Erschienen in: KNOTH, Sven, Hrsg., Yarema OKHRIN, Hrsg., Philipp OTTO, Hrsg.. Advanced Statistical Methods in Process Monitoring, Finance, and Environmental Science : Essays in Honour of Wolfgang Schmid. 1. Cham: Springer, 2024, S. 337-366. ISBN 978-3-031-69110-2
https://doi.org/10.1007/978-3-031-69111-9_16

Can We Give the Maximum Sharpe Ratio Portfolio a Chance?

Winfried Pohlmeier and Ekaterina Kazak

Abstract This chapter studies the applicability of the maximum Sharpe ratio (MaxSR) portfolio strategy in real-world settings. As shown by Okhrin and Schmidt the plug-in estimated weights show abysmal distributional properties such that it renders an application impossible for financial practitioners. In this chapter we propose a double regularization approach for the MaxSR portfolio strategy based on the bagged pretested portfolio selection (BPPS) algorithm. We show that for certain settings the doubly shrunken portfolio weights strongly mitigate the adverse properties of the plug-in estimated weights and can beat the popular 1/N benchmark strategy.

1 Introduction

Among practitioners in the finance industry the Sharpe ratio is one of the most popular measures to evaluate portfolio performance. Its computation is straightforward and its interpretation as the expected return per unit of risk does not even require knowledge of basic modern portfolio theory. Therefore, a portfolio strategy targeting to maximize the Sharpe ratio, the maximum Sharpe ratio (MaxSR) portfolio strategy, seems to be a natural candidate of choice among the plethora of possible portfolio strategies.

However, the crucial role of estimation risk for the performance of different portfolio strategies has been pointed out for a long time by numerous scholars (e.g., Merton, 1980; Michaud, 1989; Best & Grauer, 1991; Chopra & Ziemba, 1993). Simple plug-in estimated portfolio weights, where the mean return vector and the

W. Pohlmeier (✉)
Department of Economics, University of Konstanz, CoFE, ICEA, Konstanz, Germany
e-mail: Winfried.Pohlmeier@uni-konstanz.de

E. Kazak
Department of Economics, University of Birmingham, Birmingham, UK
e-mail: e.kazak@bham.ac.uk

variance-covariance matrix are replaced by their empirical counterparts suffer from serious shortcomings such as large standard errors of the weight estimates, high transaction costs, and extremely unrealistic short positions which as a result lead to poor out-of-sample performance and low-powered performance tests (e.g., Jobson & Korkie, 1980; Kazak & Pohlmeier, 2019).

These empirical deficiencies, which generally occur for any portfolio strategy aggravate dramatically when it comes to estimating the portfolio weights of the MaxSR strategy. Even in comparison to other rather unrealistic plug-in estimated weights, the MaxSR portfolio weights for realistic data scenarios and portfolio dimensions are such that they are far from being usable for any application in the finance world.

It was the seminal contribution of Okhrin and Schmid (2006) which showed that the MaxSR portfolio weights have no moments in finite samples. Assuming iid multivariate normality of the return process, they derive the finite sample distribution of the MaxSR portfolio weights and prove that the plug-in estimator has no first and higher moments. To some extent the Okhrin-Schmidt result did not receive the attention it is ought in the empirical finance literature. In fact, this suggests the assumption that the obvious lack of empirical studies on the MaxSR strategy is a consequence of its abysmal distributional properties. Although the choice of a portfolio strategy, which is directly targeted to maximize the Sharpe ratio, seems to be a natural candidate in empirical "horse race" studies with the Sharpe ratio as performance measure, it seems to be somewhat surprising that the MaxSR strategy is generally not present among the competing "horses." We believe that this observation can only be explained by the exceptionally poor performance of the MaxSR strategy, which refrains authors from including it in the empirical performance comparisons.

By deriving the finite sample properties of the plug-in estimated MaxSR portfolio weights Okhrin and Schmid provide a rigorous theoretical explanation of the exceptionally poor finite sample performance of the MaxSR portfolio strategy with large standard errors and large outliers of the estimated portfolio weights. Although their finding is based on the rather unrealistic multivariate normality assumption, there are no good reasons to assume that under more general distributional assumptions for the joint return process (e.g., multivariate t-distribution with time-varying conditional moments), the distributional properties of the MaxSR portfolio weights would become more favorable. Furthermore, Schmid and Zabolotskyy (2008) show that an unbiased estimator of the MaxSR portfolio weights does not exist at all. Even asymptotically no unbiased estimator of MaxSR weights exists within the family of estimators which are bounded by cylinder functions.

To mitigate the shortcomings of empirical portfolio models and to robustify portfolio estimates against too large estimation noise, a range of alternative estimation approaches have been proposed including (i) the regularization of the moment estimates or the portfolio weights, (ii) Bayesian estimation strategies approaches, or (iii) combination strategies. While these strategies show some improvements in portfolio performance over simple plug-in approaches, these strategies were applied to the less challenging strategies but not the MaxSR strategy.

This study takes up the challenge of estimating more realistic MaxSR portfolio weights by means of a novel double shrinkage approach, where both the mean vector and the variance covariance matrix of the returns are regularized. The optimal portfolio weights are obtained by the bagged pretested portfolio selection estimator (BPPS) recently proposed by Kazak and Pohlmeier (2023). This estimation approach achieves more stable portfolio weight estimates by the means of weighted model averaging strategy where the weights result from bootstrap aggregation of the pretest estimator. The bagging step in their estimation algorithm smooths the typical sharp thresholding effects of pretesting approaches and therefore reduces the transaction costs and the occurrence of extreme outliers substantially.

Although originally proposed as a model averaging approach which combines different portfolio strategies, the BPPS can also serve as a data-driven mechanism to choose an optimal out-of-sample penalization for a given shrinkage approach. In addition, the BPPS algorithm allows for a dynamic adjustment of the penalty parameter within a rolling window approach by memorizing the penalty values of previous estimation windows.

In this study we show that the BPPS can be extended to the case where two shrinkage parameters are selected simultaneously by means of bagged pretesting. The doubly shrunken portfolio weights nest simple alternatives like the equally weighted 1/N-portfolio strategy as a limiting case. Therefore, the BPPS can serve as natural strategy to check empirically whether there exists a competitive shrunken MaxSR approach at all.

The chapter is organized as follows. Section 2 illustrates the difficulties in estimating properties of the MaxSR weights. Rather than concentrating on the distributional properties of the MaxSR portfolio weights, as has been done by Okhrin and Schmid, we try to shed more light in the numerical features of weights. Even without any estimation risk the MaxSR weights are shown to be highly sensitive to slight and realistic changes in the population covariance matrix. By introducing regularized versions of mean and covariance we can show that several popular portfolio strategies can be treated as limiting cases of a doubly regularized MaxSR strategy. Section 3 introduces the bagged pretest estimator and shows how optimal penalty terms can be selected by the BPPS algorithm. Section 4 provides an empirical illustration of the proposed method. Section 5 summarizes the main findings and gives an outlook on future research.

2 The Mess with the Maximum Sharpe Ratio Portfolio

The focus of our study is the MaxSR strategy for a one-period investment problem for a universe of N risky assets. Let r be the $N \times 1$ vector of returns with mean $\mathrm{E}[r] = \mu$ and variance-covariance matrix $\mathrm{V}[r] = \Sigma$, such that the return $r_p = w'r$ of a portfolio has mean $\mu_p = w'\mu$ and variance, $\sigma_p^2 = w'\Sigma w$, where w denotes the $N \times 1$ vector of portfolio weights. Accordingly, for the N-risky asset case the

Sharpe ratio is given by $SR(w) = \frac{\mu_p}{\sigma_p} = \frac{w'\mu}{(w'\Sigma w)^{1/2}}$. The MaxSR portfolio weights w_{sr} maximize the Sharpe ratio subject to the adding-up constraint:

$$w_{sr} = \underset{w,\, s.t. w'\iota=1}{\arg\max} \frac{w'\mu}{(w'\Sigma w)^{1/2}} = \frac{\Sigma^{-1}\mu}{\iota'\Sigma^{-1}\mu}, \tag{1}$$

where ι denotes an N-dimensional column vector of ones.[1] The solution given by (1) is the conventional "textbook" solution assuming the portfolio constant, (e.g., Pennacchi, 2008), $a = \iota'\Sigma^{-1}\mu > 0$. For this solution the Sharpe ratio reaches its maximum and is given by $SR(w_{sr}) = \sqrt{\mu'\Sigma^{-1}\mu}$. However, this condition does not necessarily hold theoretically, and it is more likely to fail for its empirical counterpart. Maller and Turkington (2003) point out that for $a = \iota'\Sigma^{-1}\mu \leq 0$, the Sharpe ratio is minimized. Moreover, they show that $SR(w_{sr}) = \sqrt{\mu'\Sigma^{-1}\mu - (\iota'\Sigma^{-1}\mu)^2/(\iota'\Sigma^{-1}\iota)}$ is the supremum of the SR for a negative portfolio constant a.

2.1 Numerical Issues

Besides the fundamental problem of estimating and testing the MaxSR weights as pointed out by Okhrin and Schmid (2006), it is important to emphasize that the theoretical weights are highly sensitive for realistic parameter values, i.e., the weights react considerably to small changes in Σ or a and attain values that are basically impractical for real-world applications. To demonstrate this sensitivity, we have generated a multivariate return distribution which is calibrated on the daily return data of the S&P 100 constituents. We choose the data generating process to be as closely calibrated to the daily data as possible, as this numerical example sheds light on the empirical performance of the MaxSR strategy introduced in Sect. 4 below. In particular, the conditional distribution of the return distribution is multivariate normal, $r_t|\mathcal{F}_{t-1} \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma_t)$, with a conditional variance following a multivariate HAR-DRD process.[2] Due to the time-varying conditional variances the solution of the MaxSR is different at each point of time, where we keep the mean vector μ constant to disentangle the effects of the two moments. Figure 1 depicts the distribution of the theoretical portfolio weights $w_{sr}(t)$ for a portfolio of $N = 10$ assets over 504 observations, which is equivalent to 2 years of trading days. Even without any estimation noise and transaction costs, the portfolio weights take on

[1] Note that, for the case of N risky assets and a risk-free asset, a similar expression arises where μ_p is the mean of the excess returns over the risk-free rate. For that case w_{sr} is defined by the tangency point to the efficient frontier and is obtained by normalizing the utility maximizing portfolio weights such that they add up to one; see, e.g., Britten-Jones (1999).

[2] For explicit expressions see Appendix "Simulation Design".

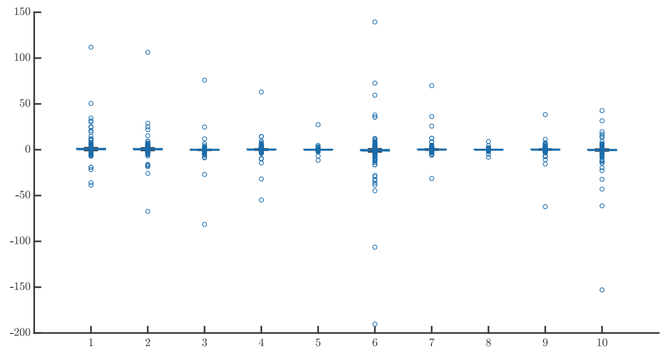

Fig. 1 Theoretical MaxSR weights from a HAR process. Boxplots of the theoretically optimal MaxSR weights over a time span of 504 periods for $N = 10$. The underlying return process follows a conditional multivariate normal distribution with constant mean and a HAR-DRD covariance matrix. The parameters of the data generating process are calibrated on daily return data of S&P100 constituents from Jan 2014 to Dec 2021. Each boxplot corresponds to the distribution of a portfolio weight $i = 1, \ldots, 10$

many absurd positive and negative values. From a practitioner's point of view the investments based on the MaxSR strategy are practically infeasible just by the pure size of the optimal weights. As shown in Fig. 2 this feature of extreme portfolio weights remains for increasing portfolio dimensions as well. The gross exposure of a portfolio vector, defined as the ℓ_1-norm of the weight vector, is a measure for the magnitude of the short sales an investor is facing (see, e.g., Brodie et al., 2009; Fan et al., 2012). It serves as an indicator for the size of the risk approximation error. Figure 2 depicts the distribution of the gross exposure of the MaxSR portfolio weights $w_{sr}(t)$ for different portfolio dimensions. A gross exposure of 1 corresponds to a portfolio with no short positions and corresponds to the horizontal red line on the graph. For all portfolio dimensions the median gross exposure is very high and varies between 3 and 5 indicating that the MaxSR strategy is generally rather aggressive. Moreover, we observe a considerable number of weight vectors with gross exposure larger than 10 calling for extremely speculative and unrealistic portfolio strategies.

Table 1 reports the descriptive statistic of the optimal MaxSR weights, averaged across N.[3] The distribution of optimal portfolio weights is extremely heavy-tailed with an average kurtosis of 36 for $N = 90$. Moreover, the distribution of the weights is skewed, where both right and left skewness can occur. For a realistic data generating process the distribution of the MaxSR weights cannot be approximated by the usual location-scale distributions. The behavior of the estimated portfolio weights can therefore be partially attributed to the distributional properties of the optimal weights.

[3] For each portfolio size N, we obtain N sample skewnesses s_i and kurtosis k_i computed over a time span of 504 observations for every weight $w_{t,i}$, $i = 1, \ldots, N$. We report the average, minimum, and maximum measures over N, e.g., $\sum_{i=1}^{N} s_i/N$.

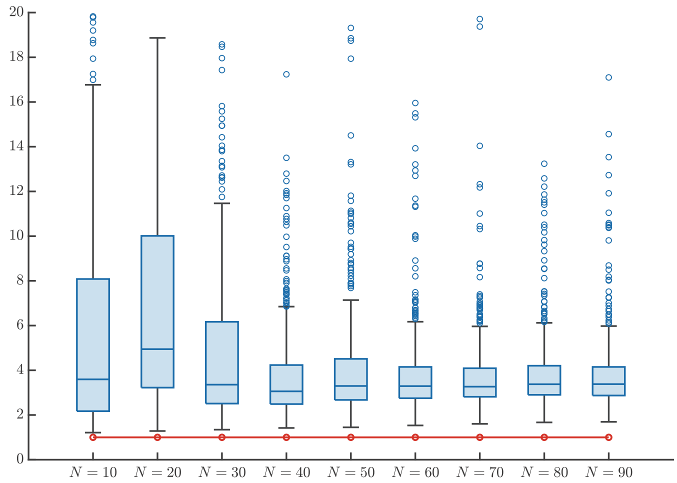

Fig. 2 Gross exposure of theoretical MaxSR weights from a HAR process. Boxplots of the gross exposure of the theoretically optimal MaxSR weights over a time span of 504 periods. The underlying return process follows a conditional multivariate normal distribution with constant mean and a HAR-DRD covariance matrix. The parameters of the data generating process are calibrated on daily return data of S&P100 constituents from Jan 2014 to Dec 2021. Each boxplot corresponds to the distribution of the gross exposure of different portfolio sizes N (x-axis). The red line plots the level of one, which is equivalent to a portfolio with no short positions. To improve readability the y-axis is limited to the range of [0, 20], even though there are outliers in the range of [0, 200]

Table 1 Descriptive statistic of theoretical MaxSR weights from a HAR process. Numbers in the table correspond to the average sample moments of the theoretically optimal MaxSR weights over a time span of 504 periods. The underlying return process follows a conditional multivariate normal distribution with constant mean and a HAR-DRD covariance matrix. The parameters of the data generating process are calibrated on daily return data of S&P100 constituents from Jan 2014 to Dec 2021. S stands for sample skewness and K for sample kurtosis. Columns correspond to portfolio size N

N	10	20	30	40	50	60	70	80	90
min S	−11.85	−13.40	−22.37	−15.28	−17.63	−22.38	−19.05	−8.48	−8.90
mean S	1.47	2.89	7.44	1.54	4.73	6.02	1.36	1.27	1.29
max S	13.07	16.11	22.38	16.77	19.66	22.38	14.63	19.75	20.32
min K	109.90	42.75	473.59	36.21	80.82	349.50	8.45	5.60	6.56
mean K	174.95	138.60	496.51	140.81	227.34	496.60	54.15	36.73	36.04
max K	288.15	313.68	501.77	331.63	418.23	501.99	399.75	430.27	445.17

Figure 3 depicts the distribution of the optimal SR of the MaxSR portfolio over time for different portfolio dimensions (in blue) together with the distribution of the Sharpe ratio of the $1/N$-portfolio (in red). By construction for the MaxSR strategy the Sharpe ratio increases with portfolio dimension. The slight increase of the Sharpe ratio for the equally weighted portfolio that can be observed in Fig. 3 can be explained by a market effect: the larger the number of stocks N, the closer the equally weighted portfolio describes the market portfolio. The extreme values for the portfolio weights and the gross exposure are reflected by the large positive outliers for the Sharpe ratios for the MaxSR strategy. Interestingly, smaller negative

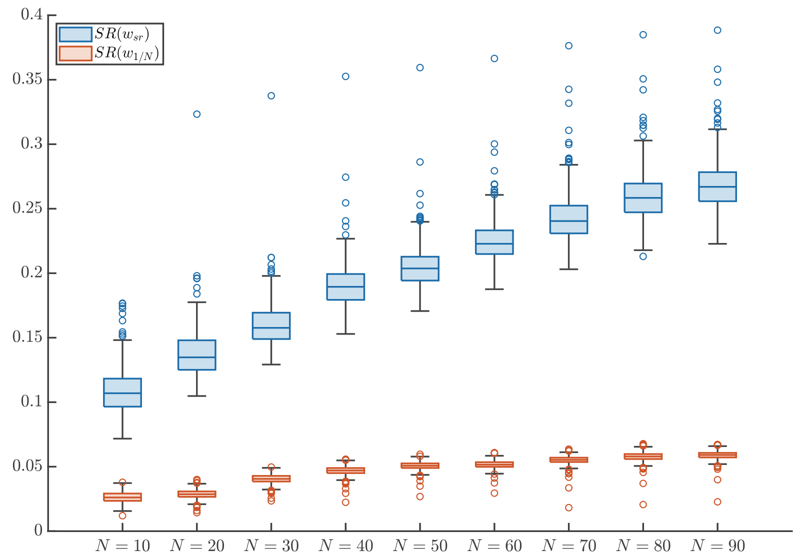

Fig. 3 Sharpe ratio of $w_{sr}(t)$ and $w_{1/N}$ portfolio from a HAR process. Boxplots of the Sharpe ratios evaluated at the true parameters (μ, Σ_t) over a time span of 504 periods. The underlying return process follows a conditional multivariate normal distribution with constant mean and a HAR-DRD covariance matrix. The parameters of the data generating process are calibrated on daily return data of S&P100 constituents from Jan 2014 to Dec 2021. Each boxplot corresponds to the distribution of the Sharpe ratio for the MaxSR portfolio (in blue) and for the equally weighted portfolio (in red) given different portfolio dimensions N (x-axis)

outliers can be observed for the equally weighted portfolios. Finally, Fig. 10 in Appendix "Simulation Design" provides some insight into the issue of a negative portfolio constant a discussed by Maller and Turkington (2003). Figure 10 plots the distribution of the portfolio constant $a_t = \iota' \Sigma_t^{-1} \mu$ over a time span of 504 periods t for different portfolio sizes N. For our calibrated HAR-DRD process the occurrence of a negative portfolio constant is not very likely, but it can occur for any portfolio dimension.

The occurrence of extreme outliers can be better understood through a covariance matrix decomposition, which disentangles the effects of variances and covariances between the stocks in a portfolio. Santos (2019) studies the relative contributions of the diagonal and the off-diagonal elements of the variance-covariance matrix of returns to detect their potential influence on the estimation error. For this the covariance Σ_t needs to be decomposed into the diagonal matrix of variances $\Sigma_{d,t}$ and the matrix of off-diagonal elements $\Sigma_{o,t}$:

$$\Sigma_t = \Sigma_{d,t} + \Sigma_{o,t} \tag{2}$$

Accordingly the precision matrix takes the following form (Henderson & Searle, 1981):

$$\Sigma_t^{-1} = \left(\Sigma_{d,t} + \Sigma_{o,t}\right)^{-1} = \tilde{\Sigma}_{d,t} + \tilde{\Sigma}_{o,t} \, , \tag{3}$$

where

$$\tilde{\Sigma}_{d,t} = \mathrm{diag}\big(\Sigma_t\big)^{-1} \tag{4}$$

$$\tilde{\Sigma}_{o,t} = -\,\Sigma_{d,t}^{-1}\Sigma_{o,t}\big(I + \Sigma_{d,t}^{-1}\Sigma_{o,t}\big)^{-1}\Sigma_{d,t}^{-1}. \tag{5}$$

Note that our decomposition is in terms of the population values, while Santos (2019) uses the empirical counterparts. We use the population variance-covariance matrix to emphasize that the decomposition is a numerical issue, related to the functional form of the theoretical MaxSR weights.

Using the decomposition (3) the MaxSR weight vector (1) takes the form

$$w_{sr}(t) = w_{sr,d}(t) + w_{sr,o}(t) \tag{6}$$

where $w_{sr,d}(t) = \frac{1}{a}\tilde{\Sigma}_{d,t}^{-1}\mu$ and $w_{sr,o}(t) = \frac{1}{a}\tilde{\Sigma}_{o,t}^{-1}\mu$. Alternatively, $w_{sr}(t)$ can also be represented in terms of scaled versions $w_{sr,d}^{*}(t)$ and $w_{sr,o}^{*}(t)$, which both satisfy the adding-up constraint:

$$w_{sr}(t) = \lambda_{d,t}w_{sr,d}^{*}(t) + (1-\lambda_{d,t})w_{sr,o}^{*}(t) \tag{7}$$

where $w_{sr,d}^{*}(t) = \frac{\tilde{\Sigma}_{d,t}^{-1}\mu}{a_{d,t}}$, $w_{sr,o}^{*}(t) = \frac{\tilde{\Sigma}_{o,t}^{-1}\mu}{a_{o,t}}$. The weight $\lambda_{d,t}$ reflects the relative contribution of the diagonal elements to the portfolio constant $a_t = a_{d,t} + a_{o,t}$: $\lambda_{d,t} = \frac{a_{d,t}}{a_t}$, $a_{d,t} = \iota'\tilde{\Sigma}_{d,t}^{-1}\mu$ and $a_{o,t} = \iota'\tilde{\Sigma}_{o,t}^{-1}\mu$. Therefore, the MaxSR portfolio can also be interpreted in terms of a two-fund portfolio. Note that the weights $w_{sr,d}^{*}$ reflect the contribution of the diagonal elements of the variance-covariance matrix. By construction $w_{sr,d}^{*}(t)$ is always positive and is equal to the rewarding-to-risk timing portfolio weighted proposed (Kirby & Ostdiek, 2012), weight $w_{sr,d,i}^{*}(t) = \frac{\mu_i/\sigma_{i,t}}{\sum_i^N(\mu_i/\sigma_{i,t}^2)}$.

Figure 4 depicts the gross exposure of the decomposed weights, $w_{sr,d}^{*}(t)$ and $w_{sr,o}^{*}(t)$, over the time span of 504 periods for the same DGP as in Fig. 2. The left panel corresponds to the ℓ_1-norm of the weights which are driven by the variances and the right panel corresponds to the ℓ_1-norm of the weights driven by the covariances. The left panel of Fig. 4 demonstrates that the gross exposure of the variance-driven weights decreases with N, as one may naturally expect through the numerical decrease in size of the individual weights. Comparing the limits of the y-axes clarifies the source of the instability of the MaxSR portfolio. The extreme outliers arise through the off-diagonal elements of the covariance matrix irrespectively of the portfolio dimension N. Therefore, a natural empirical strategy aimed at stabilizing the MaxSR solution would be to shrink the off-diagonal elements of Σ. Moreover, shrinking a covariance matrix towards its diagonal elements simply implies that $w_{sr,d}^{*}(t)$ receive more weight at the expense of $w_{sr,o}^{*}(t)$, which Santos (2019) interprets as the contribution of the zero-investment residual portfolio.

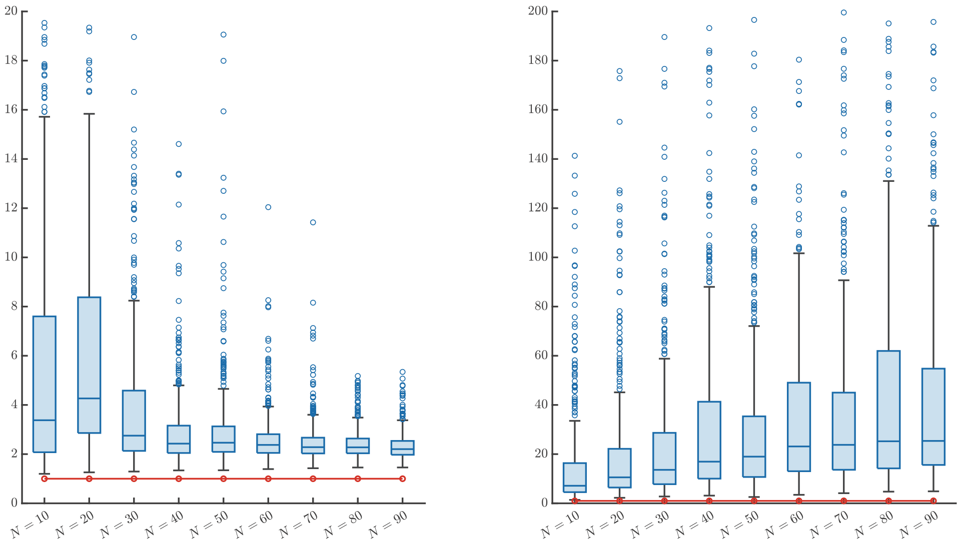

Fig. 4 Gross exposure of theoretical $w^*_{sr,d}(t)$ and $w^*_{sr,o}(t)$ weights from a HAR process. Boxplots of the gross exposure of $w^*_{sr,d}(t)$ (left panel) and $w^*_{sr,o}(t)$ (right panel) weights over a time span of 504 periods. The underlying return process follows a conditional multivariate normal distribution with constant mean and a HAR-DRD covariance matrix. The parameters of the data generating process are calibrated on daily return data of S&P100 constituents from Jan 2014 to Dec 2021. Each boxplot corresponds to the distribution of the gross exposure of different portfolio sizes N (x-axis).The red line plots the level of one, which is equivalent to a portfolio with no short positions. The limits of the y-axes are set differently: [0, 20] for $w^*_{sr,d}(t)$ and [0, 200] for $w^*_{sr,o}(t)$

2.2 *Statistical Issues*

As the optimal portfolio solution (1) depends on the unknown moments of the return distribution, they can be consistently estimated by replacing the population parameters by their empirical counterparts from a sample of returns. The plug-in estimated counterpart of (1) is given by

$$\hat{w}_{sr} = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{\iota'\hat{\Sigma}^{-1}\hat{\mu}}, \tag{8}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ denote the sample mean and the sample covariance estimates, respectively.

In the tradition of the James-Stein estimator various shrinkage approaches for the mean vector have been proposed to mitigate the impact of estimation noise on the empirical portfolio weights, e.g., Jorion (1986), Kan and Zhou (2007) and Liu and Pohlmeier (2013) and many others. The basic idea here is to find an optimal tuning parameter that yields the best trade-off in terms of bias and precision of a linear combination between the noisy empirical mean $\hat{\mu}$ and some target vector μ_0. However, the theoretically optimal shrinkage parameter has to be estimated from noisy data and is therefore a noisy estimate itself. Moreover, the estimated optimal tuning parameter is typically optimal in terms of a statistical risk function but not in terms of the out-of-sample portfolio performance measure. Shrinkage strategies for the variance covariance matrix have been proposed by Frost and Savarino (1986), Ledoit and Wolf (2003, 2004, 2020) and Kourtis et al. (2012) among others.

In the following we propose an approach to estimate the weights of the MaxSR portfolio where both, the mean vector and the covariance matrix, are shrunken towards given target values. Let the two shrinkage estimators be given by

$$\begin{aligned}\tilde{\mu}(\lambda_\mu) =&(1-\lambda_\mu)\hat{\mu} + \lambda_\mu \iota \\ \tilde{\Sigma}(\lambda_\Sigma) =&(1-\lambda_\Sigma)\hat{\Sigma} + \lambda_\Sigma I\,,\end{aligned}$$

with $0 \leq \lambda_\mu, \lambda_\Sigma \leq 1$. Without loss of generality, we assume that the shrinkage target for the mean and the covariance matrix are the all-one vector ι and the identity matrix I, respectively. A doubly shrunken MaxSR solution is given by

$$\tilde{w}_{sr}(\lambda_\mu, \lambda_\Sigma) = \frac{\tilde{\Sigma}(\lambda_\Sigma)^{-1}\tilde{\mu}(\lambda_\mu)}{\iota'\tilde{\Sigma}(\lambda_\Sigma)^{-1}\tilde{\mu}(\lambda_\mu)}. \tag{9}$$

It is helpful to express the doubly shrunken MaxSR portfolio weight (9) as a linear combination of MaxSR portfolio weight and the global minimum variance portfolio (GMVP) weight both computed with the regularized covariance matrix $\tilde{\Sigma}(\lambda_\Sigma)$:

$$\tilde{w}_{sr}(\lambda_\mu, \lambda_\Sigma) = (1-\lambda_\mu)\left(\frac{\tilde{a}(0, \lambda_\Sigma)}{\tilde{a}(\lambda_\mu, \lambda_\Sigma)}\right)\tilde{w}_{sr}(\lambda_\Sigma) + \lambda_\mu\left(\frac{\tilde{c}(\lambda_\Sigma)}{\tilde{a}(\lambda_\mu, \lambda_\Sigma)}\right) w_g(\lambda_\Sigma) \tag{10}$$

where $\tilde{a}(\lambda_\mu, \lambda_\Sigma) = \iota'\tilde{\Sigma}(\lambda_\Sigma)^{-1}\tilde{\mu}(\lambda_\mu)$ is the doubly regularized portfolio constant a and $\tilde{c}(\lambda_\Sigma) = \iota'\tilde{\Sigma}(\lambda_\Sigma)^{-1}\iota$ the regularized counterpart of portfolio constant $c = \iota'\Sigma^{-1}\iota$.

The $\tilde{w}_{sr}(\lambda_\mu, \lambda_\Sigma)$ can be seen as combination approach of various popular portfolio strategies, which are limiting cases of (10), when the two tuning parameters take on the extreme values 0 and 1.

Figure 5 depicts possible portfolio combination strategies that can be generated by an appropriate selection of the two shrinkage parameters. The $1/N$-portfolio (1,1) is the extreme shrinkage case of the plug-in unshrunken MaxSR strategy (0,0). Combinations of the GMVP and the (shrunken) MaxSR portfolio are given by the points on the λ_μ axis, while the line ((1,0), (1,1)) represents GMVPs with different degrees of shrinkage, with the equally weighted $1/N$-portfolio as the limiting case (1,1).

However, for a given data frequency, sample size, and out-of-sample evaluation horizon, it is rather difficult to come up with precise estimates of theoretically optimal shrinkage parameters $(\lambda_\mu, \lambda_\Sigma)$. Moreover, the theoretical justification is given either for the mean shrinkage or the covariance shrinkage, but to the best of our knowledge never both. To illustrate the idea, consider a simulation study where the returns are drawn from a multivariate normal distribution, where the constant mean vector and the covariance matrix are calibrated on the daily data introduced

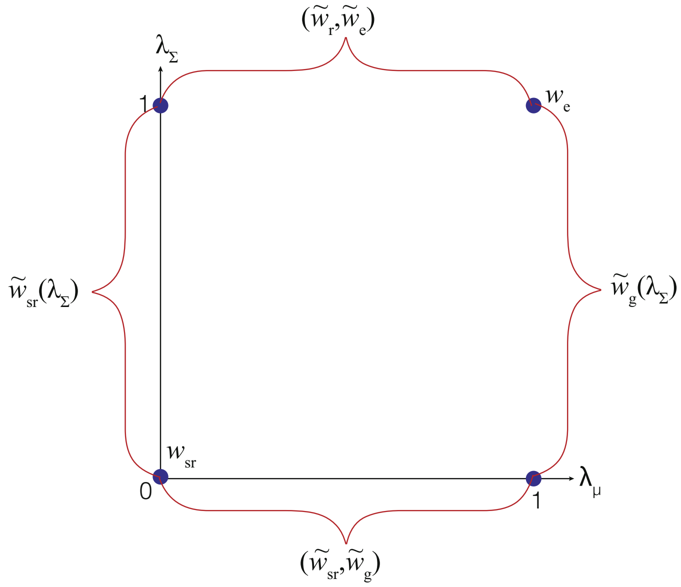

Fig. 5 Combinations of doubly shrunken portfolio weights. Loci of the doubly shrunken MaxSR portfolio weights. The point $(\lambda_\mu = 0, \lambda_\Sigma = 0)$ indicates the plug-in estimated MaxSR portfolio weight, while (1, 1) indicates the equally weighted portfolio $1/N$. Points on the λ_μ-axis represent portfolio strategies which are linear combinations of the (shrunken) MaxSR portfolio and the (shrunken) GMVP portfolio. Points on the line ((0,1),(1,1)) represent combinations of equally weighted, $1/N$-portfolio and the relative mean portfolio $\tilde{w}_r$ with weight $\tilde{w}_{r,i} = \mu_i/\bar{\mu}$ for asset i

in Sect. 4 below. The sample mean and covariance matrix estimates are based on $T = 252$ observations and distributional plots below are based on 100 random draws.

The x-axis on Fig. 6 corresponds to different shrinkage parameter combinations for the doubly shrunken MaxSR portfolio weights given in Eq. (9). Boxplots correspond to the distribution of the out-of-sample Sharpe ratio over 100 simulations for a given $(\lambda_\mu, \lambda_\Sigma)$ for a portfolio of $N = 30$ assets on the left and a portfolio of the $N = 90$ assets on the right. The first parameter combination corresponds to the original MaxSR portfolio defined in Eq. (8), which for $N = 30$ generates the highest SR over all 100 random draws. However, the very same parameter combination for a larger portfolio results in a negative SR. The same holds in an opposite direction: shrinking the covariance matrix with $\lambda_\Sigma = 0.007$ results in the robustly high SR for portfolios formed of 90 assets and a wide range of Sharpe ratios, including negative ones for smaller portfolios with $N = 30$. Needless to mention, that for a more realistic DGP, with, e.g., time-varying conditional moments of the underlying return process, the complexity of the shrinkage parameter choice increases. This very challenging problem can be seen as a model selection problem, where the investor faces different models of $\tilde{w}_{sr}(\lambda_\mu, \lambda_\Sigma)$ and has to choose among them.

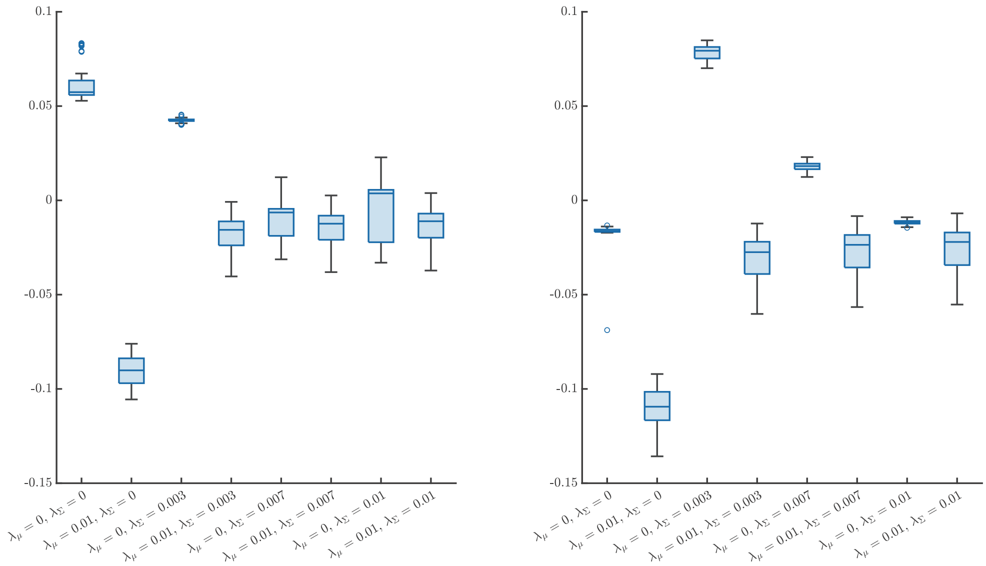

Fig. 6 Sharpe ratio of $\tilde{w}_{sr}(\lambda_\mu, \lambda_\Sigma)$ portfolio from a multivariate normal process. Boxplots of the out of sample SR for $N = 30$ (left panel) and $N = 90$ (right panel) over a time span of 504 periods. The underlying return process follows a multivariate normal distribution with constant mean and covariance matrix. The parameters of the data generating process are calibrated on daily return data of S&P100 constituents from Jan 2014 to Dec 2021. Each boxplot corresponds to the distribution of the SR for the doubly shrunken MaxSR portfolio from (9)

3 Tuning by Bagged Pretesting

In this section we propose a data-driven procedure to robustify empirical MaxSR portfolios. The so-called BPPS algorithm is based on the pretest idea to select the shrinkage parameters for the mean vector and the covariance matrix in an optimal forward-looking way, taking into account actual market conditions. The method proposed by Kazak and Pohlmeier (2023) has been successfully applied to select the optimal combination of strategies with respect to the certainty equivalent. In what follows, we study combinations of shrinkage MaxSR strategies to maximize the out-of-sample Sharpe ratio. We show that the BPPS can be extended to the case of optimally choosing two shrinkage parameters.

In the following we consider a typical rolling window setup, where for the period $t + 1$ the out-of-sample portfolio return $\hat{r}^p_{t+1}(s)$ is based on a one-step ahead forecast of the portfolio weights $\hat{w}_{t+1|t}(s)$ with period $\{t - T, \dots, t\}$ as the estimation window. Here s denotes a *strategy*, which could be a combination of different 2-tupels of the shrinkage parameters $(\lambda_\mu, \lambda_\Sigma)$. We adopt the standard assumption for static portfolio models that the last available estimate $\hat{w}_t(s)$ is used to compute the out-of-sample return for the next period: $\hat{r}^p_{t+1}(s) = \hat{w}_{t+1|t}(s)' r_{t+1} = \hat{w}_t(s)' r_{t+1}$. The estimation window is shifted one period ahead H times resulting in an $H \times 1$ vector of the out-of-sample portfolio returns $\{\hat{r}^p_{t+1}(s), \dots, \hat{r}^p_{t+H}(s)\}$.

Different portfolio strategies are then evaluated based on the out-of-sample Sharpe ratio $\widehat{SR}(\hat{w}(s))$ given by

$$\widehat{SR}(\hat{w}(s)) = \frac{\hat{\mu}_{os}(s)}{\hat{\sigma}^2_{os}(s)}, \tag{11}$$

$$\text{where} \qquad \hat{\mu}_{os}(s) = \frac{1}{H}\sum_{h=1}^{H} \hat{r}^p_{t+h}(s) = \frac{1}{H}\sum_{h=1}^{H} \hat{w}_{t+h-1}(s)' r_{t+h},$$

$$\hat{\sigma}^2_{os}(s) = \frac{1}{H-1}\sum_{h=1}^{H} \left(\hat{r}^p_{t+h}(s) - \hat{\mu}_{os}(s)\right)^2.$$

For the sake of notational simplicity assume that the investor has to decide between two alternative strategies s and $\tilde{s}$ and maximize the Sharpe ratio. The difference in the out-of-sample Sharpe ratios between the two strategies $\Delta_{os}(s, \tilde{s})$ is defined as

$$\Delta_{os}(s, \tilde{s}) = SR(\hat{w}_t(s)) - SR(\hat{w}_t(\tilde{s})).$$

The goal is to select either strategy s or strategy $\tilde{s}$ depending on the test outcome, which tests whether the difference $\Delta_{os}(s, \tilde{s})$ is significantly larger than zero. Null and alternative hypotheses are then given by

$$H_0: \quad \Delta_{os}(s, \tilde{s}) \leq 0 \qquad \text{and} \qquad H_1: \quad \Delta_{os}(s, \tilde{s}) > 0, \tag{12}$$

and thus, the pretest estimator of the portfolio weight forecasts for $t+1$ is such that it depends either on strategy s in case the null is rejected or on $\tilde{s}$ otherwise:

$$w_t(s, \tilde{s}, \alpha) = \mathbb{I}\{\hat{\Delta}_{os}(s, \tilde{s}) > \Delta^*(\alpha)\}\left(w_t(s) - w_t(\tilde{s})\right) + w_t(\tilde{s}), \tag{13}$$

with the estimated SR difference $\hat{\Delta}_{os}(s, \tilde{s}) = \widehat{SR}(s) - \widehat{SR}(\tilde{s})$ and the critical value $\Delta^*(\alpha)$ for a significance level α. Note, however, that at time t the out-of-sample SR difference for the H periods ahead, $\hat{\Delta}_{os}(s, \tilde{s})$, is unknown. We therefore consider two versions of the BPPS: based on the in-sample and pseudo-out-of-sample SR differences. For the first version, the estimated within-sample SR for the strategy s is given by

$$\widehat{SR}_{in,t}(s) = \frac{\hat{w}_t(s)'\bar{r}_t}{\hat{w}_t(s)'\hat{\Sigma}_t\hat{w}_t(s)} \tag{14}$$

where $\bar{r}_t$ denotes the sample mean and $\hat{\Sigma}_t$ the sample covariance matrix of the returns based on the estimation window $\{t-T, \ldots, t\}$. The in-sample test statistic comparing the performance of s and the benchmark $\tilde{s}$ is defined as

$$t_{in,t}(s, \tilde{s}) = \frac{\widehat{SR}_{in,t}(s) - \widehat{SR}_{in,t}(\tilde{s})}{\widehat{S.E.}\left[\widehat{SR}_{in,t}(s) - \widehat{SR}_{in,t}(\tilde{s})\right]}, \tag{15}$$

where the standard error is computed using the delta method following Ledoit and Wolf (2008) (see Appendix "SR testing" for details). Note that the strategy selection based on the pretest estimator considers not only the relative performance of the strategies in terms of the SR difference but also the precision of the estimated performance difference. This approach accounts for the accuracy with which the performance difference has been estimated, providing a more comprehensive assessment of strategy selection, thus enhancing the robustness of the strategy selection process.

Naturally, the in-sample Sharpe ratio difference does not capture the forecasting risk and the amount of portfolio turnover. An empirically more relevant estimator should be based on a forward-looking measure which we artificially construct by splitting the in-sample window $\{t, \ldots, T - t\}$ into two parts, where the first $T - \kappa$ observations are used as a "training set" to estimate MaxSR weights with all possible shrinkage parameter combinations, and the last κ observations are left for the "validation" set to compute the pseudo out-of-sample portfolio returns. On the validation set we then can calculate the transaction costs:

$$TC_t(s) = c \cdot \sum_{j=1}^{N} |\hat{w}_{j,t+1}(s) - \hat{w}_{j,t^+}(s)|, \tag{16}$$

where $TC_t(s)$ denotes transaction costs for strategy s at period t, $\hat{w}_{j,t^+}$ the portfolio weight before rebalancing at $t + 1$, and c the cost per transaction (50 basis points, DeMiguel et al. (2009)). The pseudo out-of-sample SR is then computed based on the net portfolio returns $\hat{r}_t^{p,net}(s) = \hat{r}_t^p(s) - TC_t(s)$ over the pseudo-out-of-sample period of length κ. Similar to (14) and (15) the pretest estimator can be defined in terms of the pseudo out-of-sample SR difference test. The strategy selection of the pretest estimator trained on the in-sample SR difference will be very different from the one based on the pseudo out-of-sample difference, mainly due to the fact that the portfolio turnover often offsets the theoretical gains of different estimated strategies. This is particularly relevant for the MaxSR portfolio with the unstable finite sample performance of the estimated weights.

Note that, in general, the pretest rule defined in (13) depends on the significance level α: the lower the nominal level, the stricter is the pretest rule and the larger should be the difference $\hat{\Delta}_{os}(s, \tilde{s})$ for selecting the strategy s over $\tilde{s}$. In other words, α can be seen as a tuning parameter which determines how risky can we be in terms of selecting an, e.g., MaxSR portfolio over the conservative benchmark of the equally weighted portfolio. For the portfolio optimization problem, this tuning parameter α must be adaptive to the market conditions and must be data-driven. If the goal is to select a shrinkage parameter combination which produces a higher out-of-sample Sharpe ratio, then we no longer enter the classical statistical problem of controlling the size of the test and aiming at a high power. Therefore, for the pretest estimator we use a data-driven rule to select the significance level.

At each time point t for a grid of significance levels α_j, we can determine the pretest choice (s or $\tilde{s}$) together with the corresponding Sharpe ratio, $\widehat{SR}_t(s, \tilde{s}, \alpha_j)$.

A data-driven significance level is determined as the smallest on the grid with the highest SR:

$$\alpha^*_{t+1} = \arg\max_{\alpha} \widehat{SR}_t(s, \tilde{s}, \alpha_j). \tag{17}$$

Note that empirically the choice of α^*_{t+1} is data-driven and depends on the unstable estimates of the portfolio weights. Thus, in the second step we suggest smoothing α^*_{t+1}-series adaptively according to

$$\alpha^s_{t+1} = (1 - \delta)\alpha^*_{t+1} + \delta\alpha^s_t, \tag{18}$$

where the tuning parameter $\delta \in (0, 1)$ is selected to control the degree of smoothness. The resulting tuning parameter α^s_{t+1} is therefore determined via a simple grid search, as both α and δ grids are limited. Through δ the estimator controls how much influence should the past information have to determine the current significance level with geometrically decaying weights. Therefore, the propensity of the pretest estimator to deviate from the benchmark is time-adaptive and data-driven.

As a final step for the optimal shrinkage parameter selection we implement bootstrap aggregation of the pretest decision. For the daily return data introduced in Sect. 4 below we implement a circular block bootstrap by Politis and Romano (1992) to obtain B bootstrap samples of the in-sample estimation window. By taking the average test decision over many bootstrap samples, we construct a combination of the considered strategies. The bagged pretested portfolio selection for the Sharpe ratio maximization is given by the following procedure:

1. For period t define the estimation window of length T: $\{r_{t-T}, \ldots, r_t\}$.
2. For every shrinkage parameter combination compute the Sharpe ratio, either in-sample or pseudo out-of-sample one.
3. For each pair of strategies $(s_i, \tilde{s})$, $i = 1, \ldots, M$, compute the test statistic for testing the difference in the Sharpe ratios against zero.
4. For a grid of α-values of length J compute the pretest estimates and select α^*_{t+1} which results in the largest out-of-sample SR on the grid.
5. For a grid of δ values compute the smoothed $\alpha^s_{os,t+1}$ using the previous optimal significance level and select the one maximizing the SR of the pretest estimator according to (18).

   For every bootstrap iteration $b = 1, \ldots, B$ repeat steps (6) and (7):
6. Randomly sample the rows of the in-sample $T \times N$ data with circular block bootstrap and repeat step (2) for the test statistic computation *keeping the weight estimates fixed*.

(continued)

7. For the significance level selected in step (4) compute the strategy selection indicators $\mathbb{I}^b(s_i, \alpha^s_{t+1})$ for every strategy i and bootstrap iteration b.
8. The bagged probability of the strategy s_i is then defined as $\hat{p}(s_i, \alpha^s_{t+1}) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}^b(s_i, \alpha^s_{t+1})$.
9. Finally, the bagged pretest weight estimator for the period $t+1$ is defined as an average of the weights estimated on the whole sample weighted by the bootstrap probabilities:

$$\hat{w}_t^B(s_1, \ldots, s_M, \tilde{s}, \alpha^s_{t+1}) = \sum_{i=1}^{M+1} \hat{p}(s_i, \alpha^s_{t+1})\, \hat{w}_t(s_i). \tag{19}$$

Bagging is a powerful tool to improve finite sample performance of unstable test decisions (Bühlmann & Yu, 2002). In the portfolio selection context, it allows to combine different portfolio weight estimators s_i, proportionally to the probability that a given strategy s_i significantly dominates the benchmark. The proposed combination is time-adaptive through the flexible significance level choice; it optimizes for an economic criterion, the Sharpe ratio, and not for a statistical measure as often suggested in the literature. Kazak and Pohlmeier (2023) show that the BPPS assigns larger probabilities to the dominating strategy and reduces the variance of the strategy selection. Note, however, that the strategy selection can be beneficial if there is some shrinkage parameter combination which outperforms the benchmark. For the scenario where the benchmark produces the highest Sharpe ratio, the bagged combination would perform similarly to the benchmark. The BPPS can be seen as a tool which insures against estimation risk and a tool for a data-driven shrinkage parameter choice if the shrinkage estimator is superior to the benchmark.

4 Empirical Evidence

In what follows we evaluate the performance of the BPPS method applied to the MaxSR strategy based on daily financial data. We use the stocks contained in the S&P100 together with the SPY index and use the data from January 2019 until December 2021, where different portfolio strategies are estimated during 2019 and evaluated with a rolling window of $T = 252$ within-sample observations over 2020 and 2021 ($H = 504$ out-of-sample observations). We consider a range of portfolio sizes $N \in \{10, 20, \ldots, 90\}$, where for a given portfolio dimension, we randomly draw a subset of N assets from the given pool of 100 assets and report the empirical performance over 500 randomly drawn portfolios. This ensures that the reported results are robust to the asset selection and allows to compare portfolio performance across different portfolio sizes N.

For the BPPS we choose the equally weighted portfolio (i.e., the MaxSR strategy with extreme shrinkage $\lambda_\mu = \lambda_\Sigma = 1$), as the benchmark and consider a range of shrinkage parameters for the MaxSR portfolio as defined in (9), with $\lambda_\Sigma \in \{0, 0.003, 0.007, 0.01\}$, $\lambda_\mu \in \{0, 0.003, 0.007, 0.01\}$. The size of the bagged weights therefore indicate to what extent is it possible to use a shrunken MaxSR strategy to improve portfolio allocation compared to the equally weighted portfolio. All of the 16 combinations of the shrinkage parameters are used for pretesting. The combination of parameters $\lambda_\mu = \lambda_\Sigma = 0$ corresponds to the original MaxSR portfolio defined in (8). Values of shrinkage parameters are driven by the scale of the daily returns: the average return in the data is $0.10826 \cdot 10^{-3}$ and the average covariance equals $0.06056 \cdot 10^{-3}$. Compared to such scaling, the shrinkage targets of an identity matrix for Σ and a vector of ones for μ are very large, and thus, a relatively small shrinkage parameter value would imply a very strong regularization.

To improve readability, we report the performance of 8 most interesting combinations only. For the bagging step we set the number of bootstrap replications to $B = 200$ and we use the block bootstrap, where the block length corresponds to 66 days (3 trading months). We consider two versions of the BPPS estimator: (i) *BPPS gross* is based on the Sharpe ratio testing of gross portfolio returns within sample and (ii) *BPPS net* is based on the SR of the portfolio returns net of the transaction costs. For the computation of the pseudo out-of-sample test statistic, we split the sample into two equal parts, i.e., $\kappa = T/2$. For both pretest estimators we allow for a data-driven choice of the significance level, where the selection is performed to maximize within-sample SR or pseudo-out-of-sample SR correspondingly.

Figure 7 plots the time evolution of the data-driven significance level for the *BPPS net* across portfolio sizes N. A larger α_t^* (higher significance level) corresponds to a softer pretesting rule, where a SR of a given shrinkage parameter couple $(\lambda_\Sigma, \lambda_\mu)$ has to be only slightly larger than the SR of the equally weighted portfolio in order for the MaxSR to be chosen by the pretest. Notably, the pretest estimator "tightens" during the COVID-19 pandemic. The y-axis on the figure is equivalent to the critical level range of approximately $(2.33; 0.67)$ from bottom to the top. The test statistic, or the standardized difference in the Sharpe ratios between a given MaxSR strategy and the equally weighted portfolio, had to be larger during March to October 2020 compared to the October 2020 to March 2021 period for the MaxSR to be selected.[4] This illustrates that BPPS adapts to the market conditions: during more volatile periods the algorithm tends to select the conservative benchmark, whereas when the market is calm, it allows for the risky strategy selection. This essentially is due to the fact that through the denominator of the test statistic, the pretest estimator takes into account not only the estimated

[4] Figure 11 in Appendix "Further Empirical Evidence" plots the average squared daily return for the out-of-sample period as a proxy for the daily variance. For the considered data the periods of the highest market volatility correspond to the lowest significance levels.

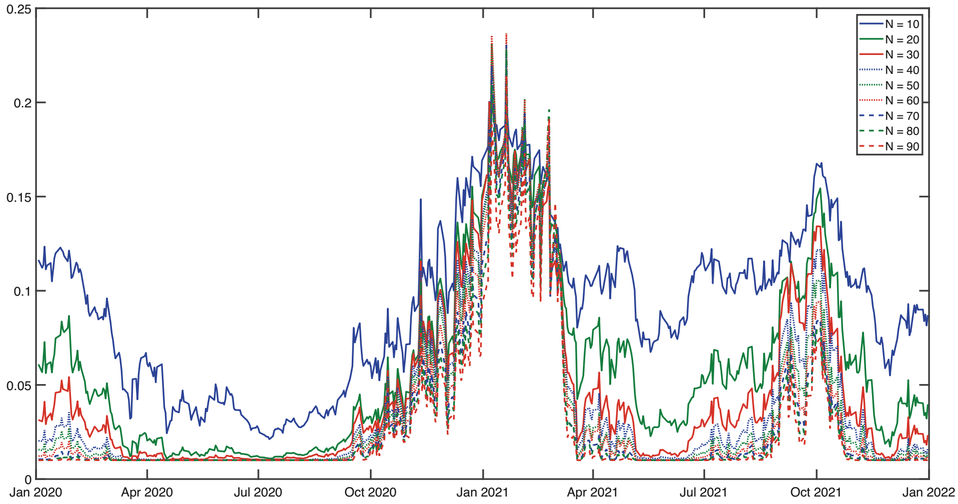

Fig. 7 Time evolution of the data-driven significance level for the BPPS. Plotted lines represent the average significance level, which was selected for the bagged pretest estimator, over 500 randomly drawn portfolios. Different lines correspond to different portfolio sizes N. The x-axis represents the out-of-sample evaluation period. The y-axis corresponds to the range of significance levels: from 0.01 to 0.25 (equivalent to (2.33; 0.67) range of critical values) for the one-sided test defined in (12)

SR difference, but also the precision with which this difference is estimated. During periods of high market volatility the difference in the Sharpe ratios between two competing strategies would have a larger standard error, which forces the pretest estimator to stick to the conservative equally weighted portfolio benchmark.

Another feature of the pretest estimator is its ability to adapt to the given portfolio size N. For larger portfolio dimensions, as N approaches the market size, the covariance between portfolio returns of different strategies increases, as naturally there are fewer opportunities to produce portfolio weights which are far off from $1/N$. The covariance between the portfolio returns enters the denominator of the test statistic: the larger the covariance, the larger is the denominator of the test statistic. Thus, for larger portfolio dimensions the BPPS is more conservative in the sense of selecting the risky MaxSR portfolio if the test statistic is large.

We first consider the quality of the strategy selection based on the gross portfolio returns. Table 2 reports annualized Sharpe ratios, averaged over 500 randomly drawn portfolios of size N in different rows. The first 8 columns correspond to shrinkage parameter combinations of the MaxSR strategy $(\lambda_\mu, \lambda_\Sigma)$, and the ninth column corresponds to the equally weighted portfolio benchmark. The last two columns report the annualized out-of-sample Sharpe ratio of the BPPS strategy, trained on the in-sample SR difference (*gross*) and on the pseudo out-of-sample one taking into account portfolio rebalancing (*net*). Numbers in red bold denote the largest SR for a given N, and numbers in black bold correspond to the second largest SR. Note that, among the shrinkage parameter combinations, there is no unique pair $(\lambda_\mu, \lambda_\Sigma)$ which would demonstrate superior performance across all portfolio dimensions. The *BPPS gross* estimator on the other hand is able to optimally select

Table 2 Out-of-sample SR for gross returns. Numbers in the table correspond to the annualized average out-of-sample SR computed on gross portfolio returns over 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 504$ observations, in-sample estimation window length $T = 252$. The first 8 columns correspond to shrunken MaxSR weights with covariance shrinkage λ_Σ and mean shrinkage λ_μ. Column 9 reports the SR for the benchmark equally weighted portfolio. The last two columns correspond to the BPPS trained on in-sample SR differences (*gross*) and pseudo out-of-sample differences net of the transaction costs (*net*). Numbers in bold correspond to the second largest SR for a given portfolio size N. Numbers in bold italic correspond to the largest SR for a given portfolio size N

	$\lambda_\Sigma = 0$		$\lambda_\Sigma = 0.003$		$\lambda_\Sigma = 0.007$		$\lambda_\Sigma = 0.01$			BPPS	BPPS
N	$\hat{w}_{sr}$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$w_{1/N}$	gross	net
10	**0.04**	−0.09	−0.02	−0.03	−0.04	−0.03	−0.04	−0.03	−0.02	***0.04***	−0.01
20	−0.07	−0.18	−0.04	−0.03	**0.01**	−0.03	0.00	−0.03	−0.01	***0.23***	−0.03
30	−0.05	−0.20	−0.01	−0.04	**0.02**	−0.03	−0.03	−0.02	−0.01	***0.32***	−0.03
40	−0.09	−0.23	−0.01	−0.03	**0.03**	−0.02	0.03	−0.02	0.00	***0.39***	−0.03
50	−0.07	−0.31	0.05	−0.05	0.02	−0.04	**0.06**	−0.03	−0.02	***0.46***	−0.04
60	−0.09	−0.33	−0.01	−0.04	**0.02**	−0.03	−0.01	−0.03	−0.01	***0.55***	−0.04
70	−0.08	−0.37	**0.02**	−0.05	−0.02	−0.04	−0.02	−0.04	−0.01	***0.58***	−0.04
80	−0.03	−0.45	−0.07	−0.05	−0.01	−0.04	**0.00**	−0.04	−0.01	***0.71***	−0.05
90	0.02	−0.50	**0.03**	−0.06	−0.04	−0.05	−0.03	−0.04	−0.02	***0.83***	−0.05

different strategies over time, such that the resulting combination produces the largest SR based on the gross returns for all N. Furthermore, the resulting out-of-sample Sharpe ratios of the BPPS increase with N, which should be expected theoretically as the diversification opportunities arise, empirically however, as reported in the Table 2; this is not the case for the other strategies. The last column reports the performance of the *BPPS net* estimator, which is "trained" on a different criterion: SR net of the transaction costs and therefore its performance is closest to the equally weighted portfolio.

Table 2 gives only a limited overview of the performance of different strategies as it reports only averages over 500 randomly drawn portfolios. Figure 8 plots the distribution of the out-of-sample Sharpe ratios for the largest considered portfolio dimension $N = 90$. Each boxplot depicts the distribution of the SR for considered weight estimation strategies. These boxplots provide insights into the robustness of different strategies in terms of the asset selection. For instance, even a slightest shrinkage of the mean with $\lambda_\mu \neq 0$ stabilizes the performance of MaxSR portfolio across portfolios formed of different assets. In other words, the limiting case of the equally weighted portfolio is achieved very quickly as long as some shrinkage of the covariance matrix is present. Note that, in the limiting case of strong covariance shrinkage, the MaxSR portfolio converges to the estimator which assigns the weight proportional to the means of different stocks. It turns out that for larger portfolio

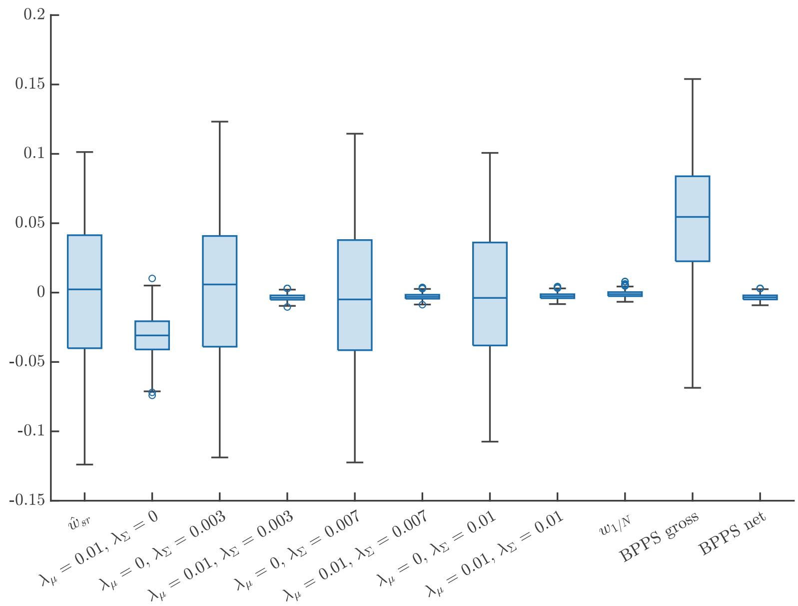

Fig. 8 Distribution of the Sharpe ratio for gross returns, $N = 90$. Boxplots of the out-of-sample SR computed on gross portfolio returns over 500 randomly drawn portfolios of size $N = 90$. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 504$ observations, in-sample estimation window length $T = 252$. X-axes denote different ways of computing portfolio weights: the first 8 boxplots correspond to shrunken MaxSR weights with covariance shrinkage λ_Σ and mean shrinkage λ_μ. Boxplot 9 reports the distribution of the SR for the benchmark equally weighted portfolio $1/N$. The last two boxplots correspond to the BPPS trained on in-sample SR differences (*gross*) and pseudo out-of-sample differences net of the transaction costs (*net*)

dimensions this shrinkage target is not too far from the equally weighted portfolio as the means of returns across stocks are not too different from each other.

The *BPPS gross* dominates the equally weighted portfolio in terms of the average Sharpe ratio; however, its performance varies much more across different stocks compared to the benchmark. For some periods the BPPS assigns a nonzero weight to the strategy with no mean shrinkage, which produces extreme outliers. This instability problem can be reduced by considering the Sharpe ratio calculated on net portfolio returns, which by construction implies a penalty on the portfolio turnover. Table 4 in Appendix "Further Empirical Evidence" reports the annualized average out-of-sample Sharpe ratio calculated on net portfolio returns. Unsurprisingly, all of the considered strategies produce negative annualized Sharpe ratios, as daily rebalancing implies unrealistic turnover.[5] Therefore, we consider a more empirically relevant scenario where portfolio weights are estimated on the daily basis, but rebalancing is performed monthly.[6]

Table 3 reports the annualized out-of-sample SR net of the transaction costs averaged over 500 randomly drawn portfolios with monthly rebalancing. The equally weighted portfolio produces the largest Sharpe ratio for all portfolio dimensions $N > 10$. The portfolio performance tends to deteriorate with the increase in portfolio dimension N, as the amount of portfolio rebalancing outweighs the diversification gains. Note that, despite monthly rebalancing, out-of-sample Sharpe ratios tend to be negative even for the $1/N$. For out-of-sample testing the *BPPS net* is able to mimic the performance of the benchmark. However, it is not able to produce larger out-of-sample Sharpe ratio.

Figure 9 provides more insights into the performance of different shrinkage strategies in terms of net portfolio returns. The performance of the equally weighted portfolio, MaxSR portfolio with double shrinkage of both mean and covariance, and *BBPS net* is very similar across all randomly drawn portfolios. However, in terms of the net portfolio returns the equally weighted portfolio still performs better than the other strategies. In other words, for the BPPS estimator there was no alternative strategy to choose from and it would select the benchmark in most of the cases.

5 Conclusions

The theoretical contributions by Okhrin and Schmid (2006) and Schmid and Zabolotskyy (2008) on the abysmal distributional properties of the maximum Sharpe ratio portfolio weights give rise to enormous challenges for empirical finance. On the one hand, the maximum Sharpe ratio portfolio is a natural choice of the investment strategy to maximize the Sharpe ratio, and on the other hand,

[5] We impose 50 basis points as cost per transaction (DeMiguel et al., 2009).

[6] Appendix "Further Empirical Evidence" reports the results for weekly rebalancing in Table 5 and Fig. 12.

Table 3 Out-of-sample SR for net returns with monthly rebalancing. Numbers in the table correspond to the annualized average out-of-sample SR computed on net portfolio returns with monthly rebalancing over 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 504$ observations, in-sample estimation window length $T = 252$. The first 8 columns correspond to shrunken MaxSR weights with covariance shrinkage λ_Σ and mean shrinkage λ_μ. Column 9 reports the SR for the benchmark equally weighted portfolio. The last two columns correspond to the BPPS trained on in-sample SR differences (*gross*) and pseudo out-of-sample differences net of the transaction costs (*net*) Numbers in bold correspond to the second largest SR for a given portfolio size N. Numbers in bold italic correspond to the largest SR for a given portfolio size N

	$\lambda_\Sigma = 0$		$\lambda_\Sigma = 0.003$		$\lambda_\Sigma = 0.007$		$\lambda_\Sigma = 0.01$			BPPS	BPPS
N	$\hat{w}_{sr}$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$w_{1/N}$	gross	net
10	−0.52	−0.21	−0.36	−0.09	−0.38	−0.09	−0.37	***−0.03***	**−0.03**	−0.17	−0.07
20	−0.75	−0.39	−0.37	−0.10	−0.36	−0.09	−0.32	**−0.03**	***−0.03***	−0.17	−0.04
30	−0.79	−0.55	−0.39	−0.10	−0.39	−0.09	−0.45	**−0.02**	***−0.02***	−0.17	−0.04
40	−0.96	−0.68	−0.40	−0.09	−0.43	−0.08	−0.43	**−0.02**	***−0.02***	−0.27	−0.03
50	−1.00	−0.89	−0.43	−0.11	−0.39	−0.10	−0.39	**−0.04**	***−0.03***	−0.35	−0.04
60	−1.03	−1.05	−0.40	−0.11	−0.36	−0.10	−0.35	**−0.03**	***−0.03***	−0.33	−0.04
70	−1.10	−1.22	−0.45	−0.12	−0.43	−0.11	−0.45	**−0.04**	***−0.03***	−0.37	−0.05
80	−1.10	−1.41	−0.53	−0.12	−0.48	−0.11	−0.44	**−0.04**	***−0.03***	−0.34	−0.04
90	−0.94	−1.60	−0.67	−0.13	−0.64	−0.11	−0.55	−0.04	***−0.04***	−0.22	**−0.04**

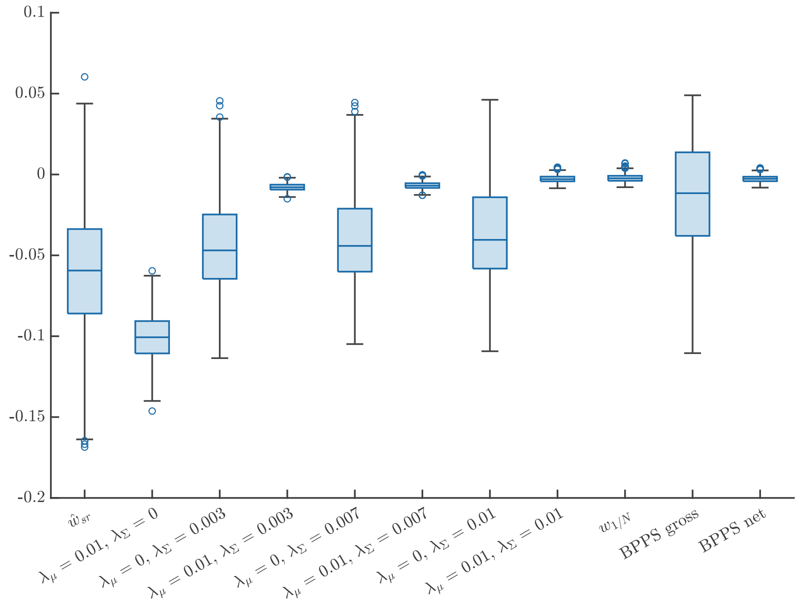

Fig. 9 Distribution of the Sharpe ratio for net returns with monthly rebalancing, $N = 90$. Boxplots of the out-of-sample SR computed on net portfolio returns with monthly rebalancing over 500 randomly drawn portfolios of size $N = 90$. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 504$ observations, in-sample estimation window length $T = 252$. X-axes denote different ways of computing portfolio weights: the first 8 boxplots correspond to shrunken MaxSR weights with covariance shrinkage λ_Σ and mean shrinkage λ_μ. Boxplot 9 reports the distribution of SR for the benchmark equally weighted portfolio $1/N$. The last two boxplots correspond to the BPPS trained on in-sample SR differences (*gross*) and pseudo out-of-sample differences net of the transaction costs (*net*)

the finite sample properties of the estimated weights make this strategy hardly applicable. This study offers an empirical investigation of the challenges to work with the MaxSR portfolio strategy. We investigate further details of the sources of the numerical instability of the MaxSR portfolio. By disentangling the MaxSR weights into different components, we provide numerical insights into the sources of instability of the MaxSR weights and how this instability can be controlled by regularizing the mean and the covariance of the return process. We propose a double regularization strategy for the MaxSR portfolio which can be seen as a combination approach of various popular standalone strategies.

For a realistic data generating process it is rather difficult to come up with an optimal shrinkage rule, and such rule depends on the noisy estimates of the moments of the return process. Therefore, we adapt the BPPS algorithm by Kazak and Pohlmeier (2023) to the shrinkage parameter selection in the context of maximizing out-of-sample Sharpe ratio both gross and net of the transaction costs. In our empirical study we show that a single regularization strategy does not dominate competitors uniformly over time and portfolio dimensions in terms of gross SR. The bagged portfolio selection is capable of combining the shrinkage parameters in an optimal way. This finding demonstrates that if appropriately regularized, the MaxSR can still get a chance to be applied in practice. There is, however, a caveat to these findings. Namely, if one is interested in maximizing the out-of-sample Sharpe ratio which takes into account the amount of portfolio turnover, the equally weighted portfolio cannot be robustly outperformed by a combination of doubly regularized MaxSR portfolios.

Finally, this study shows that the BBPS algorithm may be a promising tool for the dynamic data-driven selection of a set of tuning parameters in multidimensional regularization problems.

Acknowledgments The authors would like to thank Lea Taylor, Roxana Halbleib, and Yarema Okhrin for helpful comments. All remaining errors are ours.

Appendix

SR testing

The significance of the SR difference is tested by the means of the Student t-test. The standard error of the SR difference is computed using the delta method (the bootstrap option is also available but is more computationally costly). The delta method for testing the difference in SR of the two strategies $(s, \tilde{s})$ is given by

$$\Delta_{os} = \phi(\vartheta) = \frac{\mu_{os}(s)}{\sigma^2_{os}(s)} - \frac{\mu_{os}(\tilde{s})}{\sigma^2_{os}(\tilde{s})}$$

$$\sqrt{H}\big(\hat{\Delta}_{os}(s,\tilde{s}) - \Delta_{os}(s,\tilde{s})\big) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial\phi(\vartheta)}{\partial\vartheta'} V[\hat{\vartheta}] \frac{\partial\phi(\vartheta)}{\partial\vartheta}\right),$$

where $\vartheta = \big(\mu_{os}(s), \mu_{os}(\tilde{s}), \sigma^2_{os}(s), \sigma^2_{os}(\tilde{s})\big)'$ is the vector of the mean and variance of portfolio returns with the covariance matrix $V[\hat{\vartheta}]$ and $\hat{\sigma}_{os}(s,\tilde{s})$ denotes the sample covariance between the out-of-sample portfolio returns $\hat{r}_p(s)$ and $\hat{r}_p(\tilde{s})$.

However, as noted in Ledoit and Wolf (2008) working with the uncentered moments might be more convenient. Given the out-of-sample portfolio returns $\hat{r}_p(s), \hat{r}_p(\tilde{s})$ of length H, define $\hat{y} = (\hat{r}_p(s) - \bar{\hat{r}}_p(s), \hat{r}_p(\tilde{s}) - \bar{\hat{r}}_p(\tilde{s}), \hat{r}^2_p(s) - \overline{\hat{r}^2}_p(s), \hat{r}^2_p(\tilde{s}) - \overline{\hat{r}^2}_p(\tilde{s}))$, where $\bar{\hat{r}}(\cdot)$ denotes the sample average. The standard error of the out-of-sample SR difference is then computed as

$$S.E. = \sqrt{\frac{\hat{\nabla}'\hat{\Psi}\hat{\nabla}}{H}} \quad \text{with} \quad \hat{\Psi} = \frac{1}{H}\hat{y}'\hat{y} \quad \text{and}$$

$$\hat{\nabla} = \Bigg[\frac{\overline{\hat{r}^2}_p(s)}{(\overline{\hat{r}^2}_p(s) - \bar{\hat{r}}^2_p(s))^{3/2}}, \frac{\overline{\hat{r}^2}_p(\tilde{s})}{(\overline{\hat{r}^2}_p(\tilde{s}) - \bar{\hat{r}}^2_p(\tilde{s}))^{3/2}},$$

$$-\frac{1}{2}\frac{\bar{\hat{r}}_p(s)}{(\overline{\hat{r}^2}_p(s) - \bar{\hat{r}}^2_p(s))^{3/2}}, -\frac{1}{2}\frac{\bar{\hat{r}}_p(\tilde{s})}{(\overline{\hat{r}^2}_p(\tilde{s}) - \bar{\hat{r}}^2_p(\tilde{s}))^{3/2}}\Bigg]'.$$

Simulation Design

We simulate the following HAR-DRD return process r_t with the covariance matrix Σ_t and dynamic conditional correlations:

$$\Sigma_t = D_t R_t D_t, \quad D_t = \begin{pmatrix} \sigma_{11,t} & . & 0 \\ 0 & \ddots & 0 \\ 0 & . & \sigma_{NN,t} \end{pmatrix}, \quad R_t = \begin{pmatrix} 1 & \rho_{12,t} & \dots & \rho_{1N,t} \\ \rho_{12,t} & 1 & \dots & \rho_{2N,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N,t} & .. & \rho_{(N-1)N,t} & 1 \end{pmatrix}, \tag{20}$$

$$\ln \sigma^2_{jj,t} = \beta_{j,0} + \beta_{j,1} \ln \sigma^2_{jj,t-1} + \beta_{j,2} \frac{1}{5} \sum_{i=1}^{5} \ln \sigma^2_{jj,t-i} + \beta_{j,3} \frac{1}{22} \sum_{i=1}^{22} \ln \sigma^2_{jj,t-i} + \varepsilon_{j,t}, \tag{21}$$

$$vech(R_t) = \alpha_0 + \alpha_1 vech(R_{t-1}) + \alpha_2 \frac{1}{5} \sum_{i=1}^{5} vech(R_{t-i}) + \alpha_3 \frac{1}{22} \sum_{i=1}^{22} vech(R_{t-i}) + \varepsilon_t, \tag{22}$$

$$r_t \sim \mathcal{N}(\mu, \Sigma_t), \quad \varepsilon_{j,t} \sim t(\nu_j), \quad \varepsilon_t \sim t(\nu), \tag{23}$$

where $j = 1, \dots, N$, the $vech(\cdot)$ operator denotes the vector form of the lower triangular part of a matrix, and $t(\nu)$ denotes a Student-t distributed random variable with ν degrees of freedom. The parameters of the data generating process are calibrated on the S&P100 data described in Sect. 4.

Further Empirical Evidence

(Figures 10, 11 and 12 and Tables 4 and 5)

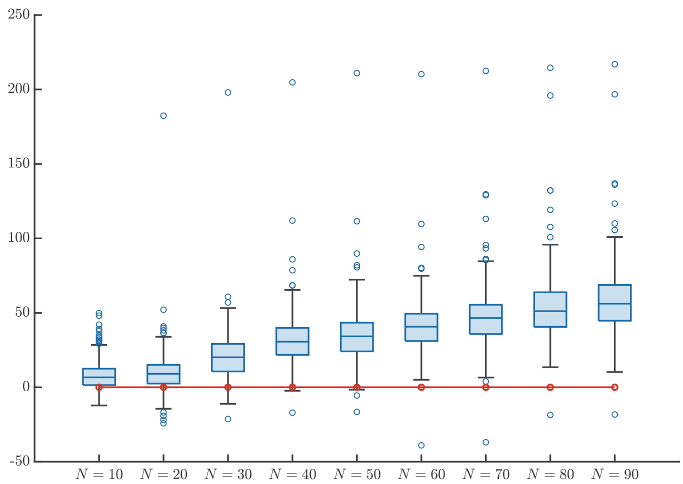

Fig. 10 Portfolio constant *a from a HAR process.* Boxplots of the portfolio constant $a = \iota' \Sigma^{-1} \mu$ over a time span of 504 periods. The returns follow a conditional multivariate normal distribution with constant mean and a HAR-DRD covariance matrix. The parameters of the data generating process are calibrated on daily return data of S&P100 constituents from Jan 2014 to Dec 2021. Each boxplot corresponds to the distribution of the portfolio constant a of different portfolio dimensions N (x-axis).The red line plots the level of zero

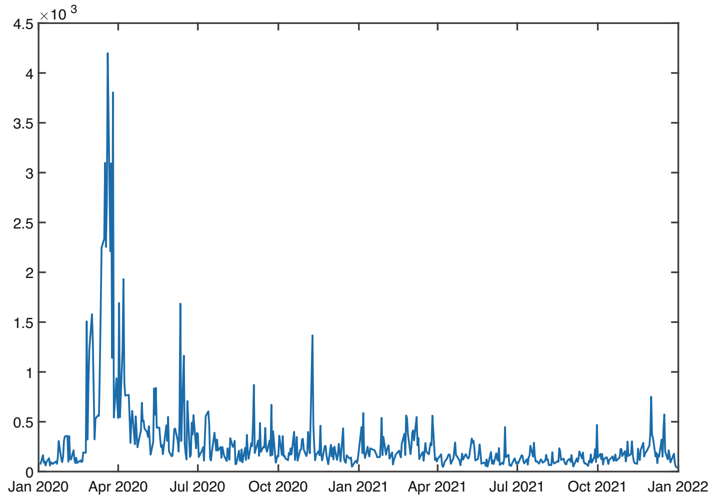

Fig. 11 Average squared return for the out-of-sample period. Plotted line represent the time series of the average squared daily return of the 100 considered stocks of S&P100 with SPY

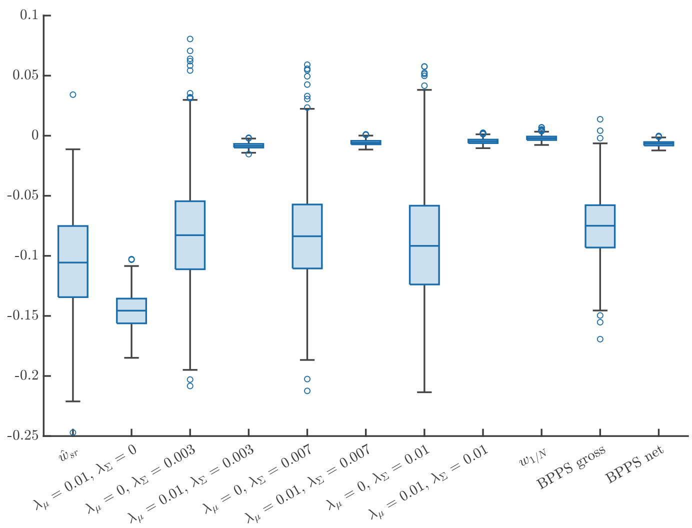

Fig. 12 SR net with weekly rebalancing, $N = 90$. Boxplots of the out-of-sample SR computed on net portfolio returns with weekly rebalancing over 500 randomly drawn portfolios of size $N = 90$. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 504$ observations, in-sample estimation window length $T = 252$. X-axes denote different ways of computing portfolio weights: the first 8 columns correspond to shrunken MaxSR weights with covariance shrinkage λ_Σ and mean shrinkage λ_μ. Column 9 reports the distribution of SR for the benchmark equally weighted portfolio. The last two columns correspond to the BPPS trained on in-sample SR differences (*gross*) and pseudo out-of-sample differences net of the transaction costs (*net*)

Table 4 Out-of-sample SR for net returns with daily rebalancing. Numbers in the table correspond to the annualized average out-of-sample SR computed on net portfolio returns with daily rebalancing over 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 504$ observations, in-sample estimation window length $T = 252$. The first 8 columns correspond to shrunken MaxSR weights with covariance shrinkage λ_Σ and mean shrinkage λ_μ. Column 9 reports the SR for the benchmark equally weighted portfolio. The last two columns correspond to the BPPS trained on in-sample SR differences (*gross*) and pseudo out-of-sample differences net of the transaction costs (*net*) Numbers in bold correspond to the second largest SR for a given portfolio size N. Numbers in bold italic correspond to the largest SR for a given portfolio size N

	$\lambda_\Sigma = 0$		$\lambda_\Sigma = 0.003$		$\lambda_\Sigma = 0.007$		$\lambda_\Sigma = 0.01$			BPPS	BPPS
N	$\hat{w}_{sr}$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$w_{1/N}$	gross	net
10	−1.98	−0.46	−1.81	−0.08	−1.82	**−0.08**	−1.85	***−0.08***	−0.09	−2.65	−0.32
20	−2.16	−0.94	−2.04	−0.10	−2.01	**−0.09**	−1.99	***−0.08***	−0.09	−3.24	−0.13
30	−2.21	−1.35	−2.12	−0.11	−2.08	−0.09	−2.02	***−0.08***	**−0.09**	−3.02	−0.12
40	−2.18	−1.76	−2.14	−0.12	−2.16	−0.09	−2.12	***−0.08***	**−0.08**	−2.76	−0.13
50	−2.15	−2.24	−2.12	−0.15	−2.16	−0.11	−2.15	**−0.10**	***−0.10***	−2.51	−0.15
60	−2.17	−2.67	−2.31	−0.16	−2.24	−0.11	−2.21	**−0.10**	***−0.09***	−2.40	−0.15
70	−2.13	−3.13	−2.22	−0.18	−2.18	−0.12	−2.19	**−0.11**	***−0.10***	−2.31	−0.16
80	−2.20	−3.64	−2.23	−0.19	−2.23	−0.13	−2.27	**−0.11**	***−0.09***	−2.24	−0.17
90	−2.16	−4.14	−2.34	−0.21	−2.34	−0.14	−2.31	**−0.12**	***−0.10***	−2.13	−0.18

Table 5 Out-of-sample SR for net returns with weekly rebalancing. Numbers in the table correspond to the annualized average out-of-sample SR computed on net portfolio returns with weekly rebalancing over 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 504$ observations, in-sample estimation window length $T = 252$. The first 8 columns correspond to shrunken MaxSR weights with covariance shrinkage λ_Σ and mean shrinkage λ_μ. Column 9 reports the SR for the benchmark equally weighted portfolio. The last two columns correspond to the BPPS trained on in-sample SR differences (*gross*) and pseudo out-of-sample differences net of the transaction costs (*net*). Numbers in bold correspond to the second largest SR for a given portfolio size N. Numbers in bold italic correspond to the largest SR for a given portfolio size N

	$\lambda_\Sigma = 0$		$\lambda_\Sigma = 0.003$		$\lambda_\Sigma = 0.007$		$\lambda_\Sigma = 0.01$			BPPS	BPPS
N	$\hat{w}_{sr}$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$\lambda_\mu = 0$	$\lambda_\mu = 0.01$	$w_{1/N}$	gross	net
10	−1.10	−0.29	−0.91	−0.05	−0.95	−0.04	−0.93	**−0.04**	***−0.03***	−0.83	−0.15
20	−1.30	−0.55	−1.08	−0.06	−1.04	−0.05	−1.06	**−0.04**	***−0.03***	−1.05	−0.07
30	−1.46	−0.77	−1.07	−0.07	−1.06	−0.05	−1.10	**−0.04**	***−0.02***	−1.10	−0.07
40	−1.55	−0.98	−1.18	−0.07	−1.20	−0.05	−1.11	**−0.04**	***−0.02***	−1.18	−0.07
50	−1.55	−1.25	−1.19	−0.09	−1.17	−0.07	−1.13	**−0.06**	***−0.03***	−1.15	−0.08
60	−1.59	−1.47	−1.19	−0.10	−1.25	−0.07	−1.27	**−0.06**	***−0.02***	−1.14	−0.08
70	−1.62	−1.72	−1.28	−0.11	−1.32	−0.08	−1.31	**−0.07**	***−0.03***	−1.15	−0.09
80	−1.68	−2.02	−1.31	−0.12	−1.32	−0.08	−1.32	**−0.07**	***−0.03***	−1.24	−0.10
90	−1.67	−2.31	−1.30	−0.13	−1.31	−0.09	−1.43	**−0.07**	***−0.03***	−1.20	−0.11

References

Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies, 4*(2), 315–342.

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

Brodie, J., Daubechies, I., Mol, C. D., Giannone, D., & Loris, I. (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences, 106*(30), 12267–12272. PMID: 19617537.

Bühlmann, P., & Yu, B. (2002). Analyzing bagging. *Annals of Statistics, 30*(4), 927–961.

Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management, 19*(2), 6–11.

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

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

Frost, P. A., & Savarino, J. E. (1986). An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis, 21*(03), 293–305.

Henderson, H. V., & Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review, 23*(1), 53–60.

Jobson, J. D., & Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association, 75*(371), 544–554.

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

Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis, 42*(03), 621–656.

Kazak, E., & Pohlmeier, W. (2019). Testing out-of-sample portfolio performance. *International Journal of Forecasting, 35*(2), 540–554.

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

Kirby, C., & Ostdiek, B. (2012). It's all in the timing: Simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis, 47*(2), 437–467.

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

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

Ledoit, O., & Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management, 30*(4), 110–119.

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

Ledoit, O., & Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Annals of Statistics, 48*(5), 3043–3065.

Liu, H., & Pohlmeier, W. (2013). Risk preferences and estimation risk in portfolio choice. In *Working paper series* (Vol. 47). The Rimini Centre for Economic Analysis.

Maller, R. A., & Turkington, D. A. (2003). New light on the portfolio allocation problem. *Mathematical Methods of Operations Research (ZOR), 56*(3), 501–511.

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

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal, 45*, 31–42.

Okhrin, Y., & Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of Econometrics, 134*(1), 235–256.

Pennacchi, G. (2008). Theory of asset pricing. In *The Addison-Wesley series in finance*. Boston: Pearson Addison-Wesley.

Politis, D. N., & Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In *Exploring the limits of bootstrap* (pp. 263–270)

Santos, A. A. P. (2019). Disentangling the role of variance and covariance information in portfolio selection problems. *Quantitative Finance, 19*(1), 57–76.

Schmid, W., & Zabolotskyy, T. (2008). On the existence of unbiased estimators for the portfolio weights obtained by maximizing the sharpe ratio. *AStA Advances in Statistical Analysis, 92*(1), 29–34.