Diversifying estimation errors: An efficient averaging rule for portfolio optimization

May 5, 2021

Abstract

We propose an averaging rule that combines established minimum-variance strategies to minimize the expected out-of-sample variance. Our rule overcomes the problem of selecting the "best" strategy ex-ante and diversifies remaining estimation errors of the strategies included in the averaging. Extensive simulations show that the contributions of estimation errors to the out-of-sample variances are uncorrelated between the considered strategies. This implies that averaging over multiple strategies offers sizable diversification benefits. Across all data sets we find that our rule achieves a significantly lower out-of-sample standard deviation than any competing strategy and that the Sharpe ratio is at least 25% higher than for the 1/N portfolio.

JEL Classification: G11
Keywords: Averaging; diversification; estimation error; portfolio optimization; shrinkage.

1 Introduction

Existing literature on portfolio selection offers 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, 2017, 2020), the inverse covariance (see, e.g., Kourtis et al., 2012; DeMiguel et al., 2013; Ledoit and Wolf, 2018; Shi et al., 2020) and the 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 in order to regularize portfolio weights (see, e.g., Brodie et al., 2009; DeMiguel et al., 2009; Xing et al., 2014; Li, 2015; Yen, 2016). The plethora of evolving strategies challenges practitioners and academics alike to select one of the proposed strategies. In addition to the challenge of identifying the "best" strategy ex-ante, the limitation to a single strategy may forego potential diversification benefits. If estimation errors are not perfectly correlated across individual strategies, a combination of different approaches offers the potential to further diversify estimation errors. The resulting portfolio weights are less prone to estimation errors and achieve, on average, an improved out-of-sample performance.

In this paper, we propose an averaging rule that overcomes the strategy selection problem and exploits the potential to diversify estimation errors between individual strategies. Thus, our approach takes the principle of diversification from the portfolio to the estimation level. Following the consensus in the literature, we apply our averaging rule to minimum-variance portfolios because expected returns are notoriously difficult to estimate (see, e.g., Merton, 1980).

Our rule aims to minimize the expected out-of-sample portfolio variance. For this purpose and to reduce the impact of estimation errors, we combine established approaches on three different levels: *i)* the covariance level, where averaging is applied before one computes the inverse of the covariance matrix, *ii)* the inverse covariance level, where averaging is applied on the individual inverses of the various covariance matrices, and *iii)* the portfolio weight level, where averaging is applied directly on the portfolio weights. Our averaging rule combines the unbiased sample estimator with a multitude of structured estimators.[1] Thus, our approach

[1]Throughout the paper, we average over three different levels (covariance, inverse, and weight level), use the Sample and the structured (target) estimators (jointly denoted as strategies) for averaging on the aforementioned levels, and benchmark our rule against existing shrinkage approaches (denoted as benchmarks).

extends the idea of shrinkage from a single to a multi-target specification.

In extensive simulations, we find that estimation errors between established minimum-variance strategies are uncorrelated. This correlation structure offers sizable and persistent diversification benefits. Our averaging rule utilizes this structure to diversify estimation errors, leading to a decrease in the out-of-sample standard deviation as the number of strategies increases. The average out-of-sample standard deviation of our rule is lower than those of any single strategy over all simulations. Our averaging rule also achieves higher Sharpe ratios for smaller estimation windows. In addition, we find that our rule compares favorably to the selected benchmarks, not only in terms of lower out-of-sample standard deviation, but also higher out-of-sample Sharpe ratio. This includes the most recent non-linear shrinkage approach of Ledoit and Wolf (2017).

Empirical results on five data sets confirm the findings from the simulation study in which our rule achieves a lower out-of-sample standard deviation than any single strategy. The standard deviation reduction in comparison to the non-linear shrinkage strategy of Ledoit and Wolf (2017) is statistically significant on all five data sets. Our rule achieves the second highest Sharpe ratio on all data sets in comparison to the considered strategies. Relative to the 1/N strategy, the out-of-sample Sharpe ratio of our rule is on all data sets at least 25% higher, being statistically significant in four out of five sets. The turnover and short interest of our averaging rule is modest. These properties make our averaging rule both statistically and economically appealing.

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 the combination of multiple targets for the shrinkage estimation of covariance matrices. 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, exploring whether estimation errors can be diversified more comprehensively if multiple resources for the diversification are taken into account.

We contribute to the existing literature in two ways: First, in a battery of simulations, we find that estimation errors between a variety of established minimum-variance strategies are uncorrelated which is undocumented and unexploited so far. The resulting diversification benefits are sizable and persistent for different estimation windows. Second, we provide a framework for combining multiple strategies to alleviate the impact of estimation errors. Our averaging rule builds on the principle of diversification, is easy to calibrate, and extends to an arbitrary number of strategies, not limited to those selected for this paper. On simulated and empirical data our rule compares favorably to selected benchmarks in terms of out-of-sample standard deviation and Sharpe ratio. Thus, averaging over multiple strategies provides a promising avenue to reduce the impact of estimation errors on out-of-sample portfolio performance.

The paper is organized as follows. Section 2 reviews the considered minimum-variance strategies and introduces our averaging rule. 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 Portfolio optimization and estimation errors

2.1 Minimum-variance portfolio optimization

Throughout the paper we focus on the estimation of the minimum-variance portfolio resulting from the following optimization problem:

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ \text{s.t.} \quad & w'\mathbf{1}_N = 1, \end{aligned} \tag{1}$$

where Σ is the true but unknown covariance matrix, $\mathbf{1}_N$ is a N-dimensional vector of ones, and N is the number of investable risky assets, respectively. The allocation to the N assets within the minimum-variance portfolio results from solving the problem in Equation (1), with the optimal portfolio weights given by:

$$w = \Sigma^{-1}\mathbf{1}_N(\mathbf{1}_N'\Sigma^{-1}\mathbf{1}_N)^{-1}. \tag{2}$$

Because w 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 $\hat{\Sigma}$ or its inverse, translate into sub-optimal allocations with an adverse effect on the portfolio performance.

2.2 Alleviating estimation errors: Existing approaches

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 literature suggests structured estimators of the covariance matrix. Examples include the constant correlation model (see, e.g., Elton and Gruber, 1973) and (approximate) factor models (see, e.g., Chan et al., 1999; Fan et al., 2013). The second branch proposes shrinkage estimators of the covariance matrix (see, e.g., Ledoit and Wolf, 2003, 2004a,b, 2017, 2020), its inverse (see, e.g., Kourtis et al., 2012; DeMiguel et al., 2013; Ledoit and Wolf, 2018; Shi et al., 2020), and portfolio weights (see, e.g., Kan and Zhou, 2007; Tu and Zhou, 2011; DeMiguel et al., 2013). The third branch regularizes portfolio weights through the imposition of weight constraints. Frost and Savarino (1986) as well as Jagannathan and Ma (2003) study the effects of short-sale constraints. Brodie et al. (2009), DeMiguel et al. (2009), Xing et al. (2014), Li (2015), and Yen (2016) suggest the imposition of norm constraints on portfolio weights.

Throughout the paper we report results for eleven minimum-variance strategies, covering at least one representative of the aforementioned branches from the existing literature. Each of the minimum-variance strategies can be represented by an estimator of the covariance matrix or its inverse. The portfolio weights of these strategies are given by Equation (2), with $\hat{\Sigma}$ being one of the following estimators, clustered into two groups.[2] The first group comprises the sample as well as structured estimates of the covariance matrix and represents the building blocks for our averaging rule:

- Sample: The (unbiased) sample covariance matrix, $\hat{\Sigma}_S$.
- 1F: The covariance matrix implied by the single factor model of Sharpe (1963), $\hat{\Sigma}_{1F}$.
- CC: The covariance matrix recovered from a constant correlation matrix with the average of all pairwise correlations off the main diagonal and the assets' individual standard deviations, $\hat{\Sigma}_{CC}$ (Elton and Gruber, 1973).

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

- ID: A diagonal matrix with the average variance of all assets on the main diagonal, $\hat{\Sigma}_{ID}$. The resulting minimum-variance portfolio corresponds to the 1/N portfolio, which represents a benchmark in terms of risk-adjusted performance (DeMiguel et al., 2009).

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

- POET: The covariance matrix based on the approximate factor model by Fan et al. (2013) with risk factors defined by the principal components of the sample covariance matrix, $\hat{\Sigma}_{POET}$. We use the first three principal components throughout the paper.

The second group of estimators covers combinations of the sample and another covariance estimator serving as benchmarks for our averaging rule:

- LW1F: A linear combination of $\hat{\Sigma}_S$ and $\hat{\Sigma}_{1F}$. The combination intensity is determined by the linear shrinkage estimator of Ledoit and Wolf (2003).

- LWID: A linear combination of $\hat{\Sigma}_S$ and $\hat{\Sigma}_{ID}$. The combination intensity is determined by the linear shrinkage estimator of Ledoit and Wolf (2004a).

- LWCC: A linear combination of $\hat{\Sigma}_S$ and $\hat{\Sigma}_{CC}$. The combination intensity is determined by the linear shrinkage estimator of Ledoit and Wolf (2004b).

- LWNLS: A non-linear shrinkage estimator, correcting for over-dispersed eigenvalues of the sample covariance matrix following Ledoit and Wolf (2017). This approach represents our benchmark in terms of out-of-sample variance.

- KDM: A linear combination of $\hat{\Sigma}_S^{-1}$, identity I, and $\hat{\Sigma}_{1F}^{-1}$. The combination intensity is determined following Kourtis et al. (2012).

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.

2.3 Estimation error diversification: An averaging approach

Our approach to estimation error diversification generalizes the idea of shrinkage. We suggest blending the sample with a multitude of structured estimators. If estimation errors can be reduced by combining the sample with one structured estimator, averaging over multiple estimators should offer further diversification potential. In essence, our approach brings the principle of diversification from the portfolio to the estimation level.

We suggest an averaging rule that combines individual estimators with the objective of minimizing the expected out-of-sample variance of the corresponding minimum-variance portfolio.[3] Because there is no consensus in the literature whether estimation errors can most efficiently be diversified on the covariance, inverse or portfolio weight level, we consider averaging over estimators on the three aforementioned levels. We assign the averaging intensities to the considered estimators on the respective level such that the expected out-of-sample variance of the corresponding minimum-variance portfolio is minimized:

$$\min_{\lambda_{AV-Level}} \quad E\left(\hat{w}_{AV-Level}\left(\lambda_{AV-Level}\right)' \Sigma \hat{w}_{AV-Level}\left(\lambda_{AV-Level}\right)\right) \tag{3}$$

$$\text{s.t.} \quad \lambda'_{AV-Level}\mathbf{1}_M = 1 \tag{4}$$

$$\lambda_{AV-Level} \geq 0, \tag{5}$$

where $\lambda_{AV-Level}$ are the averaging intensities on the respective averaging level, $\hat{w}_{AV-Level}$ are the corresponding portfolio weight estimates, $\mathbf{1}_M$ is a M-dimensional vector of ones, and M is the number of estimators. Equations (3) - (5) correspond to a minimum-variance portfolio optimization problem with non-negativity restrictions on the averaging intensities. We impose the non-negativity constraint because Jagannathan and Ma (2003) demonstrate that it has a shrinkage-like effect, reduces estimation errors, and improves the portfolio performance. The literature on linear regression model averaging also finds that non-negativity constraints enhance the quality of the respective combination (see, e.g., Timmermann, 2006; Hansen, 2008).[4]

If averaging is conducted on the covariance or inverse level, we combine the M estimators of the covariance matrix or its inverse in a first step and compute the corresponding minimum-

[3] We do not claim that other, potentially more sophisticated averaging rules may deliver improved results. We rather advocate in a comparatively simple setting that averaging over individual established approaches provides a promising avenue to alleviate the impact of estimation errors on portfolio performance.

[4] Britten-Jones (1999) shows that the mean-variance portfolio optimization problem can be formulated as a linear regression problem.

variance portfolio weights in a second step. Following the objective outlined in Equation (3), we combine the individual estimators such that the expected out-of-sample variance of the corresponding minimum-variance portfolio is minimized. We approximate the expected out-of-sample variance based on a jackknife procedure, considering a one-month holding period throughout the paper (see, e.g., Ledoit and Wolf, 2017).

We suggest two jackknife approximations of the expected out-of-sample variance. The first specification assumes that portfolio returns are independent and identically distributed (*i.i.d.*). In this case, we compute the expected out-of-sample variance as follows:

$$E(\hat{w}_{AV-Level}(\lambda_{AV-Level})'\Sigma\hat{w}_{AV-Level}(\lambda_{AV-Level})) = \sum_{i=1}^{r}(r_{i,AV-Level}^{JK} - \bar{r}_{AV-Level}^{JK})^2, \tag{6}$$
$$\text{with } \bar{r}_{AV-Level}^{JK} = \frac{1}{\tau}\sum_{i=1}^{\tau} r_{i,AV-Level}^{JK},$$

where $r_{i,AV-Level}^{JK}$ is the jackknife return for the i-th observation and the considered averaging level, $AV-Level$. The second specification of our jackknife approximation assumes that portfolio returns exhibit time-series characteristics. We account for time-series patterns by exponentially weighting the jackknife returns for the approximation of the expected out-of-sample variance:

$$\begin{aligned} &E(\hat{w}_{AV-Level-E}(\lambda_{AV-Level-E})'\Sigma\hat{w}_{AV-Level-E}(\lambda_{AV-Level-E})) = \\ &\frac{\sum_{i=1}^{r} e^{\omega i}(r_{i,AV-Level-E}^{JK} - \bar{r}_{AV-Level-E}^{JK})^2}{\sum_{i=1}^{\tau} e^{\omega i}}, \\ &\text{with } \bar{r}_{AV-Level-E}^{JK} = \frac{\sum_{i=1}^{\tau} e^{\omega i} r_{i,AV-Level-E}^{JK}}{\sum_{i=1}^{\tau} e^{\omega i}}, \end{aligned} \tag{7}$$

where ω is the decay rate. We follow Basak et al. (2009) and set $\omega = 0.01$ when using daily data.[5]

We compute jackknife returns by dropping the excess returns falling into the i-th month from our in-sample period and compute the covariance or its inverse, respectively, using the remaining excess returns within the in-sample period. We denote the m-th estimator of the covariance or its inverse, by $\hat{\Sigma}_{-i,m}$ and $\hat{\Sigma}_{-i,m}^{-1}$, respectively. The minimum-variance portfolio weights corresponding to the combination of the M estimators of the covariance matrix,

[5]We set $\omega = 0.21$ for additional results with monthly data in Section A.4 of the Internet Appendix.

$\hat{w}_{-i,AV-Cov}$, and its inverse, $\hat{w}_{-i,AV-Inv}$, are given by:

$$\hat{w}_{-i,AV-Cov} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_{-i,m}\right)^{-1} \mathbf{1}_N \left(\mathbf{1}_N' \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_{-i,m}\right)^{-1} \mathbf{1}_N\right)^{-1},$$

$$\hat{w}_{-i,AV-Inv} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}_{-i,m}^{-1}\right) \mathbf{1}_N \left(\mathbf{1}_N' \sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}_{-i,m}^{-1} \mathbf{1}_N\right)^{-1},$$

where $\hat{\lambda}_{m,Cov}$ and $\hat{\lambda}_{m,Inv}$ are our estimates of the averaging intensities of the m-th strategy. We then compute the jackknife excess returns for the i-th month, $r_{i,Cov}^{JK}$ and $r_{i,Inv}^{JK}$, representing the excess returns that would have been achieved from holding $\hat{w}_{-i,Cov}$ and $\hat{w}_{-i,Inv}$, respectively, in the hold-out month i. Repeating these steps for all months within our in-sample period yields the time series of jackknifed portfolio excess returns. We use the sample variance of this time series as an estimate of the objective in Equation (3). The portfolio weights resulting from this averaging procedure are given by:

$$\hat{w}_{AV-Cov} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_{m}\right)^{-1} \mathbf{1}_N \left(\mathbf{1}_N' \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_{m}\right)^{-1} \mathbf{1}_N\right)^{-1}, \tag{8}$$

$$\hat{w}_{AV-Inv} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}_{m}^{-1}\right) \mathbf{1}_N \left(\mathbf{1}_N' \sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}_{m}^{-1} \mathbf{1}_N\right)^{-1}, \tag{9}$$

where $\hat{\Sigma}_m$ and $\hat{\Sigma}_m^{-1}$ are the m-th estimators of the covariance or its inverse, based on all in-sample observations, respectively.

If averaging is conducted on the portfolio weight level, we combine the M estimators of the minimum-variance portfolio weights such that the expected out-of-sample variance in Equation (3) is minimized. We follow the outlined jackknife procedure and drop the excess returns falling into the i-th month. Thereafter, we compute the minimum-variance portfolio weights $\hat{w}_{-i,m}$ for each of the M estimators. The minimum-variance portfolio weights of the combination, $\hat{w}_{-i,AV-Wgt}$, in the hold-out month i are given by:

$$\hat{w}_{-i,AV-Wgt} = \sum_{m=1}^{M} \hat{\lambda}_{m,Wgt} \cdot \hat{w}_{-i,m},$$

where $\hat{\lambda}_{Wgt}$ are the estimated averaging intensities. Repeating the procedure for all months and computing the sample variance of the time series of jackknife excess returns again gives the estimate of the expected out-of-sample variance in Equation (3). The portfolio weights corresponding to averaging on the weight level are given by:

$$\hat{w}_{AV-Wgt} = \sum_{m=1}^{M} \hat{\lambda}_{m,Wgt} \cdot \hat{w}_m. \tag{10}$$

Table 1 provides an overview of our proposed averaging rule on the three levels in Panel A and the strategies in Panel B. Panel C lists the five selected benchmarks of which LWNLS is the latest and most sophisticated, and thus our target to beat in terms of out-of-sample variance.

Throughout the paper we consider averaging over the sample and the 1F, CC, ID, SC, and POET estimates on the three aforementioned levels.[6] The selected set of estimators is, as any other set, an ad-hoc selection for which the motivation is twofold. First, we seek to build our averaging rule on an established set of simple estimators. Second, we want to include a sufficiently large number of estimators to explore the trade-off between a higher potential to diversify estimation errors and the newly arising estimation problem. Including more estimators in the averaging rule offers a greater potential to diversify estimation errors. Yet, including more strategies requires the estimation of more averaging intensities, giving rise to a new source of estimation errors. We explore this trade-off in the following simulation study.

[TABLE 1 ABOUT HERE]

3 Simulation study

3.1 Simulation procedure

The data generating process of our simulation study follows Tu and Zhou (2011). We consider $N = 25$ assets with the simulated excess return of security j at time t, $r_{j,t}$, coming from the Fama-French three-factor model with mispricing:

$$r_{j,t} = \alpha_j + \beta_{j,MKT}\, r_{t,MKT} + \beta_{j,SMB}\, r_{t,SMB} + \beta_{j,HML}\, r_{t,HML} + u_{j,t}, \tag{11}$$

[6]We do not include existing shrinkage strategies in our averaging rule, because the shrinkage intensities are estimated empirically and are thus inherently estimation error prone. Including shrinkage estimators in the averaging rule thus potentially induces an error-in-errors problem. We also do not include norm-constrained strategies in our paper because the moment-shrinkage representation of norm constraints does not hold for arbitrary levels of the constraint. Thus, it is not possible to use the aforementioned approaches on the three considered averaging levels.

where $\beta_{j,MKT}$, $\beta_{j,SMB}$, and $\beta_{j,HML}$ are the factor loadings, α_j is the mispricing factor, and $u_{j,t}$ is the residual at time t. The premia of the market $r_{t,MKT}$, size $r_{t,SMB}$, and value factor $r_{t,HML}$ at t follow a multivariate normal distribution. Table 2 shows the means and standard deviations of the factors as well as the correlation matrix between the factors.

[TABLE 2 ABOUT HERE]

The assets' factor loadings are randomly paired and evenly spread between 0.9 and 1.2 for β_{MKT}, -0.3 and 1.4 for β_{SMB}, -0.5 and 0.9 for β_{HML}, as well as -2.0% and 2.0% for the annualized mispricing factor α, in each simulation. The error term of each asset at t, $u_{j,t}$, 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 10% and an upper bound of 30%, such that the average idiosyncratic volatility in the cross-section equals 20%.

We run 10,000 simulations. In each run k we draw $\tau = \{60, 120, 240, 480, 960\}$ monthly observations from the outlined data generating process and save the population mean vector, μ_k, and the covariance matrix, Σ_k. These parameters result from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. We then estimate the portfolio weights of the m-th strategy, $\hat{w}_{k,m}$, based on the τ simulated excess returns and evaluate the estimated portfolio weights on the out-of-sample standard deviation, $\hat{\sigma}_{k,m}$, and Sharpe ratio, $\hat{\Psi}_{k,m}$, of this simulation run[7]:

$$\hat{\sigma}_{k,m} = \sqrt{\hat{w}'_{k,m}\Sigma_k \hat{w}_{k,m}}, \tag{12}$$

$$\hat{\Psi}_{k,m} = \frac{\hat{w}'_{k,m}\mu_k}{\sqrt{\hat{w}_{k,m}\Sigma_k \hat{w}_{k,m}}}. \tag{13}$$

3.2 Simulation results

In a first step, we evaluate the potential to diversify the estimation errors of the individual strategies. We quantify the contribution of the estimation errors to the out-of-sample variances of the m-th strategy in the k-th simulation run as the following loss:

$$L_{k,m} = \left(\hat{w}_{k,m} - w_k\right)' \Sigma_k \left(\hat{w}_{k,m} - w_k\right), \tag{14}$$

[7]Throughout the paper we report annualized variances and Sharpe ratios if not indicated otherwise.

where w_k are the true minimum-variance portfolio weights, which we compute using the population covariance matrix of the respective run, Σ_k. Table 3 reports the means, standard deviations, and correlation coefficients between the losses of the considered strategies over the 10,000 simulation runs. The means of all strategies decrease for increasing estimation windows in Panels A - E, except for the ID strategy. As expected, this decrease is most pronounced for the unbiased Sample strategy. The ID strategy corresponds to the 1/N portfolio, such that the means and standard deviations are constant over the estimation windows. The standard deviations of the other strategies decrease with increasing sample sizes, being highest for the Sample for $\tau = 60$ and smallest for $\tau = 960$. Most importantly, the correlations are small in absolute terms, offering sizable diversification opportunities. The average pairwise correlation reduces from an already low level of 0.173 to 0.046 going from $\tau = 60$ to $\tau = 960$. The minimum and maximum pairwise correlation between the considered strategies range from -0.345 to 0.745 and -0.392 to 0.394 for $\tau = 60$ and $\tau = 960$, respectively. The correlation is the highest between the two (approximate) factor model strategies POET and 1F, highlighting the need for heterogeneous strategies to lever the diversification potential. In contrast to existing linear shrinkage approaches that combine the Sample with one or two other strategies (see, e.g., Ledoit and Wolf, 2003, 2004a,b; Kourtis et al., 2012), our results suggest that combining a multitude of strategies offers sizable diversification benefits. Yet, averaging over multiple strategies gives rise to a new estimation problem, resulting in the following trade-off: on the one hand, increasing the number of strategies improves the diversification of the estimation errors; on the other hand, it increases the number of averaging intensities that need to be estimated, creating a new source of estimation errors.

[TABLE 3 ABOUT HERE]

We evaluate this trade-off in Table 4 by applying our averaging rule to the Sample and an increasing number of strategies from Panel B of Table 1. Importantly, we consider all possible permutations that result from combining the Sample with the respective subset of strategies.[8] For each permutation in each simulation run we compute the averaging intensities according to our averaging rule and obtain the resulting portfolio weights.[9] We then evaluate the estimated portfolio weights based on their out-of-sample standard deviation according to Equation (12).

[8]When averaging over the sample and another strategy, we have $M = 2$, which gives rise to five permutations. The number of permutations for $M = 3$ and $M = 4$ amounts to ten, respectively, for $M = 5$ to five, and for $M = 6$ to one.

[9]We only consider *i.i.d.* weighted jackknife returns in our averaging rule as the data generating process in the simulation is also *i.i.d.*

Table 4 shows the annualized average out-of-sample standard deviations over all simulations and permutations for the corresponding number of strategies. Panels A-E present the results for the different estimation windows $\tau = \{60, 120, 240, 480, 960\}$, when averaging is conducted on the portfolio weight (AV-Wgt), the inverse (AV-Inv), or the covariance (AV-Cov) level. Our results show that the average out-of-sample standard deviation decreases as the number of strategies increases.[10] This result holds irrespective of the estimation window and the averaging level, demonstrating that the diversification benefits outweigh the newly created estimation error problem.

[TABLE 4 ABOUT HERE]

Investigating the relevance of each strategy within our averaging rule, we turn to Figure 1. This figure shows the mean averaging intensities of the considered strategies over all simulations when averaging is conducted over $M = 6$ strategies. Panel A displays the mean averaging intensities for AV-Wgt, Panels B and C show the respective weights for AV-Inv and AV-Cov. We observe similar patterns for the three averaging levels. Our rule utilizes, on average, the multitude of strategies, corroborating their relevance for an efficient diversification of estimation errors. Turning to the weighting of the individual strategies, the weight of the Sample is low when the estimation window is small and vice versa. This is in line with our results in Table 3. The average loss of the Sample is high (low) for smaller (higher) estimation windows, offering higher (lower) diversification gains. These higher (lower) gains are leveraged by assigning higher (lower) weights to the remaining strategies. This is in line with Table 3 where the average loss of the Sample decreases as the estimation window increases.

[FIGURE 1 ABOUT HERE]

In the context of our proposed rule, the previous findings demonstrate that averaging over a greater number of strategies is superior to averaging over fewer. This result holds irrespective of the averaging level as well as the estimation window. We further find that averaging over more strategies is not only beneficial with respect to the out-of-sample standard deviation, but

[10]We evaluate the distribution of the out-of-sample standard deviations over all simulation runs and permutations for AV-Wgt, AV-Inv, and AV-Cov using kernel densities in the left column of Figures A.2.1, A.2.2, and A.2.3 of the Internet Appendix. The distributions for $M = 6$ center around smaller out-of-sample standard deviations and collapse more tightly around this value than the distributions for smaller values of M. This result holds for all averaging levels and supports the finding that combining all six strategies is beneficial when compared to averaging over a subset.

also in terms of the out-of-sample Sharpe ratio.[11] Thus, for the rest of our paper we apply our rule to the Sample and the remaining five strategies in Panel B of Table 1, setting $M = 6$.

[TABLE 5 ABOUT HERE]

We compare the out-of-sample standard deviations of our averaging rule against the established strategies in Table 5. Our rule achieves, on average, lower out-of-sample standard deviations than the considered strategies over all estimation windows. This finding again holds for all averaging levels. We note that the out-of-sample standard deviations of our rule are, on average, lower than any of the sophisticated benchmarks in Panel C.[12]

[TABLE 6 ABOUT HERE]

Table 6 reports the risk-adjusted performance of our rule in comparison to the established strategies. AV-Wgt delivers, on average, slightly higher Sharpe ratios than AV-Inv and AV-Cov. The Sharpe ratios of AV-Wgt for $\tau = 60$ and 120 are higher than the Sharpe ratios of any strategy in Panel B. We find that only SC delivers, on average, a higher Sharpe ratio for the remaining estimation windows. The average Sharpe ratios of AV-Wgt in comparison to the benchmarks in Panel C, are higher than those of the considered strategies across all estimation windows. These results suggest that our rule compares favorably to the existing strategies not only with respect to the standard deviation, but also in terms of the Sharpe ratio.[13]

Our simulation study has three key results: First, the contributions of the estimation errors to the out-of-sample variances are uncorrelated between the considered strategies, offering sizable diversification benefits. This suggests that averaging over a multitude of strategies reduces the overall impact of estimation errors. Second, combining more strategies yields a lower standard deviation and a higher risk-adjusted performance. The performance across the averaging levels

[11] We report the results for the average out-of-sample Sharpe ratio in Table A.2.1 of the Internet Appendix. We find that the average Sharpe ratio increases with the number of strategies. We investigate the distribution of the out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov using kernel densities and report the results in the right column of Figures A.2.1, A.2.2, and A.2.3 of the Internet Appendix. We find that the distributions for $M = 6$ are centered around a higher mean value and collapse for estimation windows up to $\tau = 120$ more tightly around the mean value. This effect diminishes as the estimation window increases.

[12] We investigate the out-of-sample standard deviations of our rule in comparison to the existing strategies in more detail, exploring the kernel densities for AV-Wgt, AV-Inv, and AV-Cov in Figures A.2.4, A.2.5, and A.2.6 of the Internet Appendix. The left column in each figure shows the densities in comparison to the considered strategies. The kernel densities of each averaging level of our rule also compare favorably to the benchmarks. We find that the distributions of AV-Wgt, AV-Inv, and AV-Cov collapse as tightly as the competing benchmarks, but around a smaller mean value. This observation holds for estimation windows of up to $\tau = 480$.

[13] The kernel densities of the out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov in Figures A.2.7, A.2.8, and A.2.9 of the Internet Appendix, corroborate the favorable comparison relative to the existing strategies. The aforementioned figures present the results in comparison to the considered strategies in the left column and the benchmarks in the right column.

is similar, but slightly in favor of the portfolio weight level. We thus use AV-Wgt as the reference strategy in the empirical section of this paper. Third, we find that our rule compares favorably to the set of established strategies, suggesting that the combination of existing strategies represents an efficient way to alleviate the impact of estimation errors on out-of-sample performance.

4 Empirical study

4.1 Data

We evaluate the out-of-sample performance of our averaging rule in comparison to the eleven competing strategies on five empirical data sets based on daily excess returns.[14] Table 7 provides an overview of the data sets and their source. We use the Fama-French 6- and 25-factor as well as the 10- and 30-industry portfolios from Kenneth French's homepage. The construction of the STOCK500 data set follows Chan et al. (1999), Jagannathan and Ma (2003), and DeMiguel et al. (2009). In July of each year, we select the 500 largest stocks in terms of market capitalization 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 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 756 trading days preceding, or the 252 days subsequent to the selection date.[15] The selected stocks constitute the investment universe for one year, which is then revised based on the outlined criteria.

[TABLE 7 ABOUT HERE]

We compute the out-of-sample excess returns for each portfolio strategy from Table 1 using the common rolling-sample procedure, setting $\tau = 756$ observations. Starting in June 1973, we estimate the portfolio weights of each portfolio strategy 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

[14] Our choice of a daily data frequency follows Ledoit and Wolf (2017) and is also in line with industry standards. We report empirical results on a monthly data frequency in Tables A.4.2 - A.4.5 of the Internet Appendix. We find that the results based on monthly excess returns are qualitatively similar to the ones presented in this section.

[15] We use the common convention that a calendar month comprises 21 trading days throughout the empirical section of this paper. The 252 and 756 trading days correspond to one and three years of data.

trading days and repeat the aforementioned procedure over the entire sample period.

4.2 Performance evaluation

We compare our averaging rule with each competing portfolio strategy in terms of standard deviation (s) and Sharpe ratio (SR) of out-of-sample excess returns, the average portfolio turnover (TRN), and the average short interest (SI). The respective metrics for the i-th strategy are the same as used by DeMiguel et al. (2013) and given by:

$$s_i = \sqrt{\frac{1}{T-\tau-1}\sum_{t=\tau}^{T-1}\left(\hat{w}'_{i,t}r_{t+1} - \bar{r}_i\right)^2}, \text{ with } \bar{r}_i = \frac{1}{T-\tau}\sum_{t=\tau}^{T-1}\hat{w}'_{i,t}r_{t+1}, \tag{15}$$

$$SR_i = \frac{\bar{r}_i}{s_i}, \tag{16}$$

$$TRN_i = \frac{1}{T-\tau-1}\sum_{t=1}^{(T-21)/21}\left\|\hat{w}_{i,\tau+t\cdot 21+1} - \hat{w}_{i,(\tau+t\cdot 21)^+}\right\|_1, \tag{17}$$

$$SI_i = \frac{1}{T-\tau}\sum_{t=\tau}^{T}\frac{\|\hat{w}_{i,t}\|_1 - 1}{2}, \tag{18}$$

where T is the total number of observations in our sample period from June 1973 - June 2019, $\hat{w}_{i,t}$ are the weights of strategy i at time t, $\hat{w}_{i,t^+}$ are the respective weights at the end of t but before rebalancing, $\hat{w}_{i,t+1}$ are the congruent weights after rebalancing at time $t+1$, and $\|\hat{w}_t\|_1$ is the 1-norm of the portfolio weights at time t. We further denote the vector of excess returns at $t+1$ by r_{t+1}.[16]

Based on robust inference proposed by Ledoit and Wolf (2011), we measure the statistical significance of the differences between our averaging rule on the portfolio weight level, AV-Wgt-E, and LWNLS when investigating the out-of-sample standard deviation. We choose LWNLS as a benchmark because it represents the most sophisticated strategy in terms of portfolio variance minimization. When evaluating the out-of-sample Sharpe ratio, we compare our rule with ID using the robust procedure from Ledoit and Wolf (2008).[17] The ID portfolio serves as a benchmark for the Sharpe ratio because DeMiguel et al. (2009) find that no sophisticated portfolio strategy outperforms the 1/N portfolio.[18]

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

[17]We confine the statistical inference to the aforementioned pairwise tests to avoid multiple-testing problems in the inference (see Ledoit and Wolf (2017) and references therein).

[18]Note that the minimum-variance portfolio based on a scalar multiple of the identity matrix corresponds to the 1/N portfolio.

4.3 Out-of-sample results

Table 8 reports the annualized out-of-sample standard deviations of our proposed rule on the three averaging levels, the strategies, and benchmarks. For all considered data sets our rule achieves comparable standard deviations across all averaging levels. Averaging with exponentially smoothed jackknife returns (AV-Wgt-E) performs slightly superior to the unweighted jackknife returns, suggesting that the accommodation of time-series characteristics is beneficial on empirical data sets.[19] The standard deviation of AV-Wgt-E is lower than those of the considered strategies on all averaging levels and across all data sets.

[TABLE 8 ABOUT HERE]

Comparing AV-Wgt-E to the benchmark strategies in Panel C, we find that our rule achieves significantly lower standard deviations than LWNLS across all five data sets. Also, in comparison to the linear shrinkage strategies LW1F, LWID, LWCC, and KDM we find that AV-Wgt-E delivers lower standard deviations across all data sets. Our results show that averaging over multiple strategies leads to a more efficient diversification and compares favorably to existing shrinkage approaches in terms of out-of-sample standard deviation.

To study the relevance of each individual strategy within our AV-Wgt-E rule, we plot the estimated averaging intensities for all data sets in Figure 2. AV-Wgt-E combines at least two estimators at any point in time across all data sets. The maximum number of strategies with positive weights amounts to six strategies across all data sets, confirming the relevance of each strategy within our averaging rule.

[FIGURE 2 ABOUT HERE]

Table 9 shows the annualized out-of-sample Sharpe ratios of our averaging rule, the various strategies, and benchmarks. Our rule delivers the second highest Sharpe ratio on the Fama-French and industry portfolios compared to the strategies in Panel B. The Sharpe ratio of AV-Wgt-E on the STOCK500 data set is higher than for any strategy. AV-Wgt-E delivers a statistically significant Sharpe ratio increase relative to the ID strategy on four out of five data sets. The smallest increase in Sharpe ratio amounts to 25% for the 30Ind portfolio, being economically, but not statistically significant. We conclude that our averaging rule meets the

[19]For the remaining performance metrics we also find similar results on the different averaging levels, and thus, confine the discussion to AV-Wgt-E vis-a-vis the strategies and benchmarks for the sake of brevity.

challenge set forth by DeMiguel et al. (2009) to deliver significantly higher risk-adjusted excess returns.

[TABLE 9 ABOUT HERE]

The performance of AV-Wgt-E is comparable to the Ledoit-Wolf shrinkage strategies LW1F, LWID, LWCC, and LWNLS across all data sets.[20] AV-Wgt-E performs somewhat better on the 30Ind data set, while the shrinkage strategies seem to deliver slightly higher Sharpe ratios on the 25FF and STOCK500 data sets. The performance of AV-Wgt-E and the shrinkage strategies on the 6FF and 10Ind data set is almost identical.

We assess the practicality of our averaging rule against the established strategies in terms of turnover and short interest. Table 10 shows that the average monthly turnover of AV-Wgt-E is modest across all data sets and ranges from 4.49% on the 10Ind to 14.50% on the 25FF data set. The turnover of AV-Wgt-E is slightly lower than for the Ledoit-Wolf shrinkage strategies on the Fama-French and industry portfolios, but only half the turnover of the aforementioned strategies on the STOCK500 data set.[21] The turnover of AV-Wgt-E is generally lower for all data sets in comparison to the strategies allowing for short sales, i.e. Sample, 1F, CC and POET.

[TABLE 10 ABOUT HERE]

Table 11 reports the average daily short interest. The numbers for AV-Wgt-E range between 40.42% on the 10Ind and 173.52% on the 25FF data set, which is lower than for AV-Wgt. On average, the reported short interest of AV-Wgt-E is lower than that of the Ledoit-Wolf shrinkage strategies. The favorable comparison is mainly due to the STOCK500 data set, where AV-Wgt-E realizes a lower short interest than LWNLS and LWID of 10 and 60%, respectively. The short interest of our rule is, on average, slightly higher than for the strategies that allow short sales, except for the Sample and POET strategy.

[TABLE 11 ABOUT HERE]

[20]The benchmarks in Panel C also generate higher Sharpe ratios than the ID portfolio, which is in line with the results of Ledoit and Wolf (2017).

[21]The replacement of a stock in the investment universe is not considered for the turnover. This is in line with the turnover computation on the Fama-French data sets. The Fama-French factor and industry portfolios are constructed at the end of each June and its composition is kept stable for the next 12 months until revision of the applicable investment universe. Next June the composition changes, stocks no longer matching the criteria are dropped, and new stocks are included in the portfolios. This cannot be accounted for during the turnover computation and, consequently, we apply the same logic to the CRSP stock data set.

To further assess the practicality of our averaging rule compared to the existing portfolio strategies, we account for transaction costs in the empirical analysis. The portfolio return, net of transaction costs, is given by:

$$r_i^{tc} = (1 + r_i) \left(1 - \kappa \sum_{t=1}^{(T-21)/21} \left\| \hat{w}_{i,\tau+t\cdot 21+1} - \hat{w}_{i,(\tau+t\cdot 21)^+} \right\|_1 \right) - 1 \qquad (19)$$

where κ are the transaction costs for rebalancing the portfolio. In Table 12, we follow DeMiguel et al. (2013) and compare the out-of-sample Sharpe ratios after adjusting for proportional transaction costs of 25bps. Comparing our rule to the benchmarks, AV-Wgt-E performs somewhat better on 6FF, 30Ind, and STOCK500, while the shrinkage strategies seem to deliver slightly higher Sharpe ratios on the 25FF and 10Ind data sets. The transaction cost-adjusted Sharpe ratios of AV-Wgt-E are higher than for the ID strategy across all data sets, remaining significantly higher on four out of five data sets.[22]

[TABLE 12 ABOUT HERE]

Concluding our empirical analysis, we find that our averaging rule delivers significantly lower out-of-sample standard deviations than established estimators of the global minimum-variance portfolio. This result is important because the out-of-sample standard deviation is the primary performance criterion when estimating a minimum-variance portfolio (see Ledoit and Wolf, 2017). The risk-adjusted performance of our averaging rule is significantly higher in comparison to the 1/N portfolio and is akin to the Ledoit-Wolf shrinkage strategies, including the non-linear estimator of Ledoit and Wolf (2017). This still holds after adjusting for transaction costs. The modest turnover, especially on the STOCK500 data set, compares favorably to existing strategies and supports the practicality of our rule.

5 Conclusion

In this paper, we propose a novel approach to alleviate the impact of estimation errors on out-of-sample portfolio performance. Our approach builds on the idea of shrinkage and consists of combining the Sample with a multitude of structured estimators to minimize the expected

[22] In robustness tests in Table A.3.1 we double the transaction costs to 50bps to follow Balduzzi and Lynch (1999) who argue that 50bps are a good estimate of transaction costs for private investors. In this case the Sharpe ratio of AV-Wgt-E is after the adjustment for transaction costs no longer statistically, but still economically significantly higher than ID on the STOCK500 data set.

out-of-sample variance of the corresponding minimum-variance portfolio. 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 an improved out-of-sample performance.

We evaluate our rule through extensive simulations and find that averaging over a larger number of estimators is superior to averaging over fewer in terms of out-of-sample standard deviation and Sharpe ratio. This observation holds for all possible averaging levels, i.e. the portfolio weight, the inverse, and the covariance level. We attribute the performance gains to the leveraged benefits of estimation error diversification between the considered estimators. This diversification potential exists for large and small estimation windows. Our rule achieves, on average, lower out-of-sample standard deviations than any of the competing eleven minimum-variance strategies. The results for the risk-adjusted performance are similar for small estimation windows, suggesting that our averaging rule alleviates the impact of estimation errors on out-of-sample portfolio performance.

Our empirical results are in line with the findings from the simulation study. Our proposed averaging rule achieves a lower out-of-sample portfolio standard deviation than any competing minimum-variance strategy, including the non-linear shrinkage approach of Ledoit and Wolf (2017). In terms of out-of-sample Sharpe ratio, we find that our averaging rule outperforms the 1/N portfolio on all data sets, with statistically significant increases on four out of five. In terms of turnover and short interest, the portfolio weights of our rule are smooth over time and do not contain extreme positions.

Our results suggest that the diversification of estimation errors through averaging over the sample and multiple structured estimators provides a fruitful way to reduce the impact of estimation errors on out-of-sample portfolio performance. Future research may elaborate on alternative approaches for determining averaging intensities, or may extend the set of strategies used for averaging to explore potential limits of the suggested estimation error diversification.

References

Balduzzi, P. and Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, 52(1):47–78.

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

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance*, 54(2):655–671.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (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., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, 12(5):937–974.

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

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

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

Fan, J., Liao, Y., and Mincheva, M. (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 Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305.

Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics*, 146(2):342–350.

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

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

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

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

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

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

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

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

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

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

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

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

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

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Topics in Catalysis*, 8(4):323–361.

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

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

Timmermann, A. (2006). Chapter 4 Forecast combinations. In *Handbook of Economic Forecasting*, pages 135–196. North Holland, Amsterdam.

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

Xing, X., Hu, J., and Yang, Y. (2014). Robust minimum variance portfolio with L-infinity constraints. *Journal of Banking and Finance*, 46(1):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 on the considered levels 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 on the portfolio weight level with exponentially smoothed jackknife returns	AV-Wgt-E
2	Averaging on the inverse covariance level with exponentially smoothed jackknife returns	AV-Inv-E
3	Averaging on the covariance level with exponentially smoothed jackknife returns	AV-Cov-E
4	Averaging on the portfolio weight level with *i.i.id* jackknife returns	AV-Wgt
5	Averaging on the inverse covariance level with *i.i.id* jackknife returns	AV-Inv
6	Averaging on the covariance level with *i.i.id* jackknife returns	AV-Cov
Panel B: Strategies		
7	Minimum-variance portfolio based on the sample covariance matrix	Sample
8	Minimum-variance portfolio with a single market factor (Sharpe, 1963)	1F
9	Minimum-variance portfolio with the constant correlation model implied covariance matrix (Elton and Gruber, 1973)	CC
10	Minimum-variance portfolio with a scalar multiple of the identity matrix, representing the 1/N strategy of DeMiguel et al. (2009)	ID
11	Minimum-variance portfolio with short-sale constraints (Jagannathan and Ma, 2003)	SC
12	Minimum-variance portfolio based on the approximate factor model of (Fan et al., 2013), using the first three principal components	POET
Panel C: Benchmarks		
13	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
14	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
15	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
16	Minimum-variance portfolio with the non-linear shrinkage estimator of the covariance matrix by (Ledoit and Wolf, 2017)	LWNLS
17	Minimum-variance portfolio based on a linear combination of the inverse covariance matrix of the sample, identity matrix, and single factor matrix following (Kourtis et al., 2012)	KDM

Table 2: Summary statistics of simulation parameters

This table reports the monthly means, standard deviations (Std. dev.), 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 1963 - August 2007. The time horizon for the calibration of the parameters corresponds to Tu and Zhou (2011).

	MKT	SMB	HML
Descriptive statistics			
Mean	0.478	0.236	0.450
Std. dev.	4.379	3.215	2.809
Cross-correlations			
MKT	1.000	0.287	-0.390
SMB		1.000	-0.260
HML			1.000

Table 3: Descriptive statistics of estimation errors across estimation windows

This table shows the means, standard deviations (Std. dev.), and Bravais-Pearson correlation coefficients of the estimation errors between the single strategies. The estimation error for the m-th portfolio strategy in the k-th simulation run is defined as the following loss: $L_{k,m} = (\hat{w}_{k,m} - w_k)' \, \Sigma_k \, (\hat{w}_{k,m} - w_k)$, where $\hat{w}_{i,k}$ are the estimated minimum-variance portfolio weights and w_k are the true minimum-variance portfolio weights, which are computed using the population covariance matrix of the respective run, Σ_k. The reported values show the correlation over 10,000 simulation runs between the losses of the respective estimators. Panels A - E report the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The reported means and standard deviations are scaled by 1,000, respectively. The abbreviations for the strategies are explained in Panel B of Table 1.

	Sample	1F	CC	ID	SC	POET
Panel A: Estimation window $\tau = 60$						
Mean	1.326	0.440	0.537	0.940	0.441	0.548
Std. dev.	0.524	0.148	0.193	0.118	0.150	0.169
Cross-correlations						
Sample	1.000	0.365	0.224	-0.171	0.259	0.434
1F		1.000	0.357	-0.174	0.139	0.745
CC			1.000	-0.345	0.349	0.413
ID				1.000	-0.086	-0.257
SC					1.000	0.347
POET						1.000
Panel B: Estimation window $\tau = 120$						
Mean	0.481	0.394	0.480	0.940	0.306	0.402
Std. dev.	0.159	0.128	0.166	0.118	0.092	0.115
Cross-correlations						
Sample	1.000	0.216	0.180	-0.209	0.347	0.418
1F		1.000	0.106	-0.024	-0.043	0.609
CC			1.000	-0.370	0.327	0.232
ID				1.000	-0.060	-0.189
SC					1.000	0.191
POET						1.000
Panel C: Estimation window $\tau = 240$						
Mean	0.209	0.376	0.453	0.940	0.228	0.291
Std. dev.	0.066	0.120	0.151	0.118	0.066	0.088
Cross-correlations						
Sample	1.000	0.103	0.140	-0.220	0.282	0.334
1F		1.000	-0.025	0.106	-0.112	0.492
CC			1.000	-0.388	0.344	0.097
ID				1.000	0.013	-0.048
SC					1.000	0.053
POET						1.000

Table continues on next page

Table 3 continues here

	Sample	1F	CC	ID	SC	POET
Panel D: Estimation window $\tau = 480$						
Mean	0.099	0.369	0.438	0.940	0.185	0.212
Std. dev.	0.030	0.113	0.145	0.118	0.056	0.072
Cross-correlations						
Sample	1.000	0.024	0.117	-0.210	0.162	0.225
1F		1.000	-0.058	0.194	-0.101	0.391
CC			1.000	-0.380	0.323	0.014
ID				1.000	0.109	0.090
SC					1.000	-0.023
POET						1.000
Panel E: Estimation window $\tau = 960$						
Mean	0.049	0.366	0.432	0.940	0.162	0.164
Std. dev.	0.015	0.109	0.140	0.118	0.053	0.059
Cross-correlations						
Sample	1.000	-0.026	0.114	-0.238	0.077	0.141
1F		1.000	-0.057	0.226	-0.113	0.349
CC			1.000	-0.392	0.308	-0.027
ID				1.000	0.158	0.180
SC					1.000	-0.007
POET						1.000

Table 4: Simulated average out-of-sample standard deviations for varying numbers of strategies

This table reports the annualized average out-of-sample standard deviations of our proposed averaging rules with *i.i.d* jackknife returns on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level over 10,000 simulation runs for varying numbers of strategies, M. Panels A-E show the respective weights for the estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for $M = 2$ up to $M = 6$ represent averages over all possible permutations from combining the Sample with the five remaining strategies. The abbreviations for the averaging rules are explained in Panel A of Table 1.

	$M = 2$	$M = 3$	$M = 4$	$M = 5$	$M = 6$
Panel A: Estimation window $\tau = 60$					
AV-Wgt	0.170	0.165	0.164	0.162	0.162
AV-Inv	0.170	0.165	0.164	0.162	0.162
AV-Cov	0.168	0.166	0.165	0.164	0.164
Panel B: Estimation window $\tau = 120$					
AV-Wgt	0.162	0.160	0.159	0.157	0.157
AV-Inv	0.163	0.160	0.159	0.157	0.157
AV-Cov	0.161	0.160	0.159	0.158	0.158
Panel C: Estimation window $\tau = 240$					
AV-Wgt	0.157	0.156	0.155	0.155	0.155
AV-Inv	0.157	0.156	0.155	0.155	0.155
AV-Cov	0.157	0.156	0.156	0.155	0.155
Panel D: Estimation window $\tau = 480$					
AV-Wgt	0.154	0.153	0.153	0.153	0.153
AV-Inv	0.154	0.153	0.153	0.153	0.153
AV-Cov	0.154	0.153	0.153	0.153	0.153
Panel E: Estimation window $\tau = 960$					
AV-Wgt	0.152	0.152	0.152	0.152	0.152
AV-Inv	0.152	0.152	0.152	0.152	0.152
AV-Cov	0.152	0.152	0.152	0.152	0.152

Table 5: Simulated average out-of-sample standard deviations

This table reports the annualized average out-of-sample standard deviations of our proposed averaging rules with *i.i.d* jackknife returns on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level over 10,000 simulation runs in comparison to established minimum-variance strategies. The standard deviations are reported for the estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Wgt, AV-Inv, and AV-Cov are based on averaging over all $M = 6$ strategies. The abbreviations for the averaging rules, strategies, and benchmarks are explained in Table 1.

	$\tau = 60$	$\tau = 120$	$\tau = 240$	$\tau = 480$	$\tau = 960$
Panel A: Averaging rules					
AV-Wgt	0.162	0.157	0.155	0.153	0.152
AV-Inv	0.162	0.157	0.155	0.153	0.152
AV-Cov	0.164	0.158	0.155	0.153	0.152
Panel B: Single strategies					
Sample	0.258	0.168	0.158	0.154	0.152
1F	0.220	0.165	0.164	0.164	0.164
CC	0.216	0.168	0.167	0.167	0.167
ID	0.184	0.184	0.184	0.184	0.184
SC	0.202	0.162	0.159	0.157	0.156
POET	0.221	0.165	0.161	0.158	0.157
Panel C: Benchmarks					
LW1F	0.165	0.160	0.156	0.153	0.152
LWID	0.172	0.163	0.157	0.154	0.152
LWCC	0.167	0.161	0.157	0.154	0.152
LWNLS	0.167	0.161	0.157	0.154	0.152
KDM	0.167	0.165	0.164	0.164	0.164

Table 6: Simulated average out-of-sample Sharpe ratios

This table reports the annualized average out-of-sample Sharpe ratios of our proposed averaging rules with *i.i.d* jackknife returns on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level over 10,000 simulation runs in comparison to established minimum-variance strategies. The Sharpe ratios are reported for the estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Wgt, AV-Inv, and AV-Cov are based on averaging over all $M = 6$ strategies. The abbreviations for the averaging rules, strategies, and benchmarks are explained in Table 1.

	$\tau = 60$	$\tau = 120$	$\tau = 240$	$\tau = 480$	$\tau = 960$
Panel A: Averaging rules					
AV-Wgt	0.486	0.494	0.498	0.497	0.497
AV-Inv	0.486	0.494	0.497	0.497	0.497
AV-Cov	0.482	0.491	0.496	0.497	0.496
Panel B: Single strategies					
Sample	0.343	0.441	0.469	0.481	0.488
1F	0.398	0.442	0.440	0.438	0.438
CC	0.405	0.471	0.472	0.474	0.474
ID	0.471	0.471	0.471	0.471	0.471
SC	0.431	0.494	0.500	0.503	0.505
POET	0.395	0.460	0.472	0.480	0.485
Panel C: Benchmarks					
LW1F	0.457	0.466	0.476	0.483	0.489
LWID	0.448	0.463	0.477	0.485	0.490
LWCC	0.467	0.475	0.482	0.487	0.491
LWNLS	0.452	0.460	0.468	0.472	0.476
KDM	0.446	0.442	0.440	0.438	0.438

Table 7: List of the empirical data sets for daily data

This table lists the data sets for the empirical evaluation of our proposed averaging rule in comparison to existing minimum-variance strategies, their abbreviations, the number of assets in each data set, and the data sources. All data sets span the period from June 1973 - June 2019, comprise daily data, and apply in the case of portfolio data the value weighting scheme to the respective constituents. Data from Kenneth French is taken from his website (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`) and represents different cuts of the U.S. stock market. The STOCK500 data set contains the 500 largest single stocks in terms of market capitalization in July of every year after filtering out stocks that have a price of less than $5, or exhibit missing returns in the preceding 756 and subsequent 252 trading days to the selection date. All stock prices are taken from the Center of Research in Security Prices (CRSP).

#	Data set	Abbreviation	N	Source
1	6 Fama and French portfolios of firms sorted by size and book-to-market	6FF	6	K. French
2	25 Fama and French portfolios of firms sorted by size and book-to-market	25FF	25	K. French
3	10 industry portfolios representing the U.S. stock market	10Ind	10	K. French
4	30 industry portfolios representing the U.S. stock market	30Ind	30	K. French
5	500 Stocks with the largest market capitalization	STOCK500	500	CRSP

Table 8: Empirical out-of-sample standard deviations

This table reports the annualized out-of-sample standard deviations of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from June 1976 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. We report statistical significance for the null hypothesis wherein the log out-of-sample variance of AV-Wgt-E is greater than or equal to that of the non-linear shrinkage approach LWNLS of Ledoit and Wolf (2017). We follow Ledoit and Wolf (2011) and use their proposed bootstrap procedure with a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5, and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rules					
AV-Wgt-E	0.1180	0.0956	0.1127	0.1050	0.0851
AV-Inv-E	0.1191	0.0980	0.1141	0.1072	0.0895
AV-Cov-E	0.1228	0.0976	0.1135	0.1054	0.0837
AV-Wgt	0.1202	0.0979	0.1148	0.1064	0.0853
AV-Inv	0.1202	0.0979	0.1149	0.1064	0.0854
AV-Cov	0.1201	0.0977	0.1148	0.1067	0.0827
Panel B: Single strategies					
Sample	0.1203	0.0975	0.1156	0.1071	0.1159
1F	0.1401	0.1358	0.1203	0.1194	0.1187
CC	0.1360	0.1411	0.1204	0.1272	0.1420
ID	0.1657	0.1658	0.1562	0.1629	0.1686
SC	0.1393	0.1335	0.1210	0.1168	0.1027
POET	0.1223	0.1011	0.1173	0.1126	0.0899
Panel C: Benchmarks					
LW1F	0.1201	0.0975	0.1155	0.1069	0.0856
LWID	0.1211	0.0974	0.1156	0.1069	0.0959
LWCC	0.1217	0.1010	0.1158	0.1068	0.0900
LWNLS	0.1203***	0.0973***	0.1154***	0.1068***	0.0862**
KDM	0.1401	0.1358	0.1203	0.1194	0.1187

Table 9: Empirical out-of-sample Sharpe ratios

This table reports the annualized out-of-sample Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from June 1976 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. We use the 30-day T-bill rate as the risk-free rate. We report statistical significance for the null hypothesis wherein the Sharpe ratio of AV-Wgt-E is less than or equal to that of the ID strategy, corresponding to the 1/N portfolio. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure with a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5 and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rules					
AV-Wgt-E	1.1929	0.9361	0.7460	0.6327	0.8081
AV-Inv-E	1.1566	0.8600	0.7195	0.6258	0.7633
AV-Cov-E	1.0331	0.8744	0.7383	0.6460	0.8298
AV-Wgt	1.2230	0.9745	0.7474	0.6360	0.8087
AV-Inv	1.2215	0.9757	0.7474	0.6329	0.7915
AV-Cov	1.2059	0.9680	0.7429	0.6337	0.8813
Panel B: Single strategies					
Sample	1.2236	1.0181	0.7366	0.5973	0.6976
1F	0.9627	0.7175	0.7906	0.6319	0.6778
CC	0.9908	0.7305	0.7166	0.6200	0.5261
ID	0.5441***	0.3003***	0.5269*	0.5076	0.5750*
SC	0.8048	0.4770	0.6639	0.6413	0.8006
POET	1.2790	0.8664	0.7053	0.5921	0.6916
Panel C: Benchmarks					
LW1F	1.1999	1.0013	0.7413	0.6020	0.9334
LWID	1.1699	0.9751	0.7655	0.6255	0.8065
LWCC	1.1305	0.8729	0.7871	0.6029	0.8709
LWNLS	1.2110	0.9926	0.7376	0.6058	0.8690
KDM	0.9627	0.7175	0.7906	0.6319	0.6778

Table 10: Empirical average monthly turnover

This table reports the average monthly turnover of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from June 1973 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. Turnover is measured as the average percentage of total wealth traded in each month. The numbers are reported in percentages.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rules across three levels					
AV-Wgt-E	9.96	14.50	4.49	7.59	10.40
AV-Inv-E	9.55	12.35	4.13	6.73	8.33
AV-Cov-E	6.64	12.62	4.07	7.32	12.32
AV-Wgt	11.35	16.52	4.72	8.51	11.36
AV-Inv	11.49	16.62	4.76	8.59	11.27
AV-Cov	11.32	16.31	4.79	8.56	13.03
Panel B: Single strategies					
Sample	12.07	17.67	5.78	10.49	49.57
1F	9.39	16.22	7.61	11.00	12.27
CC	7.40	15.21	6.29	9.76	16.49
ID	1.48	1.68	2.32	2.87	5.45
SC	0.46	0.94	1.10	1.86	3.65
POET	15.96	18.25	6.62	11.28	17.75
Panel C: Benchmarks					
LW1F	11.70	17.24	5.74	10.11	22.64
LWID	9.05	15.58	5.00	9.63	33.38
LWCC	8.64	15.18	5.44	9.58	26.65
LWNLS	11.82	17.19	5.70	10.01	18.70
KDM	9.39	16.22	7.60	11.00	12.27

Table 11: Empirical average daily short interest

This table reports the average daily short interest of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from June 1973 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. Short interest is measured by the average amount of wealth that is held in short positions. The numbers are reported in percentages.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rules					
AV-Wgt-E	125.79	173.52	40.42	63.27	95.81
AV-Inv-E	120.47	147.57	35.36	53.02	96.07
AV-Cov-E	85.23	150.48	33.69	59.70	94.60
AV-Wgt	143.50	201.65	45.52	75.14	97.75
AV-Inv	145.02	202.89	45.87	76.00	96.67
AV-Cov	143.57	197.73	45.78	75.51	115.48
Panel B: Single strategies					
Sample	151.70	215.21	57.83	97.21	366.94
1F	102.92	156.50	68.81	88.75	62.30
CC	87.44	146.64	55.29	74.97	85.18
ID	0.00	0.00	0.00	0.00	0.00
SC	0.00	0.00	0.00	0.00	0.00
POET	191.03	210.65	67.52	97.94	113.66
Panel C: Benchmarks					
LW1F	147.10	209.67	57.28	92.65	156.13
LWID	114.82	189.64	48.04	87.77	240.22
LWCC	107.85	180.44	52.12	88.00	192.04
LWNLS	149.13	209.27	56.74	91.52	113.18
KDM	102.92	156.50	68.81	88.75	62.30

Table 12: Empirical out-of-sample Sharpe ratios adjusted for transaction costs of 25bps

This table reports the annualized out-of-sample Sharpe ratios after transaction costs of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from June 1973 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. We use the 30-day T-bill rate as the risk-free rate. The transaction costs are set to 25bps. We report statistical significance for the null hypothesis wherein the Sharpe ratio of AV-Wgt-E is less than or equal to that of the ID strategy, corresponding to the 1/N portfolio. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure with a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5 and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rules					
AV-Wgt-E	1.1676	0.8903	0.7341	0.6111	0.7713
AV-Inv-E	1.1325	0.8220	0.7086	0.6070	0.7354
AV-Cov-E	1.0169	0.8355	0.7276	0.6252	0.7856
AV-Wgt	1.1945	0.9236	0.7351	0.6120	0.7713
AV-Inv	1.1927	0.9245	0.7350	0.6087	0.7354
AV-Cov	1.1775	0.9176	0.7304	0.6096	0.7856
Panel B: Single strategies					
Sample	1.1934	0.9634	0.7216	0.5680	0.5426
1F	0.9425	0.6815	0.7716	0.6043	0.6483
CC	0.9745	0.6981	0.7009	0.5970	0.4908
ID	0.5415***	0.2973***	0.5225*	0.5024	0.5669*
SC	0.8038	0.4748	0.6612	0.6365	0.7882
POET	1.2396	0.8119	0.6885	0.5621	0.6104
Panel C: Benchmarks					
LW1F	1.1706	0.9478	0.7264	0.5737	0.7740
LWID	1.1475	0.9269	0.7526	0.5985	0.7810
LWCC	1.1092	0.8277	0.7730	0.5761	0.7797
LWNLS	1.1814	0.9393	0.7228	0.5777	0.7540
KDM	0.9425	0.6815	0.7716	0.6043	0.6483

Figure 1: Mean averaging intensities in the simulated data sets

This figure plots the mean averaging intensities over all 10,000 simulations that our averaging rule with *i.i.d* jackknife returns assigns to one of the considered estimators for estimation window sizes of $\tau = \{60, 120, 240, 480, 960\}$. Panel A shows the mean averaging intensities when averaging is conducted on the portfolio weight level (AV-Wgt), Panel B when averaging is conducted on the the inverse (AV-Inv), and Panel C when averaging is conducted on the covariance level (AV-Cov). The abbreviations for the six strategies are explained in Panel B of Table 1.

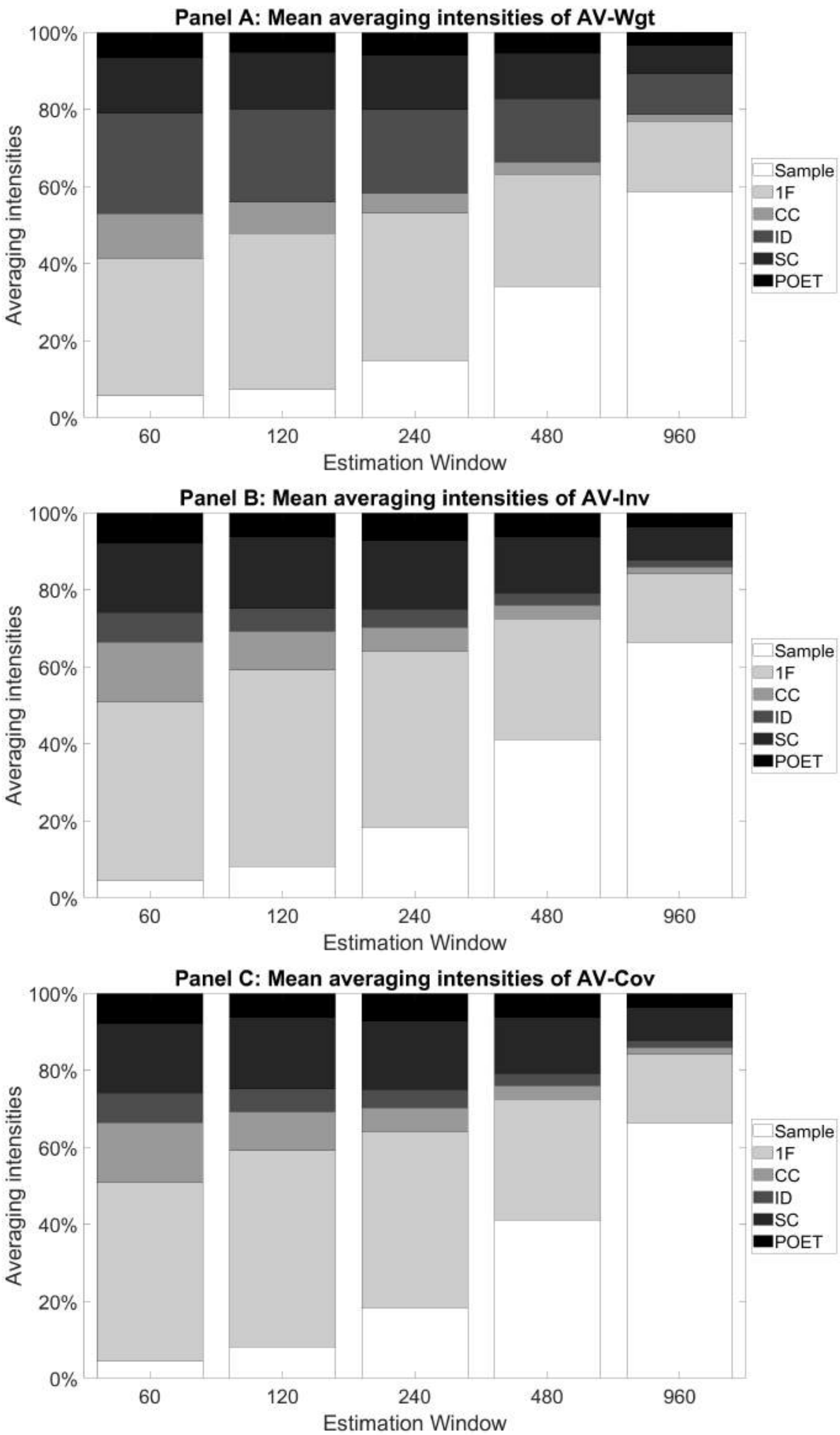
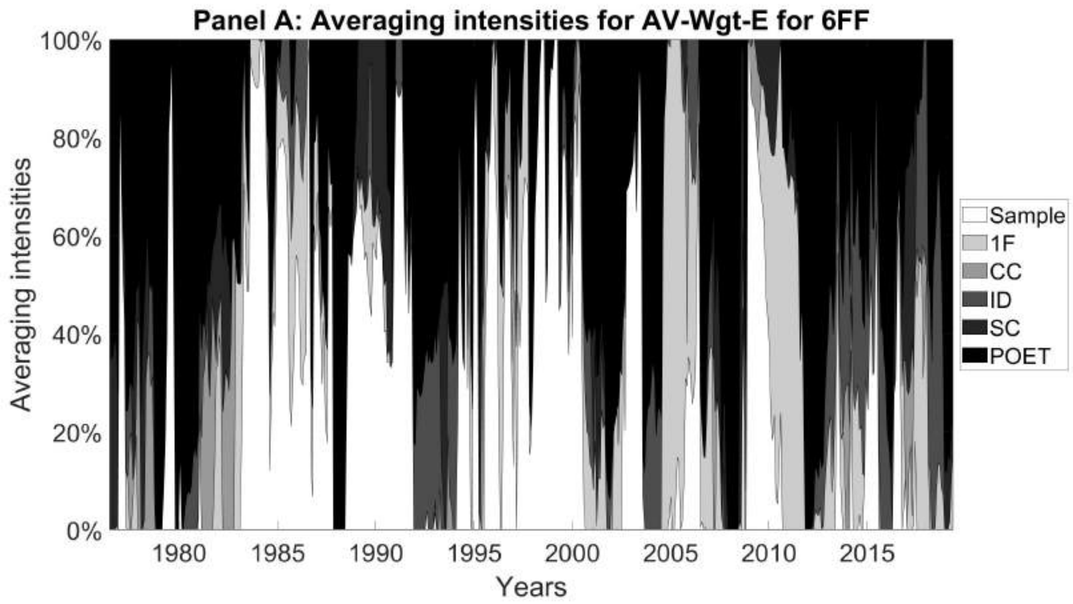
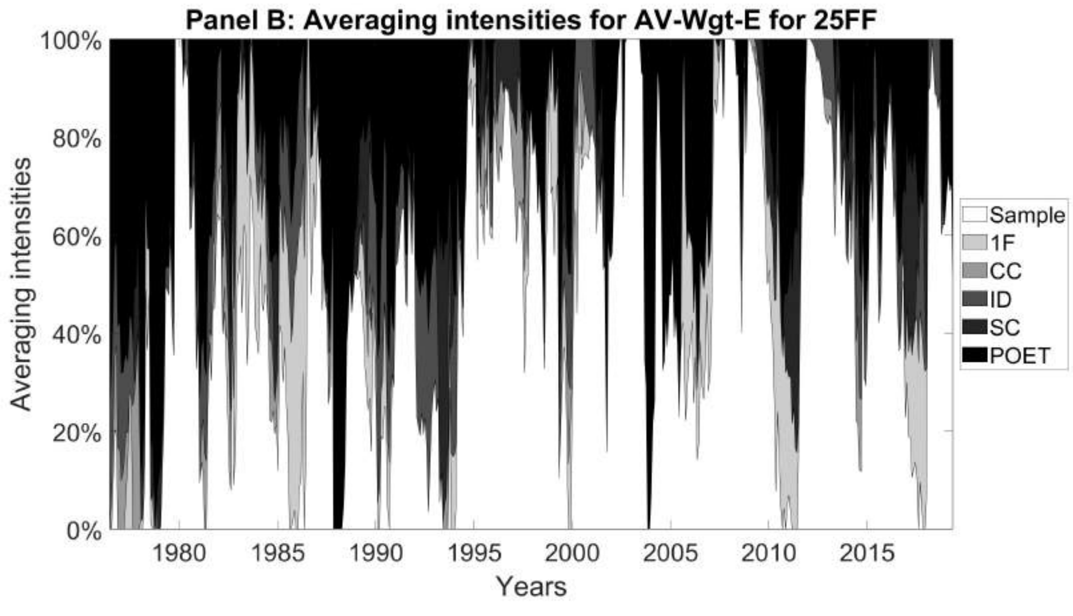
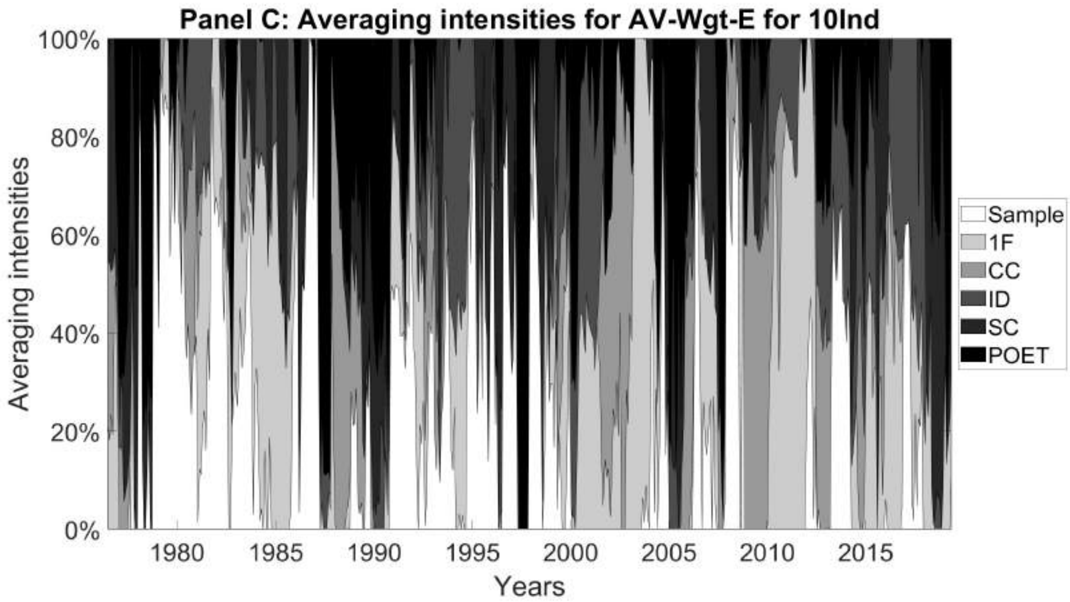


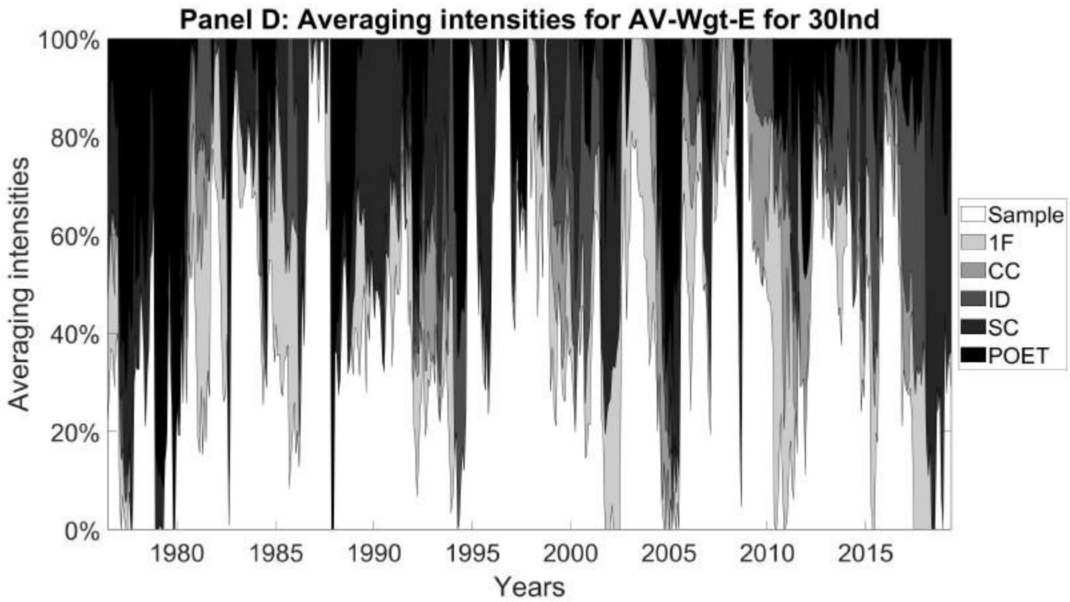
Figure 2: Averaging intensities of AV-Wgt-E for portfolios and stocks

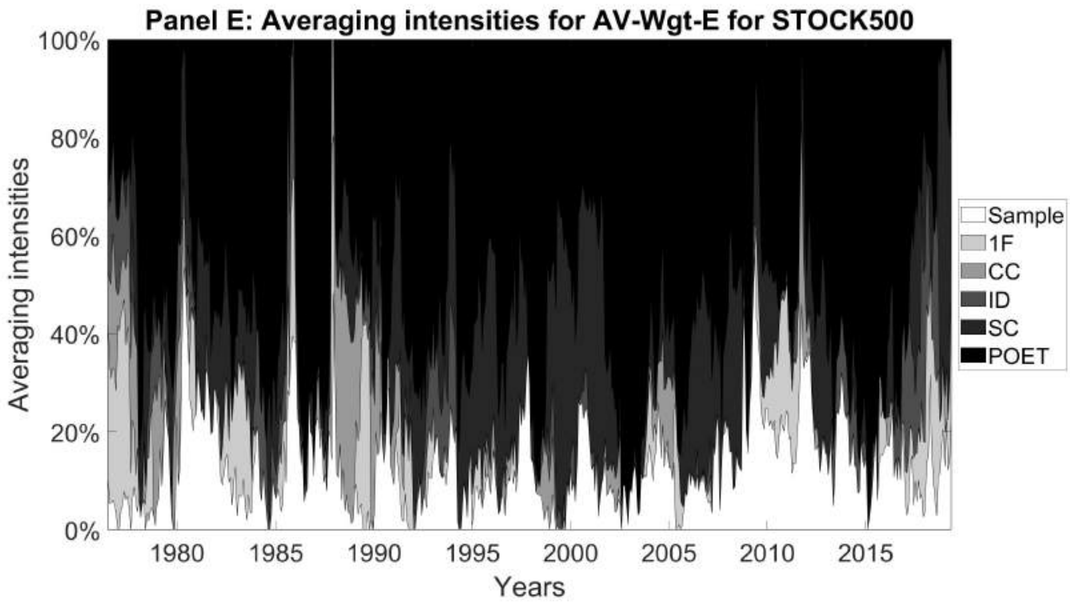
This figure plots the averaging intensities of the averaging rule on the portfolio weight level with exponentially smoothed jackknife returns (AV-Wgt-E) over the out-of-sample period from June 1973 - June 2019. The intensities represent the allocation of the respective strategy within the averaging rule over the holding period, which comprises 21 trading days. Panel A shows the plot for the 6FF, Panel B for the 25FF, Panel C for the 10Ind, Panel D for the 30Ind, and Panel E for the STOCK500 data set. The abbreviations for the strategies and data sets are explained in Tables 1 and 7.






Panel C: Averaging intensities for AV-Wgt-E for 10Ind
Averaging intensities
100%
80%
60%
40%
20%
0%
1980
1985
1990
1995
2000
2005
2010
2015
Years
Sample
1F
CC
ID
SC
POET


Panel D: Averaging intensities for AV-Wgt-E for 30Ind
Averaging intensities
100%
80%
60%
40%
20%
0%
1980
1985
1990
1995
2000
2005
2010
2015
Years
Sample
1F
CC
ID
SC
POET


Panel E: Averaging intensities for AV-Wgt-E for STOCK500
Averaging intensities
100%
80%
60%
40%
20%
0%
1980
1985
1990
1995
2000
2005
2010
2015
Years
Sample
1F
CC
ID
SC
POET

Internet Appendix for

Diversifying estimation errors:
An efficient averaging rule for portfolio optimization

This Version: May 5, 2021

Abstract

We propose an averaging rule that combines established minimum-variance strategies to minimize the expected out-of-sample variance. Our rule overcomes the problem of selecting the "best" strategy ex-ante and diversifies remaining estimation errors of the strategies included in the averaging. Extensive simulations show that the contributions of estimation errors to the out-of-sample variances are uncorrelated between the considered strategies. This implies that averaging over multiple strategies offers sizable diversification benefits. Across all data sets we find that our rule achieves a significantly lower out-of-sample standard deviation than any competing strategy and that the Sharpe ratio is at least 25% higher than for the 1/N portfolio.

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: Additional simulation results

Appendix A.3: Additional empirical results

Appendix A.4: Empirical study on monthly excess returns

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, i.e. the sample size τ:

$$m = \frac{1}{\tau} X \mathbf{1}_N \tag{20}$$

$$\hat{\Sigma}_S = \frac{1}{\tau} X \left(I - \frac{1}{\tau} \mathbf{1}_N \mathbf{1}_N' \right) X^{-1} \tag{21}$$

where X is a $N \times \tau$ matrix of τ observations and N assets, $\mathbf{1}_N$ 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) described as:

$$r_{i,t} = \alpha_i + \beta_i r_{MKT,t} + \epsilon_{i,t}, \tag{22}$$

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:

$$\hat{\Sigma}_{1F} = \hat{\sigma}^2_{MKT} \hat{\beta} \hat{\beta}' + \hat{\Delta}, \tag{23}$$

where $\hat{\sigma}^2_{MKT}$ is an unbiased estimate of the market variance, $\hat{\beta}$ is an $N \times 1$ vector comprising 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:

$$\hat{\Sigma}_{CC} = \Lambda C \Lambda', \tag{24}$$

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:

$$\hat{\Sigma}_{ID} = I\mu_{\hat{s}}, \text{ with } \mu_{\hat{s}} = \frac{1}{N}\sum_{i=1}^{N} \hat{s}_i^2, \tag{25}$$

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

Principal components

Fan et al. (2013) introduce the principal orthogonal complement thresholding method, POET, which uses an approximate factor model to define the risk factors based on principal components of the sample covariance matrix. The authors impose a threshold on the remaining principal components after taking out the first K. The covariance matrix estimator $\hat{\Sigma}_{POET}$ of the POET model is given by:

$$\hat{\Sigma}_{POET} = \sum_{i=1}^{K} \hat{\gamma}_{\tau,i}\hat{\xi}_i\hat{\xi}_i' + \hat{R}_K^T, \tag{26}$$

where $\hat{\gamma}_{\tau,1} \geq \hat{\gamma}_{\tau,2} \geq \cdots \geq \hat{\gamma}_{\tau,N}$ are the ordered eigenvalues of the sample covariance matrix $\hat{\Sigma_S}$ over the estimation window τ, $(\hat{\xi}_i)_{i=1}^N$ are the corresponding eigenvectors, $\hat{R}_K^T$ is the principal orthogonal complement of K diverging eigenvalues of $\hat{\Sigma}_S$. We set $K = 3$ in applications of the model.

Linear shrinkage estimators of the covariance matrix

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

$$\hat{\Sigma}_{LW} = \phi\hat{F} + (1-\phi)\hat{\Sigma}_S. \tag{27}$$

The shrunken 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 $\hat{\Sigma}_{LW}$. The authors consider three different candidates for $\hat{F}$: the single-factor model covariance matrix $\hat{\Sigma}_{LW1F}$ by Ledoit and Wolf (2003), a multiple of the identity matrix $\hat{\Sigma}_{LWID}$ as in Ledoit and Wolf (2004a), and the constant correlation model implied covariance matrix $\hat{\Sigma}_{LWCC}$ by Ledoit and Wolf (2004b).

Shrinkage estimators for 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:

$$\hat{\Sigma}_{KDM}^{-1} = \phi_1 \hat{\Sigma}_S^{-1} + \phi_2 I + (1 - \phi_1 - \phi_2)\hat{\Sigma}_{1F}^{-1}, \tag{28}$$

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

Non-linear shrinkage estimators for the inverse covariance matrix

Ledoit and Wolf (2017) propose a non-linear shrinkage estimator (NLS) of the covariance matrix. The procedure corrects the over-dispersed eigenvalues of the sample covariance matrix. The non-linear shrinkage estimator for the NLS covariance matrix is given by:

$$\begin{aligned} \hat{\Sigma}_{NLS} &:= U_\tau \hat{D}_\tau U_\tau', \text{ where} \\ \hat{D}_\tau &:= diag(\hat{d}_\tau(\gamma_{\tau,1}), \ldots, \hat{d}_\tau(\gamma_{\tau,N})), \text{ and} \\ \hat{d}_\tau(\gamma_{\tau,i}) &:= \begin{cases} \frac{1}{\gamma_{\tau,i}|\hat{c}(\gamma_{\tau,i})|^2} \text{ if } \gamma_{\tau,i} > 0, \\ \frac{1}{(\frac{N}{\tau}-1)\hat{c}(0)} \text{ if } N > \tau \text{ and } \gamma_{\tau,i} = 0, \end{cases} \end{aligned} \tag{29}$$

where U_τ is an orthogonal matrix from a spectral decomposition of the sample covariance matrix $\hat{\Sigma}_S$, $\hat{c}$ is the complex-valued Stieltjes transformation, and $\gamma_{\tau,i}$ are the sample eigenvalues, sorted in increasing order.

Short-sale constraints

Jagannathan and Ma (2003) provide a moment shrinkage interpretation of short-sale constraints in the minimum-variance portfolio optimization problem. The shortsale constraint augments the optimization problem in Equation (1) by the following inequality:

$$\hat{w}_i \geq 0, \text{ for } i = 1, 2, \ldots, N. \tag{30}$$

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

$$\hat{\Sigma}_{SC} = \hat{\Sigma}_S - (\delta 1' + 1\delta'), \tag{31}$$

where $\delta = (\delta_1, \delta_2, \ldots, \delta_N)$ is the vector of Lagrange multipliers for the non-negativity constraints in Equation (30).

A.2 Additional simulation results

This section contains the supporting figures for our simulation study. Complementing the results for the standard deviation in Table 4, we report the results for the average out-of-sample Sharpe ratio in Table A.2.1. Similar to the standard deviation, we find that the average Sharpe ratio increases with the number of strategies.

In the left column of Figures A.2.1, A.2.2, and A.2.3, we evaluate the distribution of the annualized out-of-sample standard deviations for AV-Wgt, AV-Inv, and AV-Cov with the kernel densities over all simulation runs and permutations. The distributions for $M = 6$ center around smaller out-of-sample standard deviations and collapse more tightly around this value than the distributions for smaller values of M. The result holds for all averaging levels and supports the finding that combining all six strategies is beneficial when compared to averaging over a subset. Investigating the distribution of the out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov, we turn to the right column of Figures A.2.1, A.2.2, and A.2.3. We find that the distributions for $M = 6$ are centered around a higher mean value and collapse more tightly around the mean value for estimation windows up to $\tau = 120$. This effect diminishes as the estimation window increases.

In Figures A.2.4, A.2.5, and A.2.6, we investigate the out-of-sample standard deviations of our averaging rule in relation to the existing strategies in more detail, exploring the kernel densities for AV-Wgt, AV-Inv, and AV-Cov, respectively. The left column in each figure shows the densities in comparison to the considered strategies. We find that the distributions for AV-Wgt, AV-Inv, and AV-Cov are centered around a smaller mean value and collapse more tightly for estimation windows up to $\tau = 480$ in comparison to the aforementioned strategies. The kernel densities of each averaging level of our rule also compare favorably to the benchmarks. We find that the distributions for AV-Wgt, AV-Inv, and AV-Cov collapse as tightly as the competing strategies, but around a smaller mean value. This observation holds for estimation windows of up to $\tau = 480$.

The kernel density plots of the annualized out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov in Figures A.2.7, A.2.8, and A.2.9 corroborate the favorable comparison relative to the existing strategies. The left column of the aforementioned figures presents the results in comparison to the considered strategies. The distribution of the out-of-sample Sharpe ratios for all averaging levels of our rule collapses more tightly around a higher mean value than the aforementioned strategies, except for SC. The right column of the aforementioned figures shows the results in comparison to the benchmarks. We find that our rule centers around a higher mean value for all estimation windows and collapses more tightly around the respective mean for estimation windows up to $\tau = 120$.

Table A.2.1: Simulated average out-of-sample Sharpe ratios for a varying number of strategies

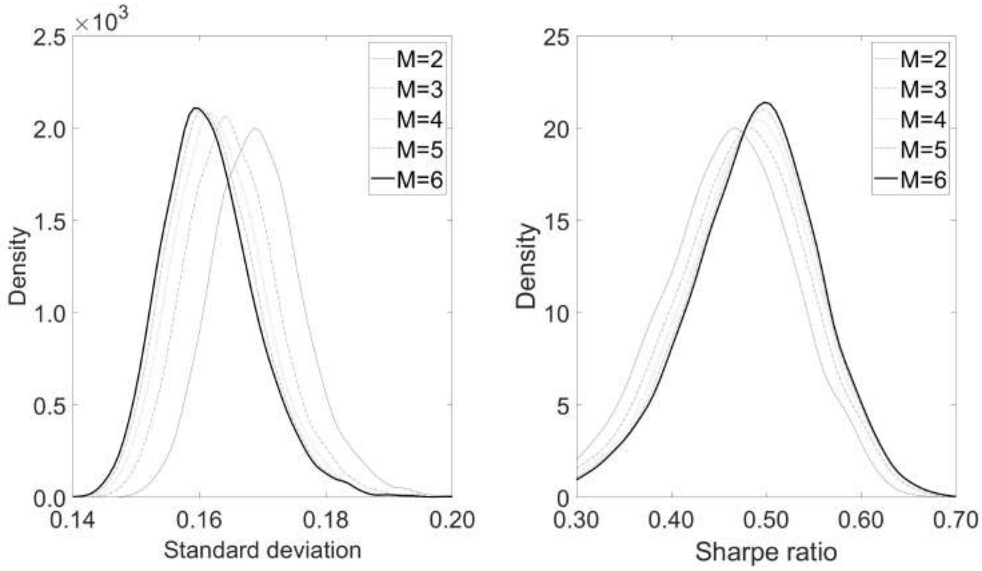
This table reports the annualized average out-of-sample Sharpe ratios of our proposed averaging rules with *i.i.d* jackknife returns on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance matrix (AV-Cov) level over 10,000 simulation runs for a varying number of strategies, M. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for $M = 2$ up to $M = 6$ represent averages over all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 1. The abbreviations for the averaging rules are explained in Panel A of Table 1.

	$M = 2$	$M = 3$	$M = 4$	$M = 5$	$M = 6$
Panel A: Estimation window $\tau = 60$					
AV-Wgt	0.458	0.474	0.481	0.486	0.486
AV-Inv	0.459	0.475	0.481	0.486	0.486
AV-Cov	0.462	0.473	0.478	0.481	0.482
Panel B: Estimation window $\tau = 120$					
AV-Wgt	0.470	0.484	0.489	0.494	0.494
AV-Inv	0.470	0.484	0.489	0.494	0.494
AV-Cov	0.474	0.484	0.488	0.491	0.491
Panel C: Estimation window $\tau = 240$					
AV-Wgt	0.480	0.489	0.493	0.497	0.498
AV-Inv	0.480	0.489	0.493	0.497	0.497
AV-Cov	0.482	0.490	0.493	0.495	0.496
Panel D: Estimation window $\tau = 480$					
AV-Wgt	0.486	0.491	0.494	0.497	0.497
AV-Inv	0.486	0.491	0.494	0.497	0.497
AV-Cov	0.488	0.492	0.495	0.496	0.497
Panel E: Estimation window $\tau = 960$					
AV-Wgt	0.490	0.493	0.495	0.497	0.497
AV-Inv	0.490	0.493	0.494	0.497	0.497
AV-Cov	0.491	0.493	0.495	0.496	0.496

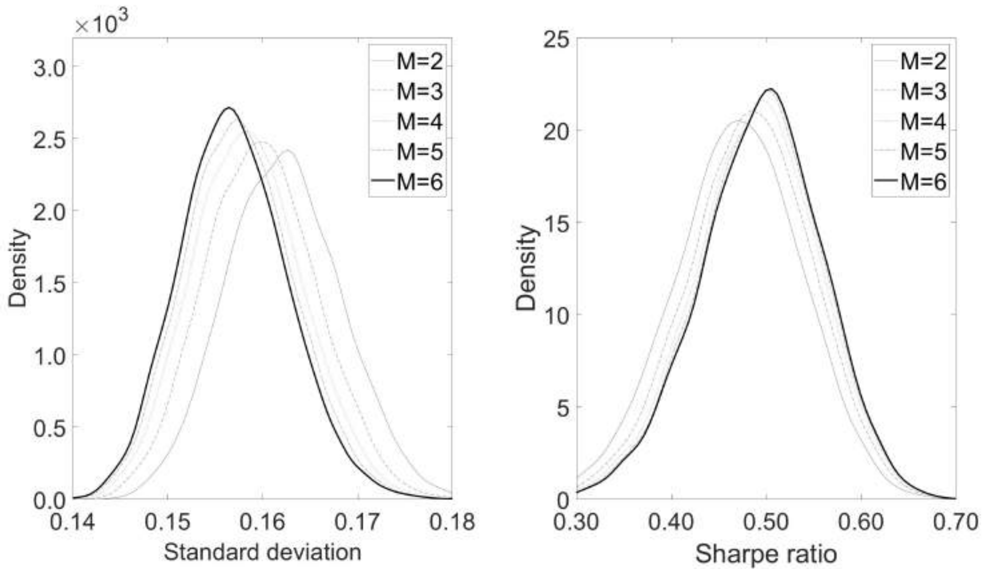
Figure A.2.1: Kernel densities of simulated standard deviations and Sharpe ratios for AV-Wgt

This figure plots the densities based on the normal kernel of the annualized out-of-sample standard deviations and Sharpe ratios of our averaging rule with *i.i.d* jackknife returns on the portfolio weight (AV-Wgt) level over 10,000 simulation runs for a varying number of strategies, M. The densities for $M = 2$ up to $M = 6$ are based on all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 1. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$.

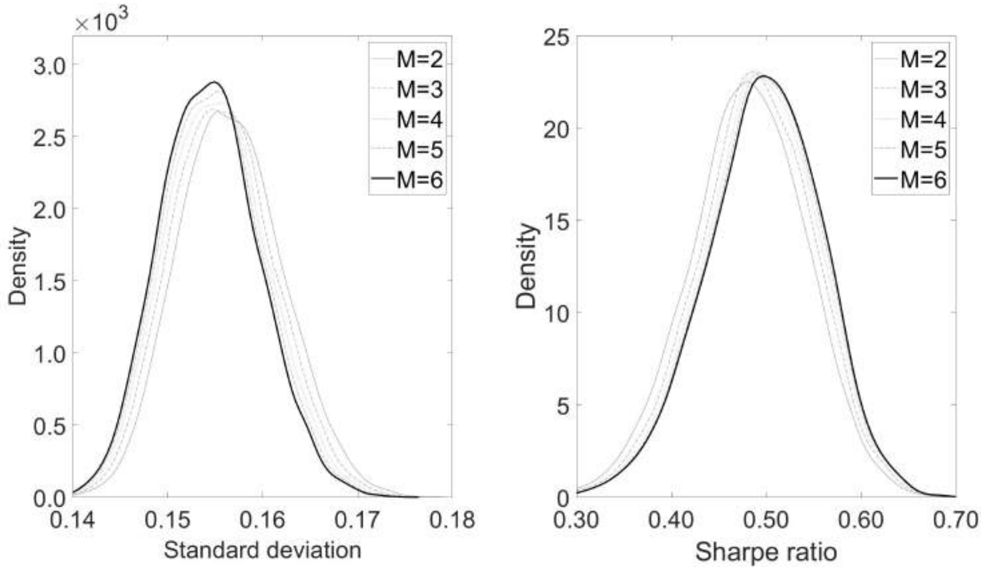
Panel A: Estimation window τ = 60



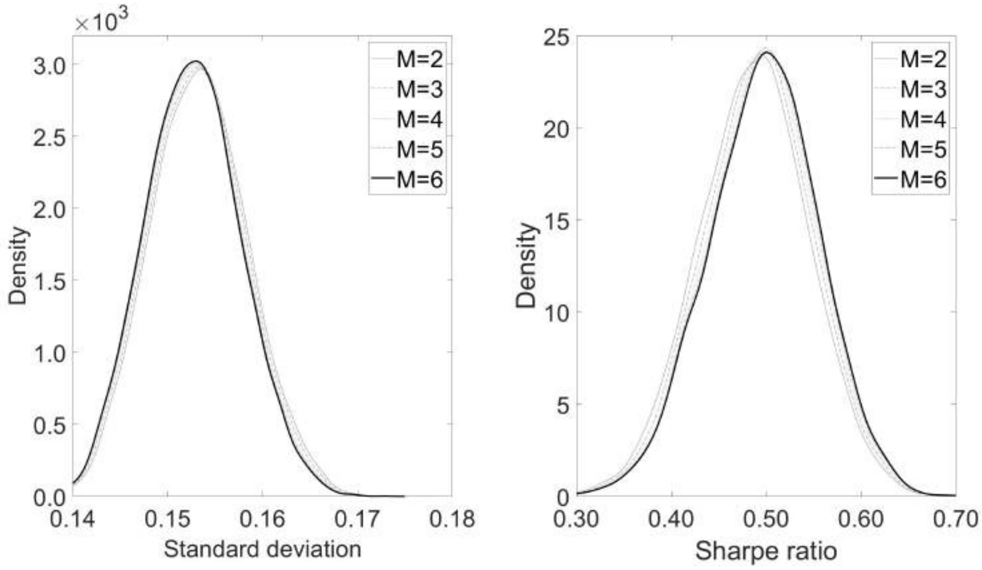
Panel B: Estimation window τ = 120



Panel C: Estimation window τ = 240



Panel D: Estimation window τ = 480



Panel E: Estimation window τ = 960

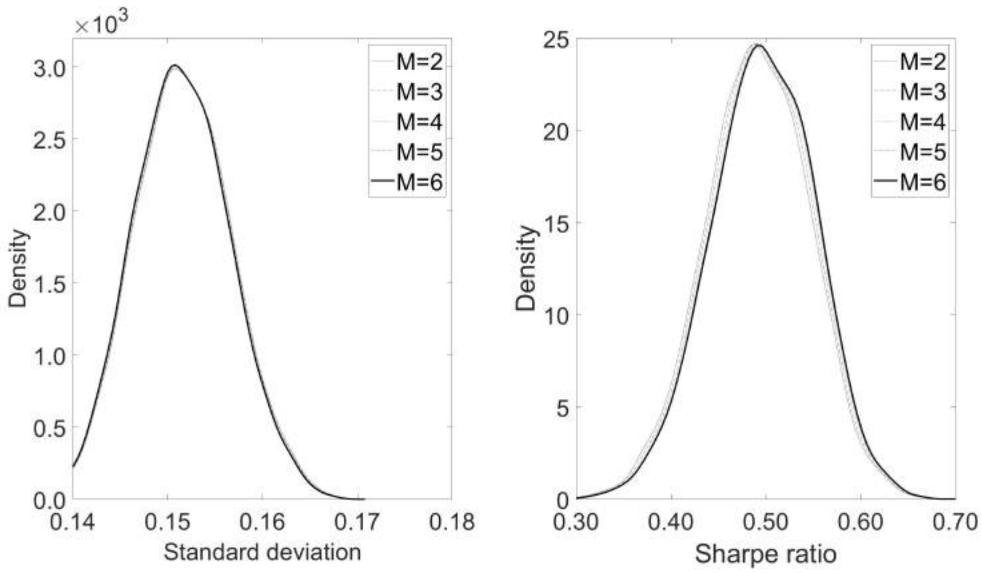
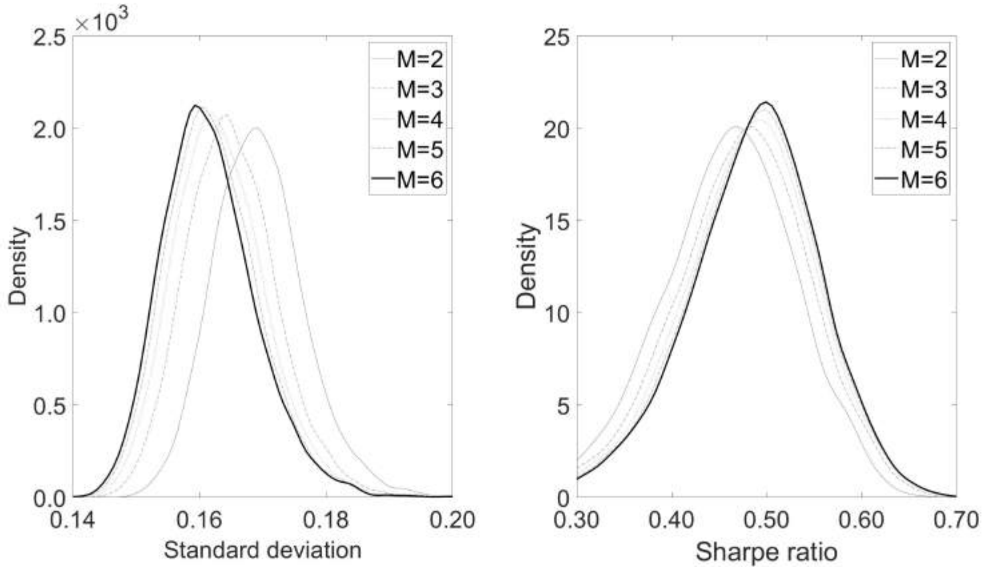


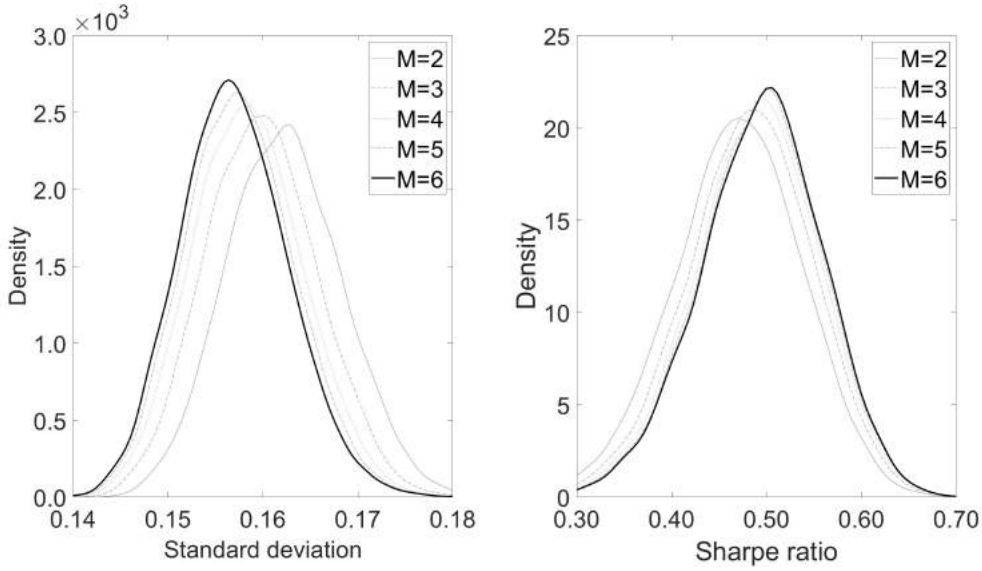
Figure A.2.2: Kernel densities of simulated standard deviations and Sharpe ratios for AV-Inv

This figure plots the densities based on the normal kernel of the annualized out-of-sample standard deviations and Sharpe ratios of our averaging rule with *i.i.d* jackknife returns on the inverse covariance matrix (AV-Inv) level over 10,000 simulation runs for a varying number of strategies, M. The densities for $M = 2$ up to $M = 6$ are based on all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 1. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$.

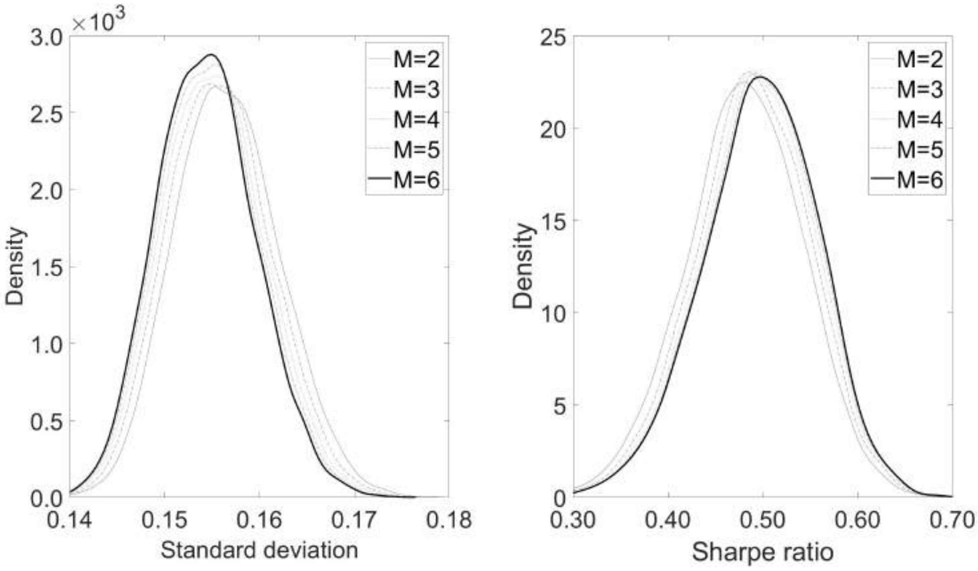
Panel A: Estimation window τ = 60



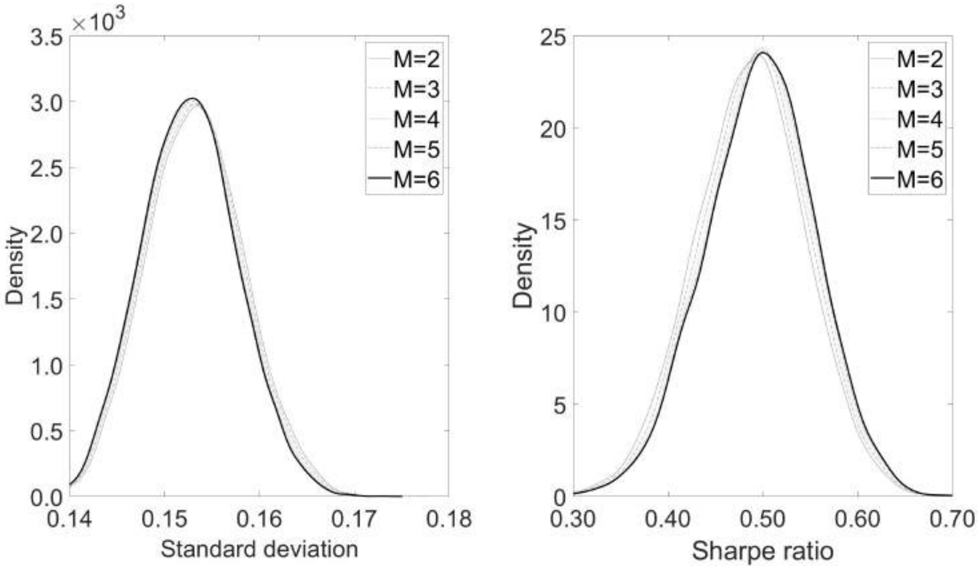
Panel B: Estimation window τ = 120



Panel C: Estimation window τ = 240



Panel D: Estimation window τ = 480



Panel E: Estimation window τ = 960

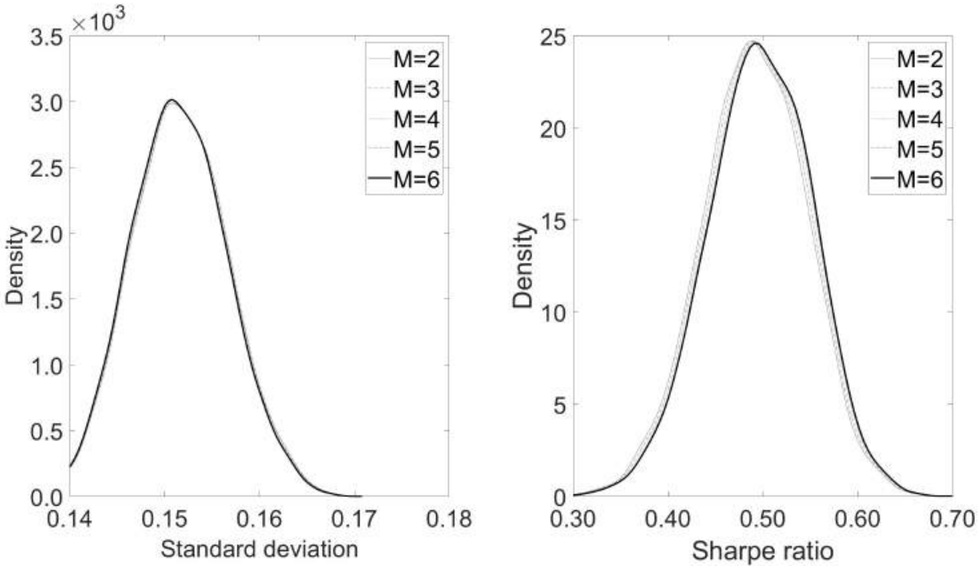
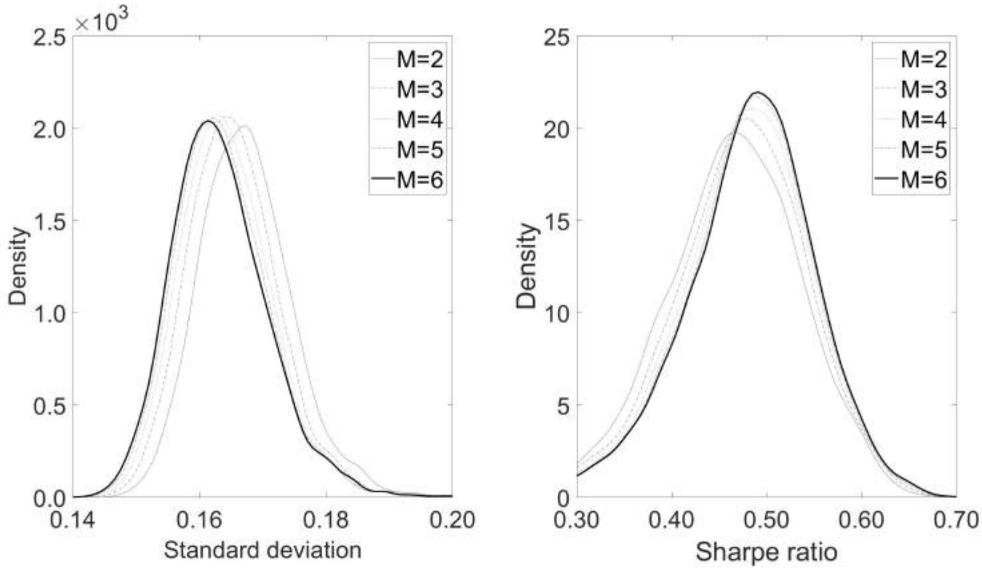


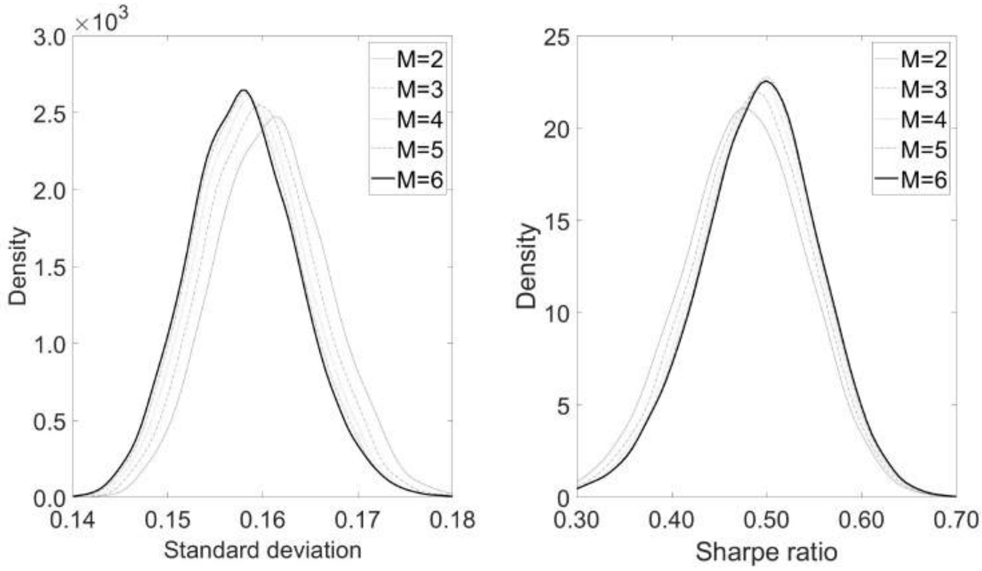
Figure A.2.3: Kernel densities of simulated standard deviations and Sharpe ratios for AV-Cov

This figure plots the densities based on the normal kernel of the out-of-sample standard deviations and Sharpe ratios of our averaging rule with *i.i.d* jackknife returns on the covariance matrix (AV-Cov) level over 10,000 simulation runs for a varying number of strategies, M. The densities for $M = 2$ up to $M = 6$ are based on all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 1. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$.

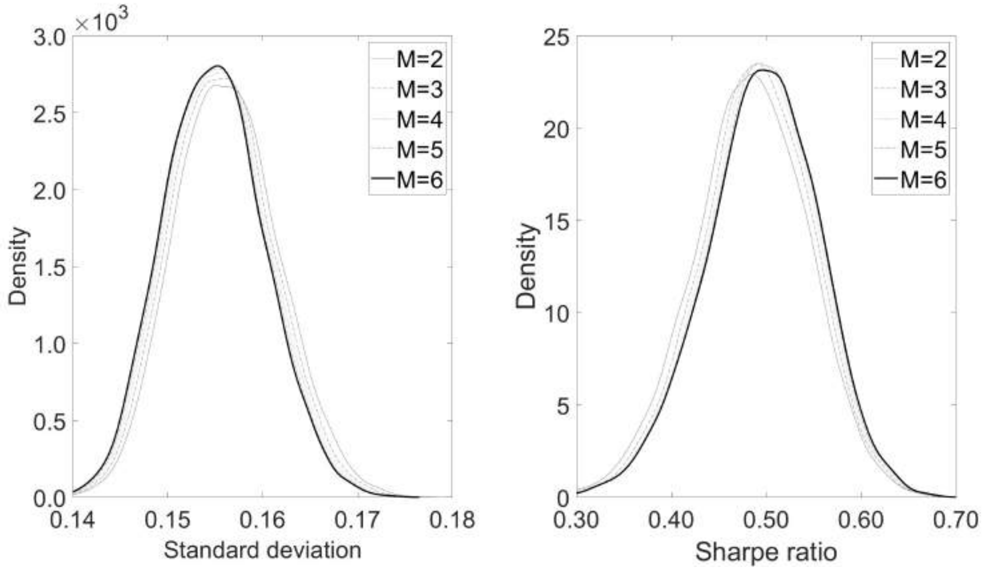
Panel A: Estimation window τ = 60



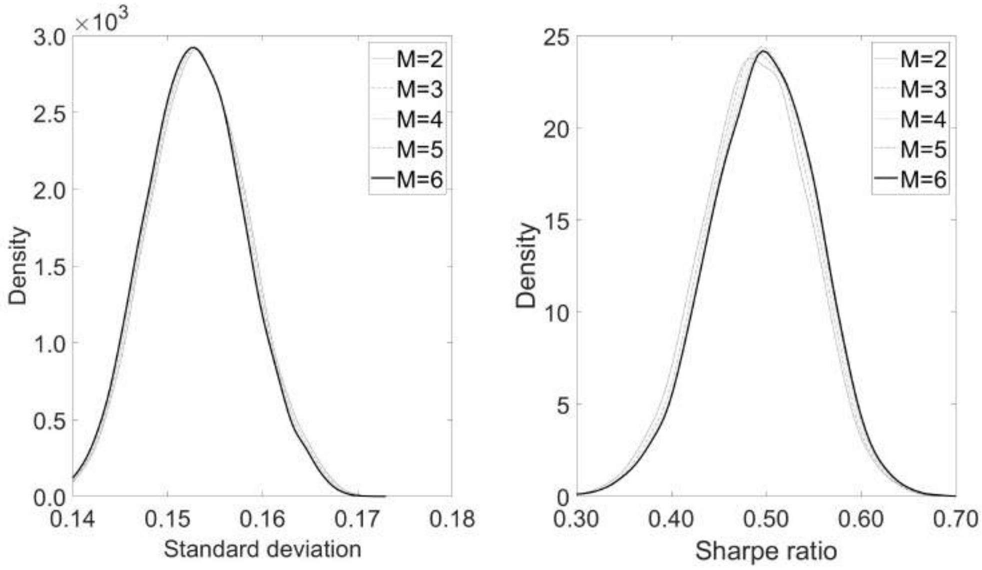
Panel B: Estimation window τ = 120



Panel C: Estimation window τ = 240



Panel D: Estimation window τ = 480



Panel E: Estimation window τ = 960

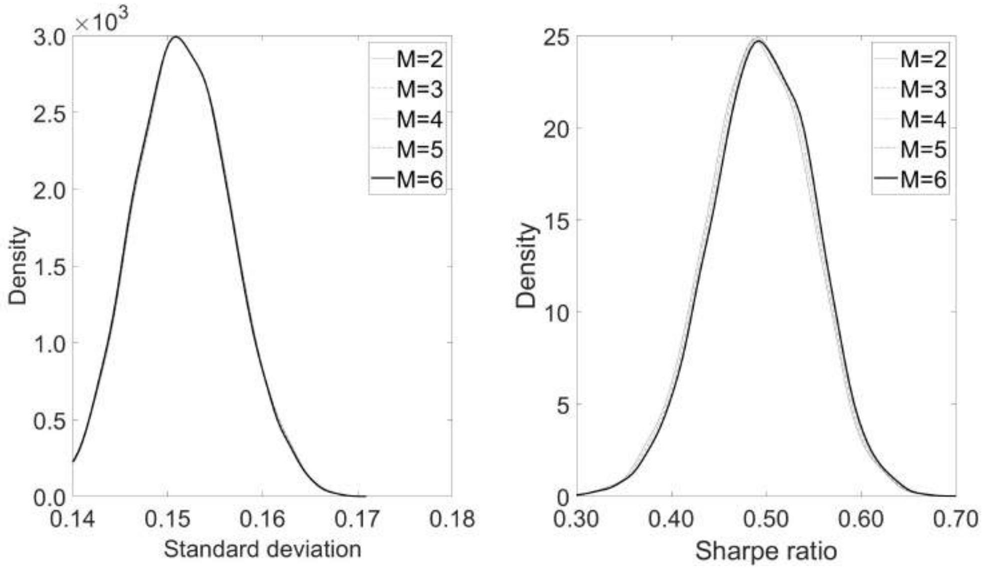
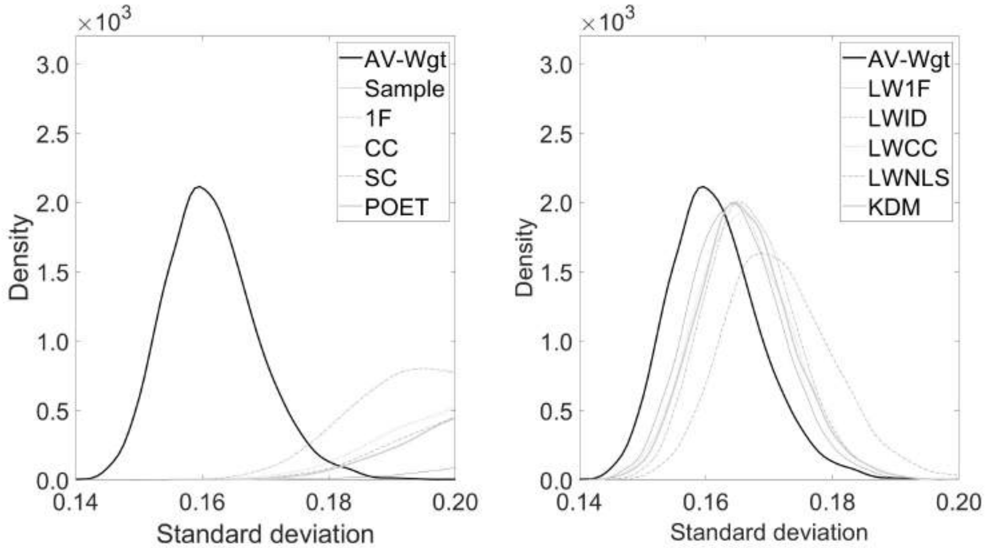


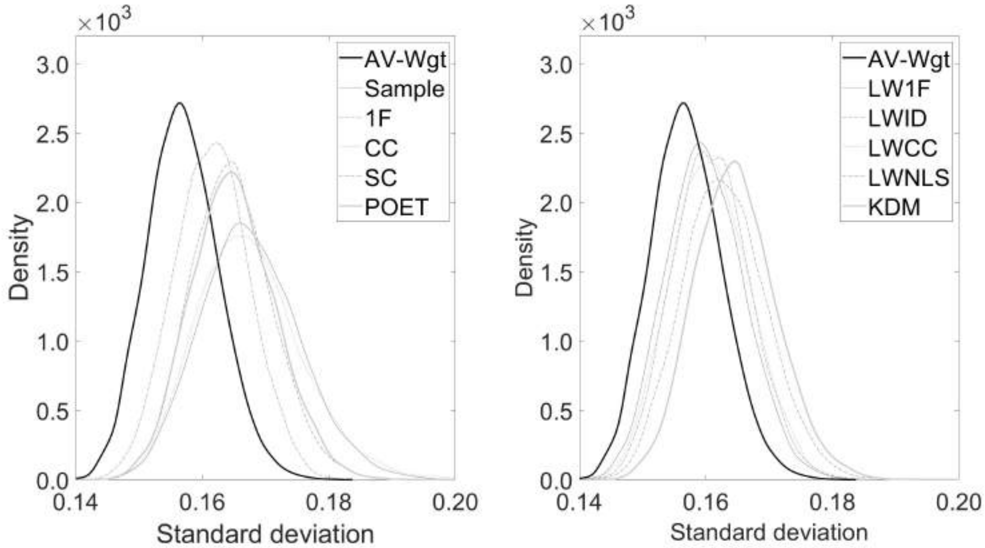
Figure A.2.4: Kernel densities of simulated standard deviations for AV-Wgt compared to established strategies and benchmarks

This figure plots the densities based on the normal kernel of the annualized out-of-sample standard deviations of our averaging rule with *i.i.d* jackknife returns on the portfolio weight (AV-Wgt) level over 10,000 simulation runs in comparison to the strategies (left column) and benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 1. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

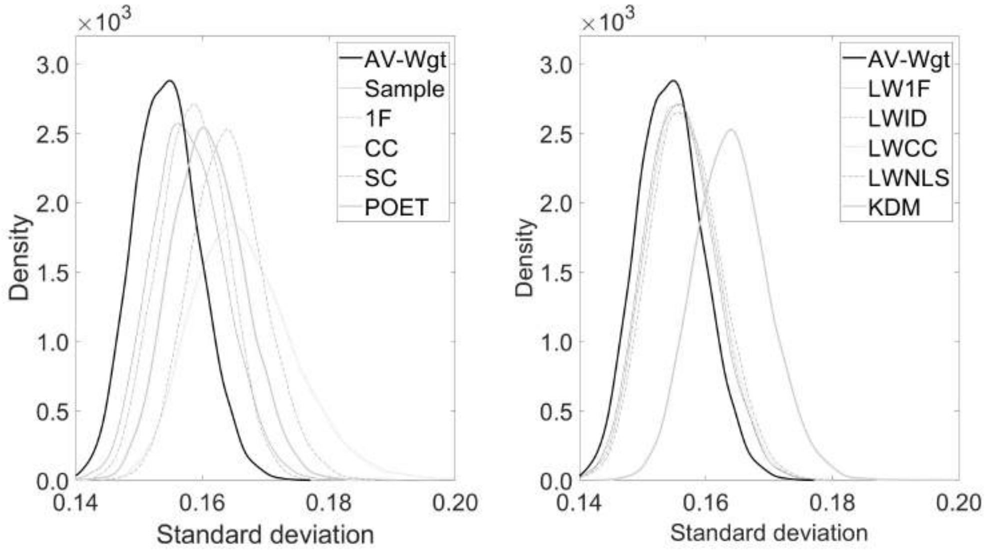
Panel A: Estimation window τ = 60



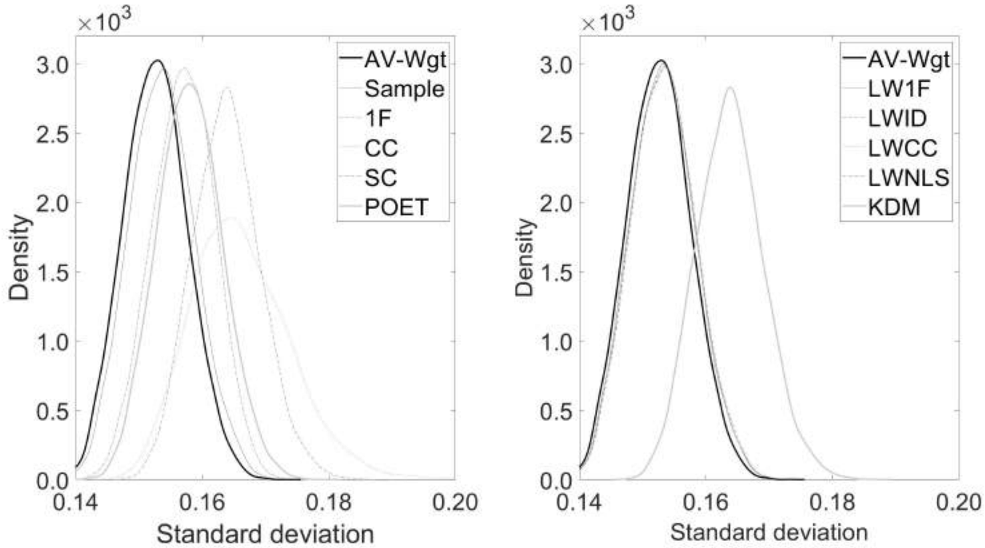
Panel B: Estimation window τ = 120



Panel C: Estimation window τ = 240



Panel D: Estimation window τ = 480



Panel E: Estimation window τ s= 960

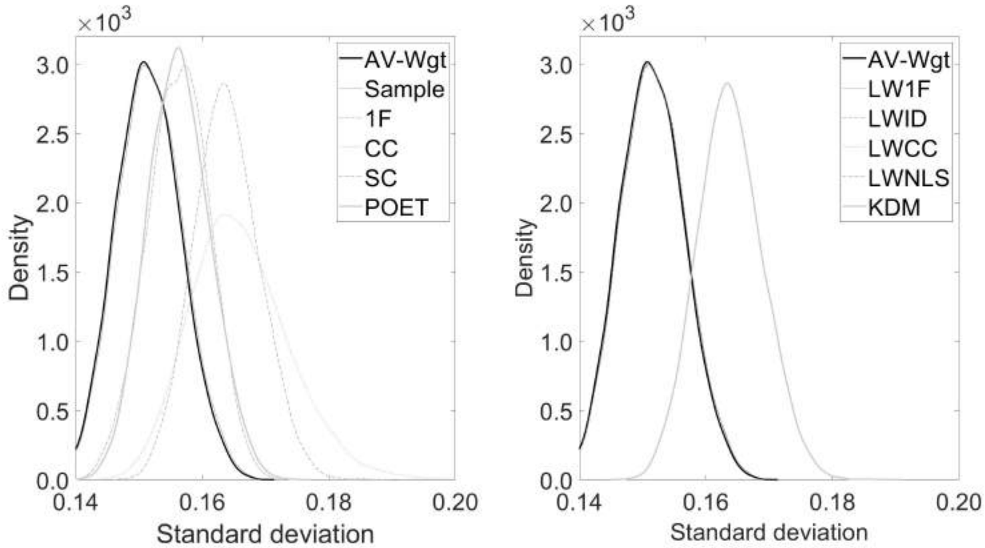
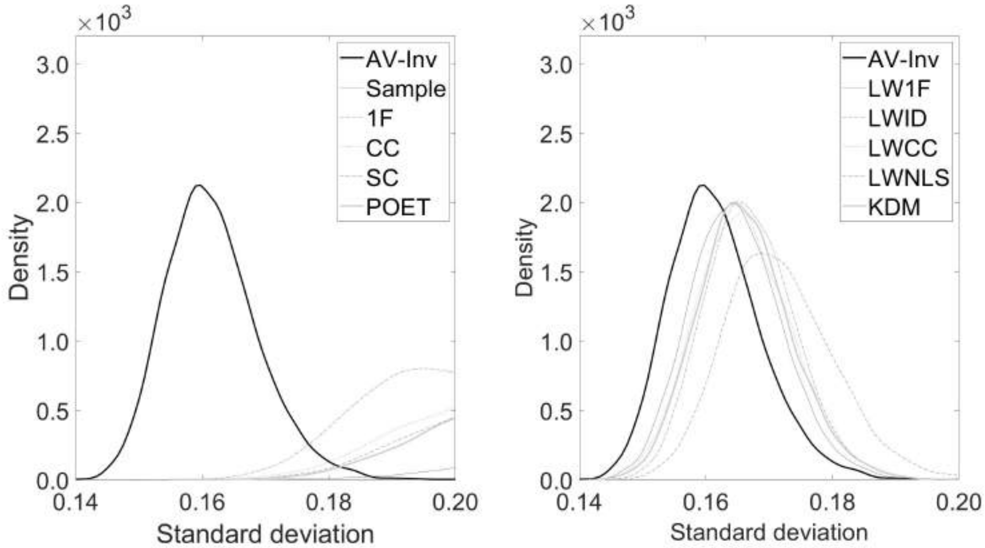


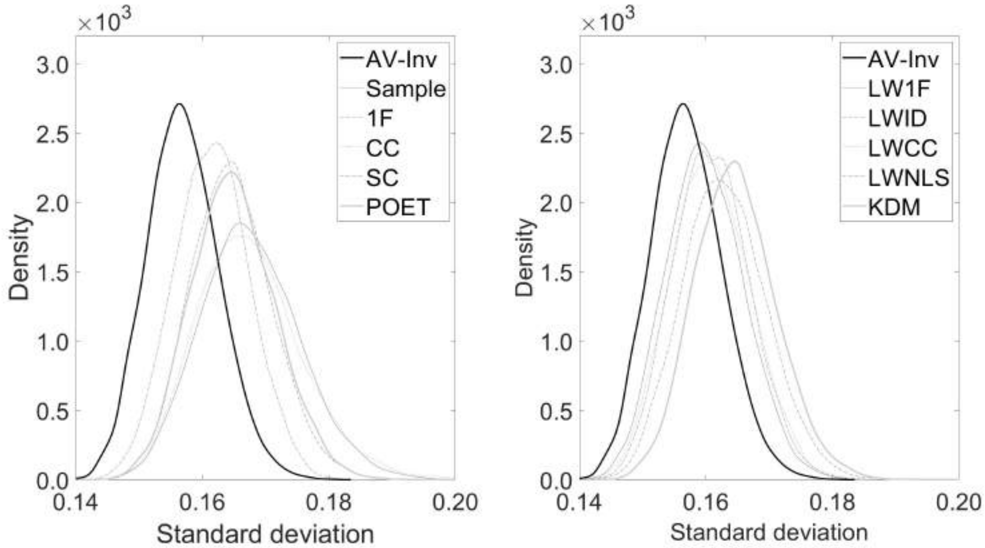
Figure A.2.5: Kernel densities of simulated standard deviations for AV-Inv compared to established strategies and benchmarks

This figure plots the densities based on the normal kernel of the annualized out-of-sample standard deviations of our averaging rule with *i.i.d* jackknife returns on the inverse covariance matrix (AV-Inv) level over 10,000 simulation runs in comparison to the strategies (left column) and benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 1. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

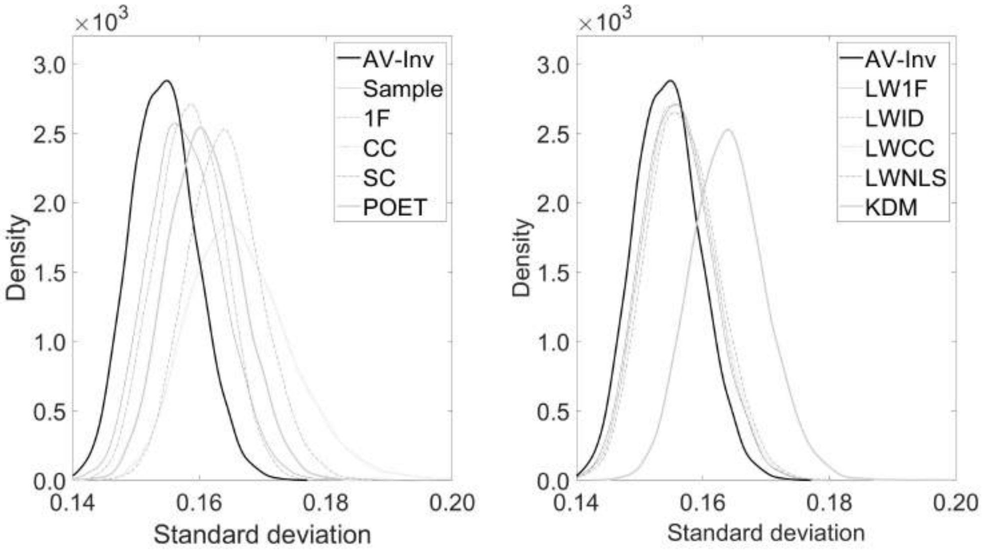
Panel A: Estimation window τ = 60



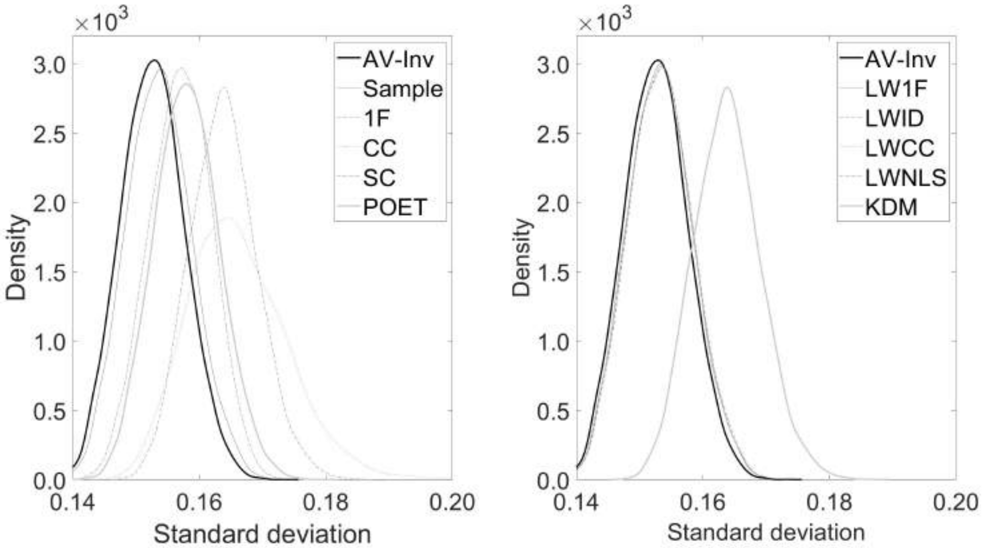
Panel B: Estimation window τ = 120



Panel C: Estimation window τ = 240



Panel D: Estimation window τ = 480



Panel E: Estimation window τ = 960

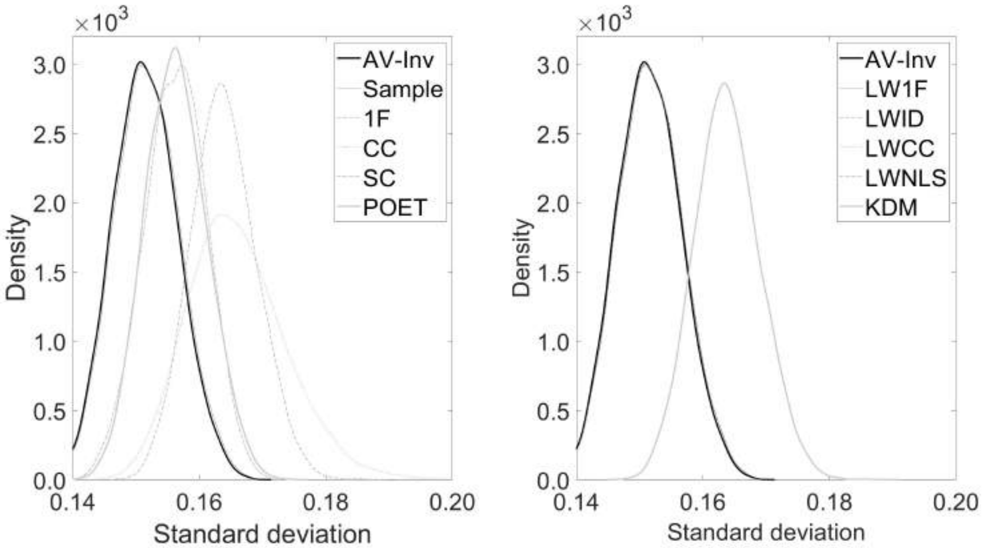
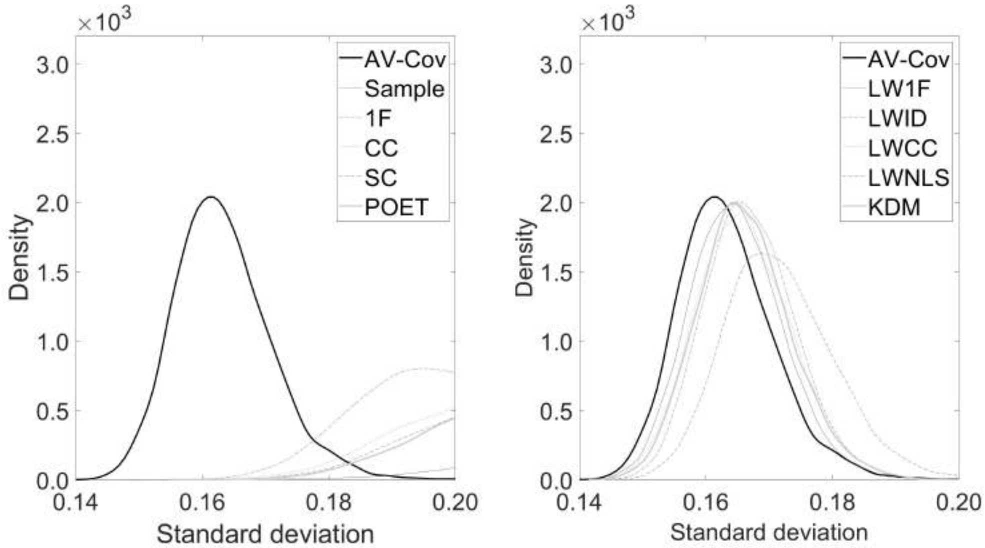


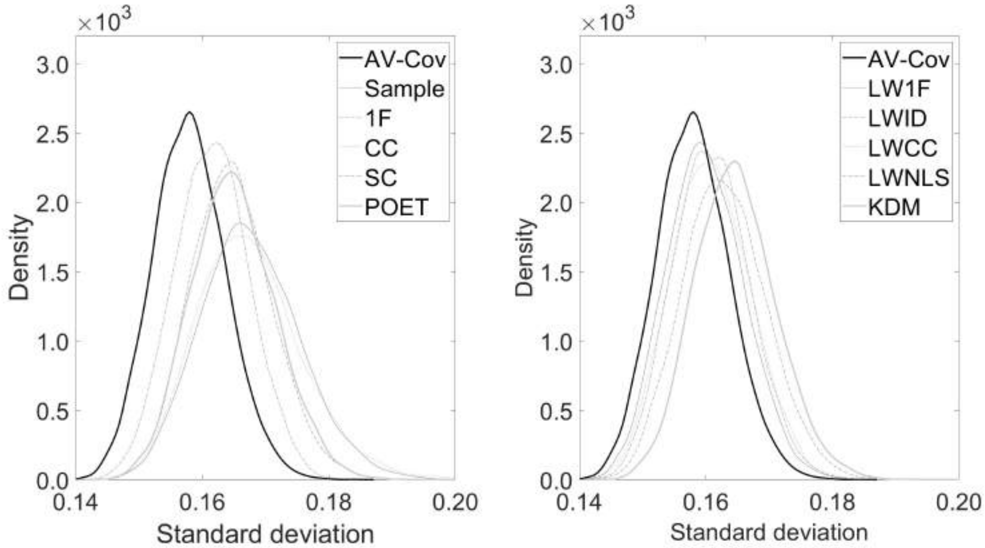
Figure A.2.6: Kernel densities of simulated standard deviations for AV-Cov compared to established strategies and benchmarks

This figure plots the densities based on the normal kernel of the annualized out-of-sample standard deviations of our averaging rule with *i.i.d* jackknife returns on the covariance matrix (AV-Cov) level over 10,000 simulation runs in comparison to the strategies (left column) and benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 1. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

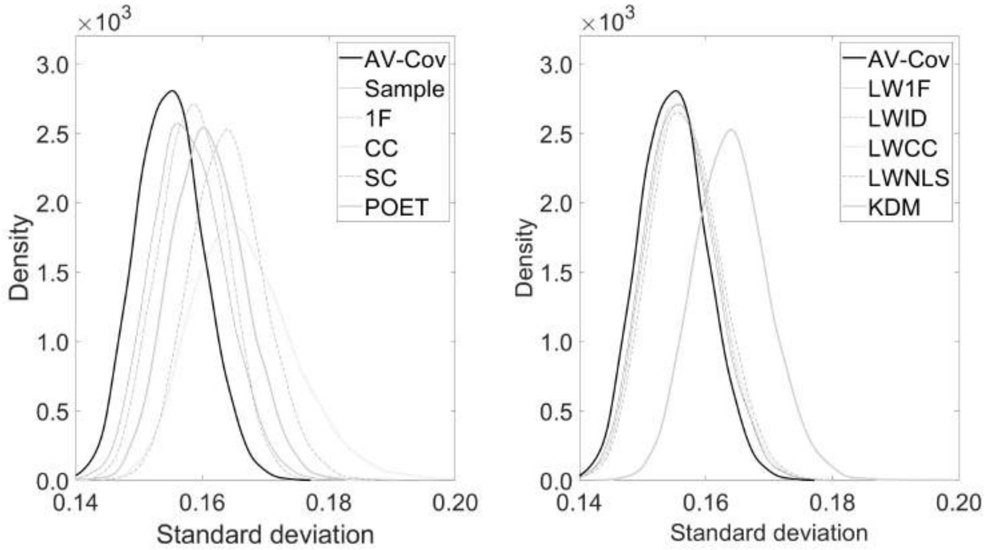
Panel A: Estimation window $\tau = 60$



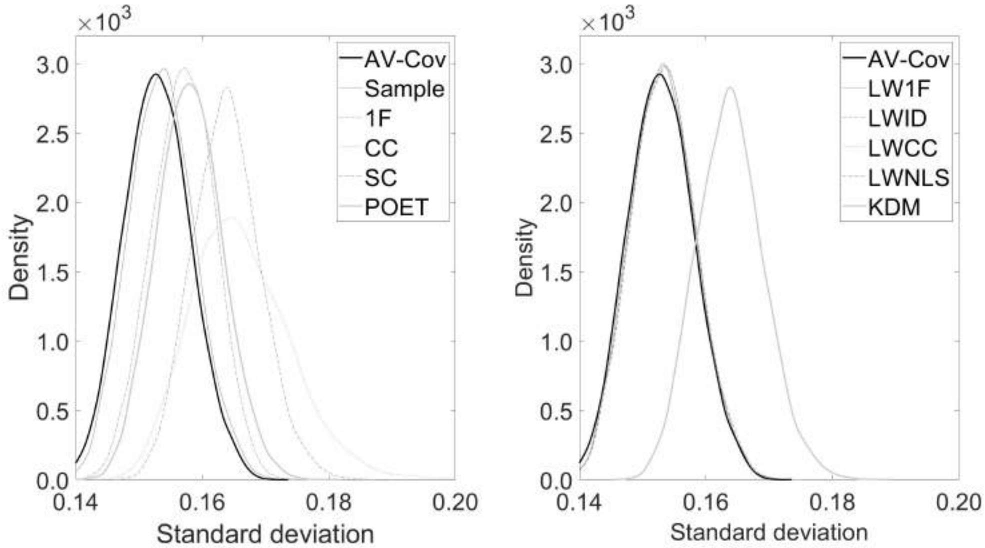
Panel B: Estimation window $\tau = 120$



Panel C: Estimation window τ = 2460



Panel D: Estimation window τ = 480



Panel E: Estimation window τ = 960

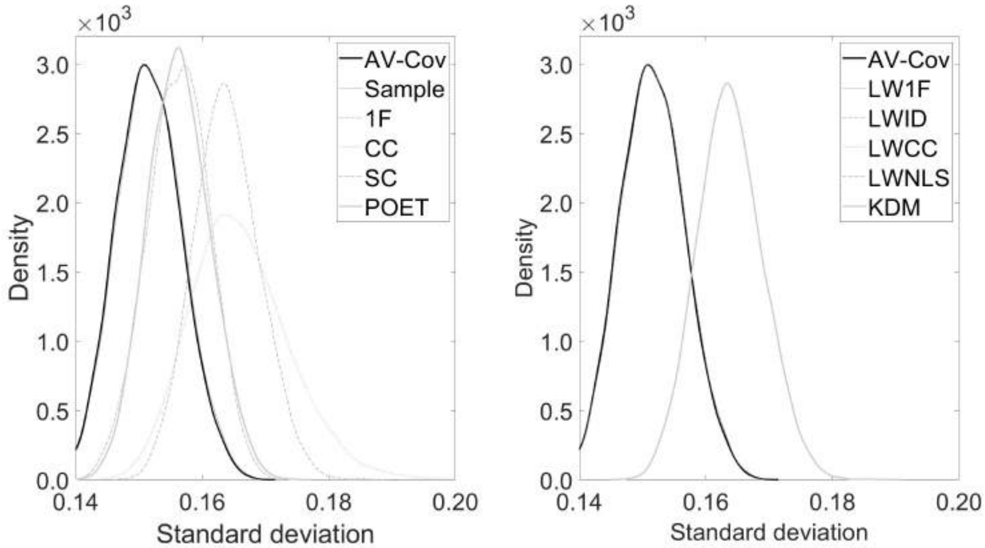
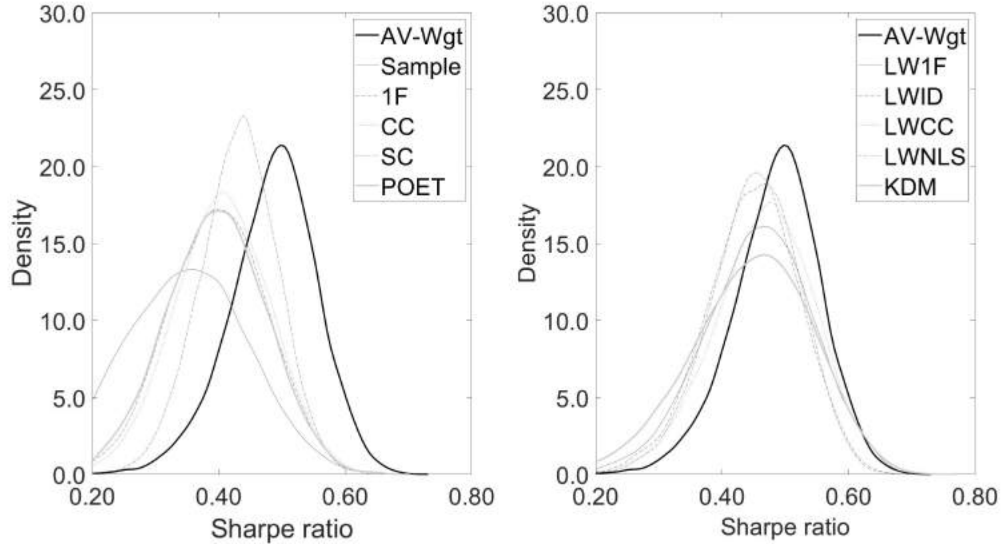


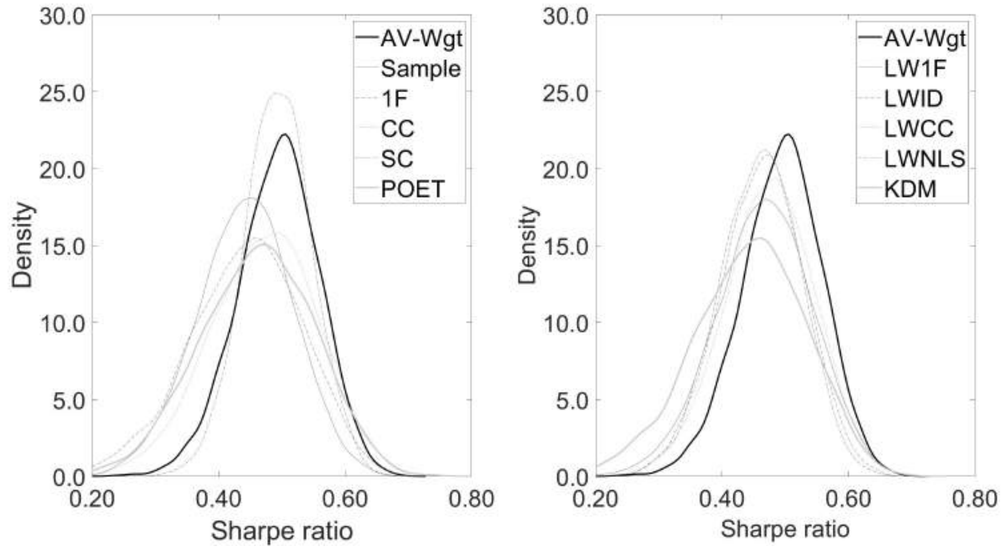
Figure A.2.7: Kernel densities of simulated Sharpe ratios for AV-Wgt compared to established strategies and benchmarks

This figure plots the densities based on the normal kernel of the annualized out-of-sample Sharpe ratios of our averaging rule with *i.i.d* jackknife returns on the portfolio weight (AV-Wgt) level over 10,000 simulation runs in comparison to the strategies (left column) and benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 1. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

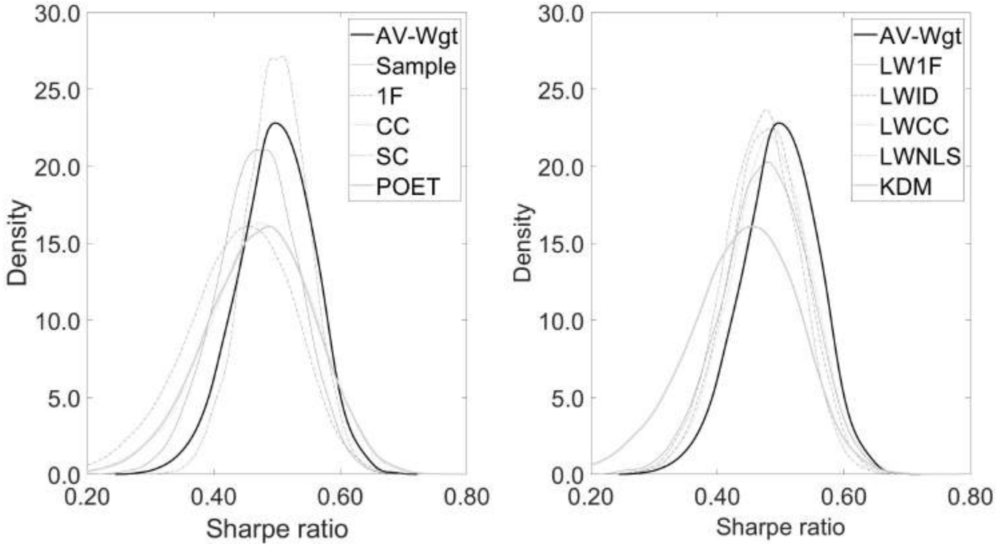
Panel A: Estimation window $\tau = 60$



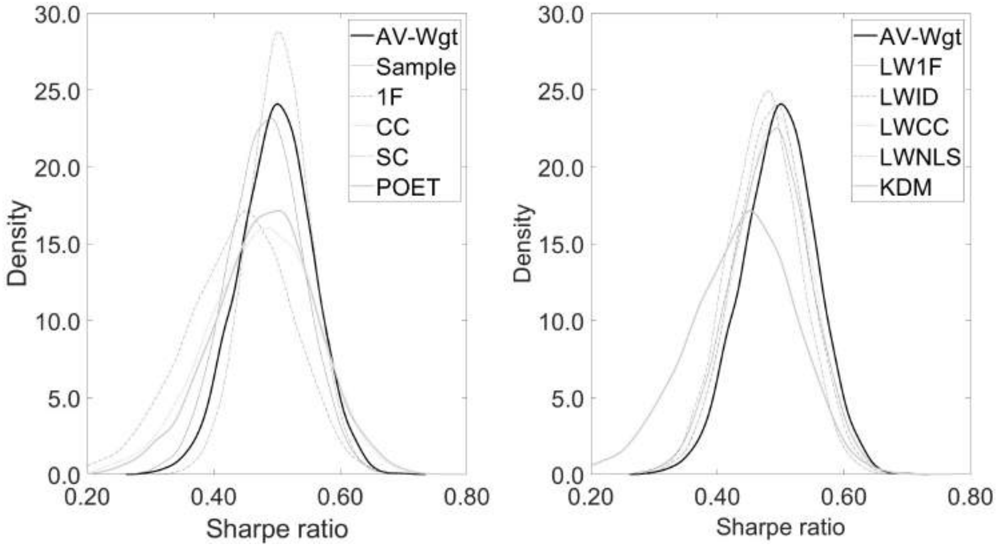
Panel B: Estimation window $\tau = 120$



Panel C: Estimation window $\tau = 240$



Panel D: Estimation window $\tau = 480$



Panel E: Estimation window $\tau = 960$

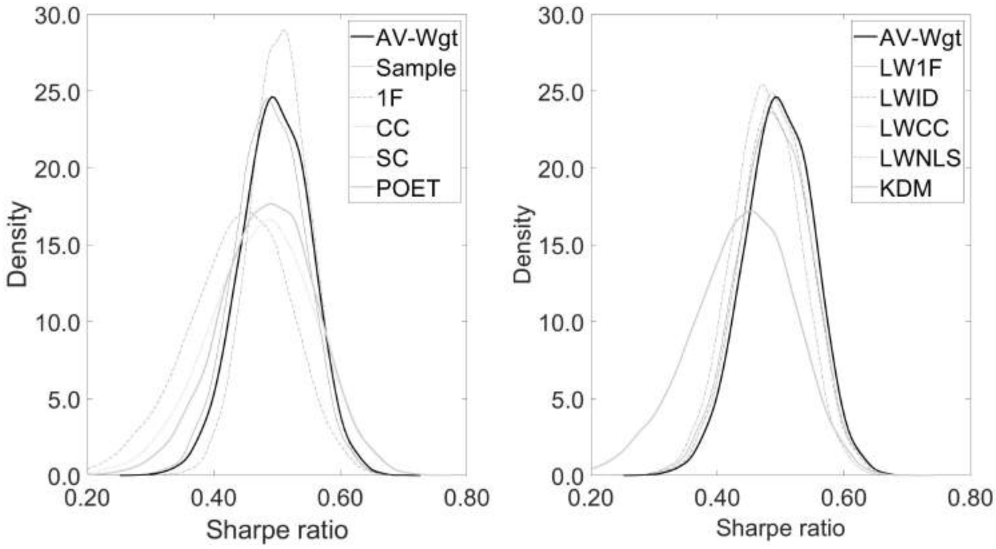
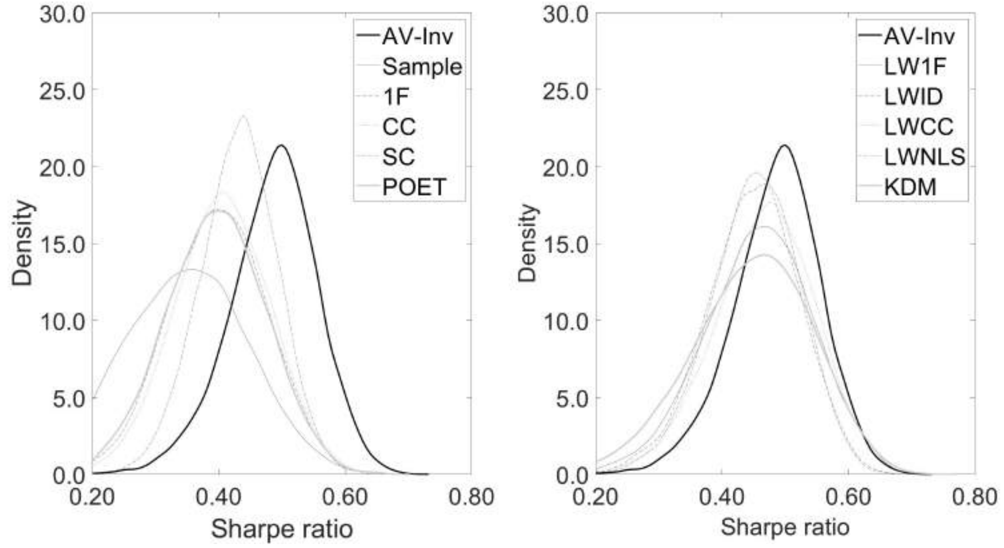


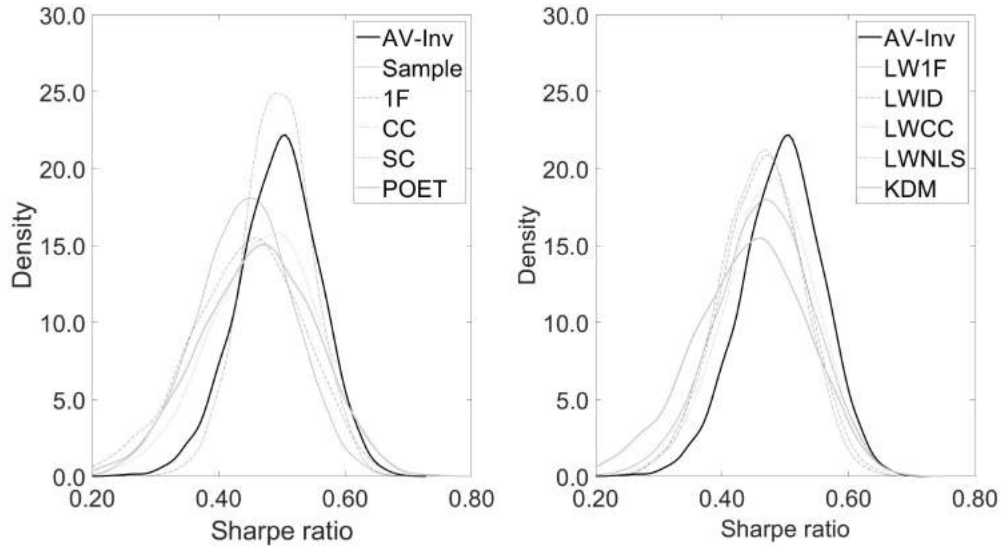
Figure A.2.8: Kernel densities of simulated Sharpe ratios for AV-Inv compared to established strategies and benchmarks

This figure plots the densities based on the normal kernel of the annualized out-of-sample Sharpe ratios of our averaging rule with *i.i.d* jackknife returns on the inverse covariance matrix (AV-Inv) level over 10,000 simulation runs in comparison to the strategies (left column) and benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Inv are based on averaging over the $M = 6$ strategies from Panel B in Table 1. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

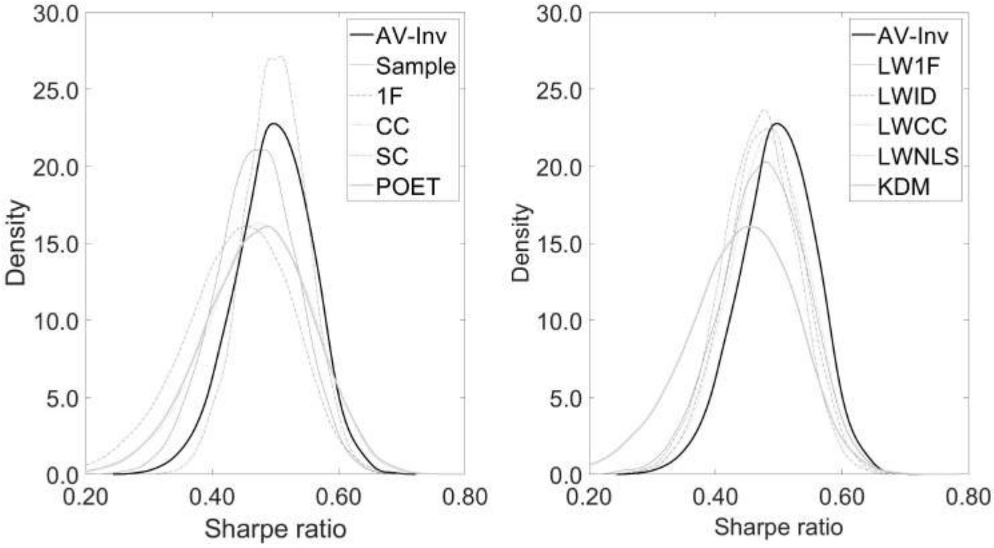
Panel A: Estimation window τ = 60



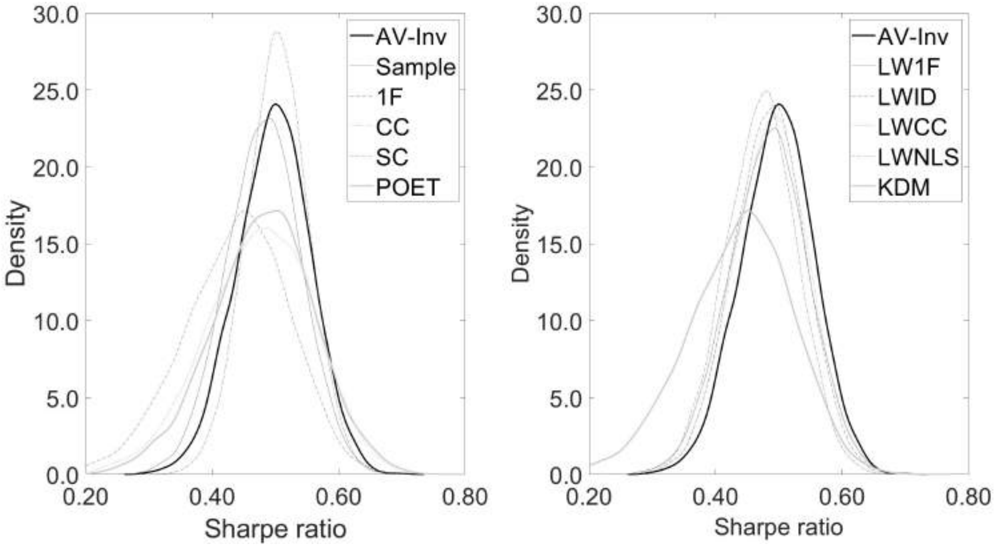
Panel B: Estimation window τ = 120



Panel C: Estimation window τ = 240



Panel D: Estimation window τ = 480



Panel E: Estimation window τ = 960

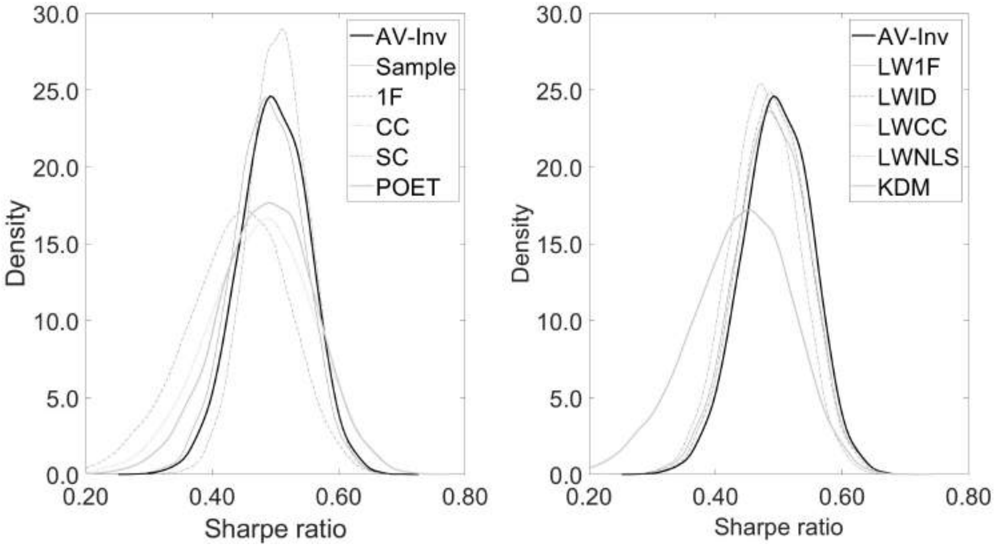
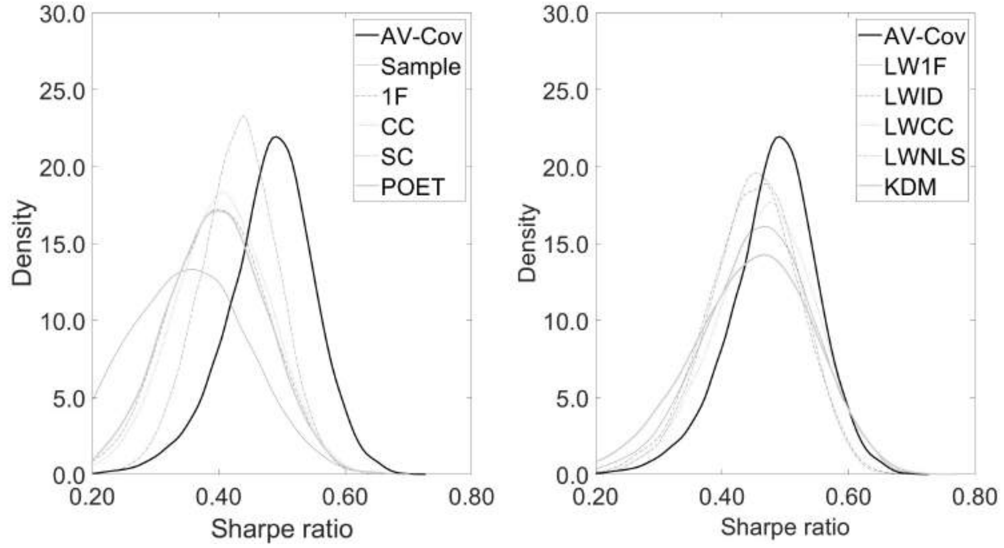


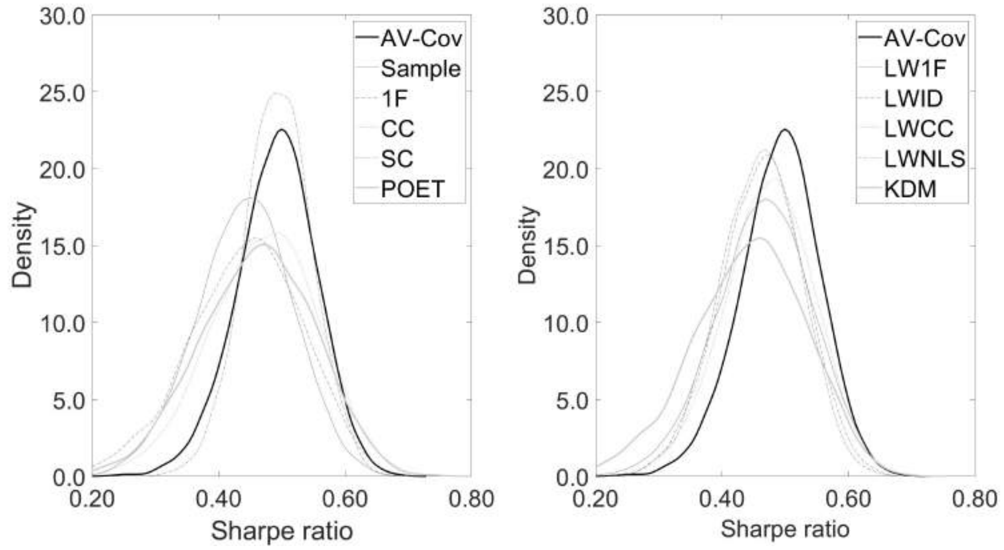
Figure A.2.9: Kernel densities of simulated Sharpe ratios for AV-Cov compared to established strategies and benchmarks

This figure plots the densities based on the normal kernel of the annualized out-of-sample Sharpe ratios of our averaging rule with *i.i.d* jackknife returns on the covariance matrix (AV-Cov) level over 10,000 simulation runs in comparison to the strategies (left column) and benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for AV-Cov are based on averaging over the $M = 6$ strategies from Panel B in Table 1. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

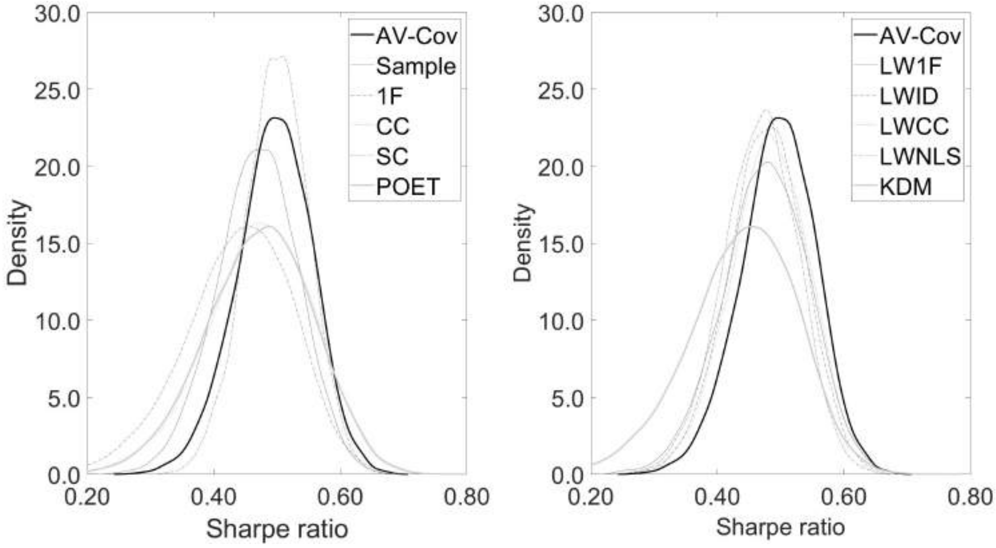
Panel A: Estimation window τ = 60



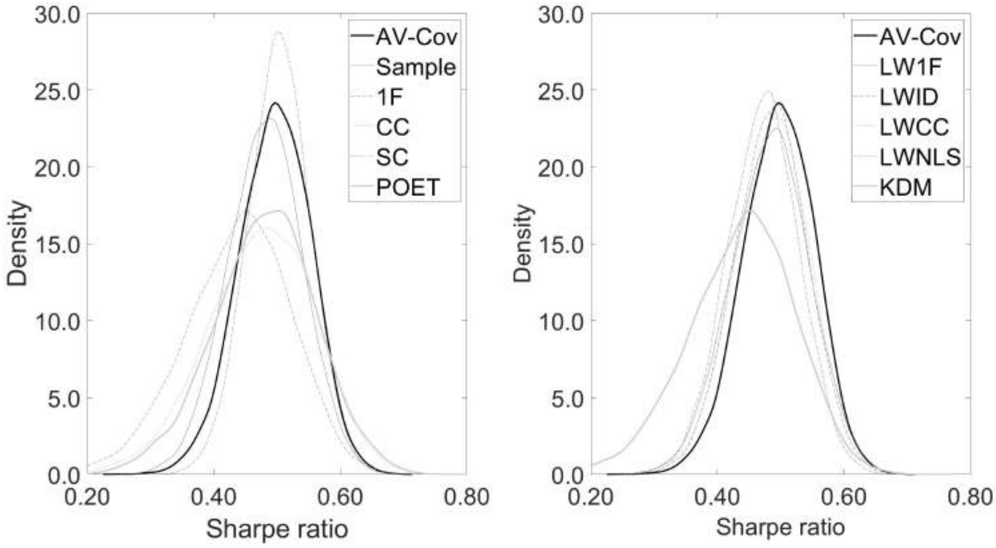
Panel B: Estimation window τ = 120



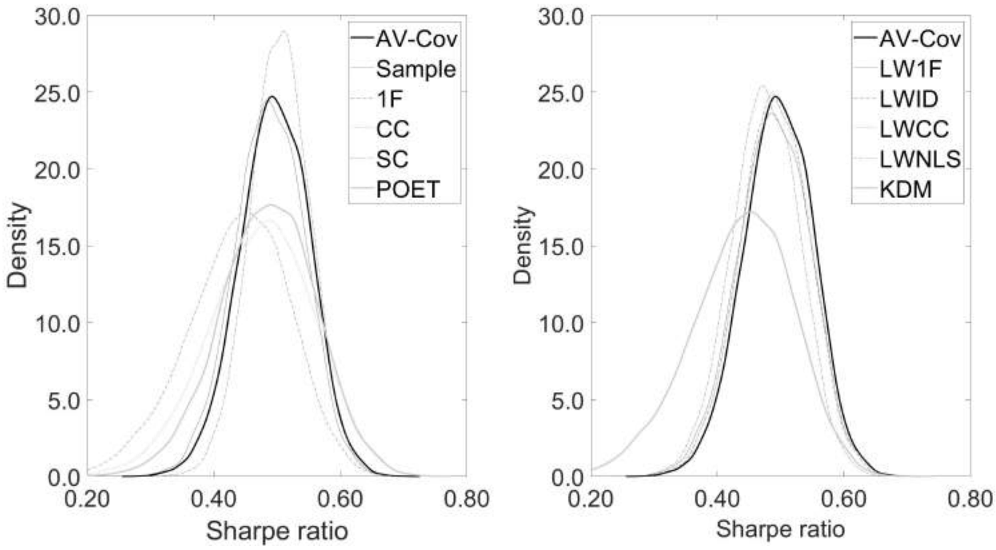
Panel C: Estimation window $\tau = 240$



Panel D: Estimation window $\tau = 480$



Panel E: Estimation window $\tau = 960$



A.3 Additional empirical results

In this section, we show additional empirical results. Table A.3.1 reports the out-of-sample Sharpe ratios under consideration of transaction costs of 50bps.

Table A.3.1: Empirical out-of-sample Sharpe ratios adjusted for transaction costs of 50bps

This table reports the annualized out-of-sample Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C, after adjustment for transaction costs. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from June 1976 - June 2019. The abbreviations for the strategies and benchmarks in Panels B and C, as well as for the data sets, are explained in Tables 1 and 7, respectively. We use the 30-day T-bill rate as the risk-free rate. The transaction costs are set to 50bps. We report statistical significance for the null hypothesis wherein the Sharpe ratio of AV-Wgt-E is less than or equal to that of the ID strategy, corresponding to the 1/N portfolio. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure with a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5 and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across three levels					
AV-Wgt-E	1.1421	0.8442	0.7222	0.5894	0.7345
AV-Inv-E	1.1083	0.7837	0.6978	0.5881	0.7074
AV-Cov-E	1.0007	0.7963	0.7169	0.6043	0.7412
AV-Wgt	1.1660	0.8723	0.7227	0.5880	0.7345
AV-Inv	1.1638	0.8730	0.7225	0.5844	0.7074
AV-Cov	1.1489	0.8670	0.7179	0.5856	0.7412
Panel B: Single strategies					
Sample	1.1631	0.9082	0.7066	0.5386	0.4429
1F	0.9223	0.6454	0.7526	0.5767	0.6163
CC	0.9581	0.6656	0.6852	0.5740	0.4568
ID	0.5389***	0.2943***	0.5181*	0.4971	0.5573
SC	0.8028	0.4727	0.6585	0.6318	0.7794
POET	1.2000	0.7572	0.6716	0.5320	0.5760
Panel C: Benchmarks					
LW1F	1.1411	0.8940	0.7114	0.5453	0.7762
LWID	1.1249	0.8784	0.7396	0.5715	0.5991
LWCC	1.0878	0.7822	0.7589	0.5491	0.6948
LWNLS	1.1517	0.8855	0.7080	0.5495	0.7281
KDM	0.9223	0.6454	0.7526	0.5767	0.6163

A.4 Results on monthly data frequency

Following Ledoit and Wolf (2017), we change the excess return frequency from daily to monthly. While keeping the out-of-sample period the same from June 1976 - June 2019, we use the longer available history to estimate the covariance matrix over $\tau = 120$ most recent months, i.e. 10 years of data, as common in the literature. The data sets as described in Table A.4.1 are the same for the Fama-French factor-mimicking and industry portfolios. We reduce the CRSP data set to the 30 largest stocks, representing the Dow Jones Industrial Average index. We thus ensure that the number of assets N remains smaller than the available monthly observations τ. In July of each year we select the 30 largest stocks in terms of market capitalization from all NYSE, AMEX, and NASDAQ stocks in the Center for Research in Security Prices (CRSP) database that fulfill the following criteria. We filter out stocks with a price of less than USD 5 as for daily observations, as well as stocks with missing excess returns in the 120 preceding months or the 12 months subsequent to the selection date. The stocks constitute the investment universe for one year. For our rolling-sample procedure, we set the estimation window to $\tau = 120$ observations, corresponding to ten years of monthly data. We estimate the portfolio weights of each strategy using only the information in the estimation window, which comprises the most recent τ excess returns. We hold the estimated portfolio weights constant for a one-month holding period, and save the corresponding out-of-sample excess returns. We then move the estimation window forward by a month and repeat the aforementioned procedure over the entire sample period. The specification is thus similar to the simulation study with an estimation window of $\tau = 120$.

We find that the results on a monthly data frequency are qualitatively similar to the ones presented in Section 4 for daily observations. Our averaging rule achieves on all averaging levels comparable standard deviations and Sharpe ratios. Comparing AV-Wgt-E to the benchmarks in Panel C of Table A.4.2, we find that AV-Wgt-E achieves on all five data sets lower standard deviations than LWNLS, being significantly lower for four out of five. The standard deviation is also always lower than the other shrinkage estimators LW1F, LWID, LWCC, or KDM. In terms of out-of-sample risk-adjusted performance, AV-Wgt-E achieves consistently higher Sharpe ratios compared to the ID strategy in Panel B of Table A.4.3. In most cases, the performance is similar to the benchmarks in Panel C. The monthly average turnover for AV-Wgt-E as shown in Table A.4.4 is smaller than for the benchmarks in the 30Ind data set but tends to be larger in the remaining data sets. The short interest of AV-Wgt-E in Table A.4.5 is the smallest of all strategies (excluding ID and SC) in the 30-industry and the STOCK30 data sets. The short interest of our rule is on a similar level as LW1F and LWNLS on the remaining data sets.

Table A.4.1: List of the empirical data sets for monthly data

This table lists the data sets for the empirical evaluation of our proposed averaging rule in comparison to existing portfolio strategies, their abbreviations, the number of assets in each data set, and the data sources. All data sets span the period from June 1966 - June 2019, comprise monthly data, and apply in the case of portfolio data the value weighting scheme to the respective constituents. Data from Kenneth French is taken from his website (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`) and represents different cuts of the U.S. stock market. The STOCK30 data set contains the 30 largest single stocks in terms of market capitalization in July of every year after filtering out stocks that have a price of less than $5, or missing returns in the preceding 120 and subsequent 12 months to the selection date. All stock prices are taken from the Center of Research in Security Prices (CRSP).

#	Data set	Abbreviation	N	Source
1	6 Fama and French portfolios of firms sorted by size and book-to-market	6FF	6	K. French
2	25 Fama and French portfolios of firms sorted by size and book-to-market	25FF	25	K. French
3	10 industry portfolios representing the U.S. stock market	10Ind	10	K. French
4	30 industry portfolios representing the U.S. stock market	30Ind	30	K. French
5	30 Stocks with the largest market capitalization	STOCK30	30	CRSP

Table A.4.2: Empirical out-of-sample standard deviations based on monthly observations

This table reports the annualized out-of-sample standard deviation of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from June 1976 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. We report statistical significance for the null hypothesis wherein the (log) expected out-of-sample portfolio variance of the averaging rule on the portfolio weight level with exponentially smoothed jackknife returns, AV-Wgt-E, is greater than or equal to that of the non-linear shrinkage approach LWNLS of Ledoit and Wolf (2017). We follow Ledoit and Wolf (2011) and use their proposed bootstrap procedure with a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5, and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt-E	0.1289	0.1209	0.1147	0.1121	0.1238
AV-Inv-E	0.1308	0.1286	0.1155	0.1173	0.1254
AV-Cov-E	0.1368	0.1248	0.1140	0.1136	0.1232
AV-Wgt	0.1295	0.1213	0.1152	0.1130	0.1241
AV-Inv	0.1295	0.1214	0.1153	0.1128	0.1241
AV-Cov	0.1309	0.1213	0.1154	0.1133	0.1242
Panel B: Single strategies					
Sample	0.1310	0.1285	0.1171	0.1230	0.1353
1F	0.1570	0.1888	0.1229	0.1249	0.1363
CC	0.1525	0.1980	0.1276	0.1406	0.1385
ID	0.1629	0.1690	0.1410	0.1577	0.1458
SC	0.1452	0.1413	0.1168	0.1177	0.1239
POET	0.1337	0.1262	0.1184	0.1187	0.1299
Panel C: Benchmarks					
LW1F	0.1319	0.1237	0.1161	0.1146	0.1266
LWID	0.1312	0.1205	0.1144	0.1144	0.1264
LWCC	0.1456	0.1519	0.1165	0.1168	0.1236
LWNLS	0.1310**	0.1238**	0.1159	0.1156**	0.1265**
KDM	0.1570	0.1888	0.1229	0.1249	0.1363

Table A.4.3: Empirical out-of-sample Sharpe ratios based on monthly observations

This table reports the annualized out-of-sample Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from June 1976 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. We use the 30-day T-bill rate as the risk-free rate. We report statistical significance for the null hypothesis wherein the Sharpe ratio of the averaging rule on the portfolio weight level with exponentially smoothed jackknife returns, AV-Wgt-E, is less than or equal to that of the ID strategy, corresponding to the 1/N portfolio. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure with a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5, and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt-E	0.9102	0.8945	0.6952	0.6512	0.5751
AV-Inv-E	0.8553	0.7778	0.6754	0.6280	0.5838
AV-Cov-E	0.7006	0.8269	0.6916	0.6675	0.5932
AV-Wgt	0.9241	0.9119	0.7049	0.6485	0.5774
AV-Inv	0.9289	0.9131	0.7031	0.6478	0.5792
AV-Cov	0.9228	0.9258	0.6945	0.6409	0.5659
Panel B: Single strategies					
Sample	0.9695	1.0133	0.7009	0.5573	0.3819
1F	0.5441	0.5569	0.7356	0.6452	0.6055
CC	0.5572	0.5045	0.6708	0.6273	0.5659
ID	0.5651***	0.5735***	0.5824	0.5318	0.5467
SC	0.6150	0.6291	0.6803	0.6320	0.6024
POET	1.0861	0.9389	0.6298	0.6005	0.6335
Panel C: Benchmarks					
LW1F	0.9277	0.9785	0.7103	0.6339	0.4857
LWID	0.7996	0.9539	0.7241	0.6869	0.4969
LWCC	0.6538	0.7455	0.7024	0.6624	0.5385
LWNLS	0.9006	0.8737	0.6974	0.6623	0.5486
KDM	0.5441	0.5569	0.7356	0.6452	0.6055

Table A.4.4: Empirical average monthly turnover based on monthly observations

This table reports the average monthly turnover of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from June 1976 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. Turnover is measured as the average percentage of total wealth traded in each month. The numbers are reported in percentages.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt-E	18.81	41.74	10.97	20.95	40.50
AV-Inv-E	17.21	31.95	10.19	20.40	37.85
AV-Cov-E	9.73	23.15	8.92	18.70	37.85
AV-Wgt	18.23	40.99	11.02	20.52	40.27
AV-Inv	18.47	41.57	10.99	20.28	40.75
AV-Cov	23.865	39.87	11.56	21.35	40.75
Panel B: Single strategies					
Sample	21.88	79.46	15.47	45.92	29.74
1F	12.36	24.54	9.25	14.79	12.69
CC	10.32	23.23	7.54	13.96	10.64
ID	1.47	1.67	2.26	2.85	4.36
SC	4.09	6.96	4.87	6.74	8.98
POET	27.22	43.96	16.37	23.96	17.85
Panel C: Benchmarks					
LW1F	19.73	55.67	13.60	29.33	19.06
LWID	9.03	32.09	10.41	26.35	18.92
LWCC	10.89	28.60	9.63	22.20	14.26
LWNLS	19.56	54.99	13.78	28.95	19.14
KDM	12.36	24.54	9.25	14.79	12.69

Table A.4.5: Empirical average monthly short interest based on monthly observations

This table reports the average monthly short interest of our averaging rule on the portfolio weight (AV-Wgt-E / AV-Wgt), the inverse (AV-Inv-E / AV-Inv), and the covariance (AV-Cov-E / AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. Averaging rules with suffix $-E$ are constructed with exponentially smoothed jackknife returns. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from June 1976 - June 2019. The abbreviations for the averaging rules, strategies, and benchmarks, as well as for the data sets, are explained in Tables 1 and 7, respectively. Short interest is measured by the average amount of wealth that is held in short positions. The numbers are reported in percentages.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt-E	106.38	193.28	36.13	56.71	27.64
AV-Inv-E	89.80	115.67	25.32	54.37	20.56
AV-Cov-E	42.47	102.57	21.01	52.27	25.61
AV-Wgt	108.23	195.87	36.47	57.28	27.98
AV-Inv	111.16	198.87	35.54	56.94	27.93
AV-Cov	107.66	199.12	35.76	60.05	29.03
Panel B: Single strategies					
Sample	140.33	328.78	64.29	159.97	85.36
1F	88.05	158.19	53.68	82.89	44.79
CC	76.54	144.93	41.03	77.17	38.58
ID	0.00	0.00	0.00	0.00	0.00
SC	0.00	0.00	0.00	0.00	0.00
POET	193.73	269.34	70.25	96.15	53.75
Panel C: Benchmarks					
LW1F	123.08	260.15	59.75	113.88	55.48
LWID	63.08	171.05	41.86	97.42	51.26
LWCC	75.19	159.06	38.56	86.43	40.78
LWNLS	116.09	253.59	56.81	105.21	51.73
KDM	88.05	158.18	53.68	82.89	44.79