13.76	13.47	13.08
	1.0000	*1.0000*	*1.0000*	*1.0000*	*1.0000*
Panel B: Strategies					
Sample	14.00	13.48	12.87	14.18	17.70
LWNLS	13.73	12.96	11.31	10.66	10.20
1F	14.89	14.36	12.89	12.37	12.10
CC	14.95	15.07	14.46	14.57	14.37
ID	17.40	17.41	17.52	17.28	17.08
SC	13.78	13.21	12.30	11.76	11.34
Panel C: Benchmarks					
LW1F	13.74	12.94	11.20	10.60	10.10
LWID	13.75	13.08	11.78	11.52	11.23
LWCC	13.68	12.98	11.33	10.89	10.47
BBCVC	13.73	13.04	11.67	11.13	10.76
KDM	14.89	14.36	12.89	12.37	12.10

Table A.4.6: Empirical out-of-sample Sharpe ratios for averaging rules with exponentially weighted jackknife returns

This table reports the annualized out-of-sample Sharpe ratio of our averaging rules with *variance* (AV-VAR-E), *ex-ante Sharpe ratio* (AV-SRA-E), and *ex-post Sharpe ratio* optimization (AV-SRP-E) based on *exponentially* smoothed jackknife returns using Sample and LWNLS as a baseline strategy in Panel A. The respective Sharpe ratios of the single strategies and benchmarks are reported in Panel B and C, respectively. The out-of-sample period for all five data sets, comprising the largest single stocks $N = \{30, 50, 100, 150, 200\}$ in the U.S. by June each year, ranges from June 1976 until June 2019. We report statistical significance for the null hypothesis that the Sharpe ratio of each averaging rule including the Sample (LWNLS) strategy is equal to or less than the Sharpe ratio of the Sample (LWNLS) strategy. All p-values are based on the HAC inference of Ledoit and Wolf (2008). We mark the portfolio with the highest Sharpe ratio for each estimation window bold.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR-E	0.529	0.564	0.510	0.562	0.590
	0.1357	*0.0394*	*0.0933*	*0.0020*	*0.0627*
AV-Sample-SRA-E	0.545	0.566	0.540	0.602	0.522
	0.2567	*0.2011*	*0.1500*	*0.0100*	*0.1705*
AV-Sample-SRP-E	0.581	0.590	**0.576**	0.643	0.524
	0.1497	*0.1379*	*0.0896*	*0.0039*	*0.1670*
AV-LWNLS-VAR-E	0.526	0.555	0.502	0.570	0.601
	0.4850	*0.1656*	*0.4104*	*0.1946*	*0.3805*
AV-LWNLS-SRA-E	0.538	0.546	0.531	0.585	0.597
	0.4380	*0.4142*	*0.3710*	*0.3429*	*0.4740*
AV-LWNLS-SRP-E	**0.594**	**0.595**	0.569	**0.657**	**0.638**
	0.2180	*0.2345*	*0.2465*	*0.1468*	*0.3327*
Panel B: Strategies					
Sample	0.486	0.479	0.410	0.266	0.372
LWNLS	0.525	0.525	0.494	0.538	0.589
1F	0.455	0.491	0.485	0.508	0.499
CC	0.464	0.523	0.456	0.440	0.475
ID	0.500	0.494	0.493	0.515	0.524
SC	0.513	0.543	0.501	0.576	0.567
Panel C: Benchmarks					
LW1F	0.501	0.521	0.516	0.549	0.620
LWID	0.505	0.505	0.455	0.437	0.546
LWCC	0.514	0.519	0.507	0.494	0.578
BBCVC	0.556	0.537	0.467	0.534	0.583
KDM	0.455	0.491	0.485	0.508	0.499

Table A.4.7: Empirical out-of-sample Sharpe ratios adjusted for transaction costs of 15bps for averaging rules with exponentially weighted jackknife returns

This table reports the annualized out-of-sample Sharpe ratio after transaction costs of 15 basis points for our averaging rules with *variance* (AV-VAR-E), *ex-ante Sharpe ratio* (AV-SRA-E), and *ex-post Sharpe ratio* optimization (AV-SRP-E) based on *exponentially* smoothed jackknife returns using either Sample or LWNLS as a baseline strategy in Panel A. The respective Sharpe ratios of the single strategies and benchmarks are reported in Panel B and C, respectively. The out-of-sample period for all five data sets, comprising the largest single stocks $N = \{30, 50, 100, 150, 200\}$ in the U.S. by June each year, ranges from June 1976 until June 2019. We report statistical significance for the null hypothesis that the Sharpe ratio of each averaging rule including the Sample (LWNLS) strategy is equal to or less than the Sharpe ratio of the Sample (LWNLS) strategy. All p-values are based on the HAC inference of Ledoit and Wolf (2008). We mark the portfolio with the highest Sharpe ratio for each estimation window bold.

N	30	50	100	150	200
Panel A: Averaging rules					
AV-Sample-VAR-E	0.453	0.459	0.351	0.369	0.396
	0.0179	*0.0004*	*0.0001*	*0.0000*	*0.0000*
AV-Sample-SRA-E	0.464	0.470	0.404	0.444	0.277
	0.1514	*0.0476*	*0.0026*	*0.0000*	*0.0000*
AV-Sample-SRP-E	0.504	0.499	0.453	0.497	0.323
	0.0724	*0.0235*	*0.0007*	*0.0000*	*0.0000*
AV-LWNLS-VAR-E	0.452	0.455	0.356	0.395	0.408
	0.2172	*0.0177*	*0.0528*	*0.0100*	*0.0366*
AV-LWNLS-SRA-E	0.463	0.458	0.422	0.476	0.473
	0.3438	*0.2439*	*0.1404*	*0.0782*	*0.1272*
AV-LWNLS-SRP-E	**0.524**	**0.513**	0.467	**0.554**	**0.524**
	0.1420	*0.1029*	*0.0633*	*0.0152*	*0.0500*
Panel B: Strategies					
Sample	0.370	0.295	0.052	-0.304	-0.543
LWNLS	0.428	0.391	0.300	0.309	0.336
1F	0.379	0.398	0.372	0.388	0.376
CC	0.395	0.436	0.347	0.326	0.357
ID	0.479	0.472	**0.470**	0.491	0.499
SC	0.467	0.489	0.435	0.504	0.488
Panel C: Benchmarks					
LW1F	0.403	0.385	0.309	0.292	0.324
LWID	0.406	0.353	0.186	0.072	0.100
LWCC	0.424	0.389	0.291	0.214	0.242
BBCVC	0.479	0.424	0.273	0.271	0.269
KDM	0.379	0.398	0.372	0.388	0.376

References Online Appendix

Basak, G. K., R. Jagannathan, and T. Ma (2009). Electronic companion: "jackknife estimator for tracking errors".

Chan, L. K., J. Karceski, and J. Lakonishok (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. Review of Financial Studies 12(5), 937–974.

De Nard, G. (2022). Oops! I Shrunk the Sample Covariance Matrix Again: Blockbuster Meets Shrinkage. Journal of Financial Econometrics 20(4), 569–611.

Elton, E. J. and M. J. Gruber (1973). Estimating the dependence structure of share prices - implications for portfolio selection. Journal of Finance 28(5), 1203–1232.

Horn, R. and C. Johnson (2013). Matrix Analysis (2nd ed.). Cambridge University Press.

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

Kourtis, A., G. Dotsis, and R. N. Markellos (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. Journal of Banking and Finance 36(9), 2522–2531.

Ledoit, O. and S. Péché (2011). Eigenvectors of some large sample covariance matrix ensembles. Probability Theory and Related Fields 151, 233—264.

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

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

Ledoit, O. and M. Wolf (2004b). Honey, I shrunk the sample covariance matrix. Journal of Portfolio Management 30(4), 110–119.

Ledoit, O. and M. Wolf (2015). Spectrum estimation: A unified framework for covariance matrix estimation and pca in large dimensions. Journal of Multivariate Analysis 139, 360–384.

Ledoit, O. and M. Wolf (2017). Numerical implementation of the quest function. Computational Statistics and Data Analysis 115, 199–223.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. Management Science 9(2), 277–293.

White, H. (2000). Asymptotic Theory for Econometricians (2nd ed.). Emerald Publishing Limited.