Optimal Portfolio Allocation with Statistical Learning*

Ekaterina Kazak†

University of Konstanz, GSDS

This version: November 10, 2018

Abstract

This paper exploits the idea of pretesting to choose between competing portfolio strategies. We propose a strategy that optimally trades off between the risk of going for a false positive strategy choice versus the risk of making a false negative choice.

Various different data driven approaches are proposed based on an optimal choice of the pretested certainty equivalent and Sharpe Ratio. Our approach belongs to the class of shrinkage portfolio estimators. However, contrary to previous approaches the shrinkage intensity is continuously updated to incorporate the most recent information in the rolling window forecasting set-up. We show that the bagged pretest estimator performs exceptionally well, especially when combined with adaptive smoothing. The resulting strategy allows for a flexible and smooth switch between the underlying strategies and is shown to outperform the corresponding stand-alone strategies.

Keywords: pretest estimation, bagging, portfolio allocation

*Financial support by Graduate School of Decision Sciences (GSDS) and the German Science Foundation (DFG) is gratefully acknowledged. All remaining errors are mine.

†Department of Economics, Universitätsstraße 10, D-78462 Konstanz, Germany. Phone: +49-7531-88-2219, fax: -4450, email: Ekaterina.Kazak@uni-konstanz.de.

1 Introduction

Finding an optimal portfolio allocation has been a target of financial research since the work of Markowitz (1952). In his work it was assumed that the key parameters of the mean-variance portfolio can be estimated with the sufficient precision. As financial markets have developed quite substantially, this crucial assumption has become unrealistic. Estimating the means and particularly covariances of asset returns is very tricky due to noise and limitations in the data. At the same time the whole field of statistical learning has been designed for recovering complicated patterns from the noisy data. This paper aims to connect the newly developed data driven methods from machine learning to the portfolio allocation problem.

Estimation risk is a well-known issue in empirical portfolio modelling. For a given performance measure, estimation risk may cause a theoretically superior portfolio strategy to be inferior compared to simple alternatives when it comes to a comparison of the performance measures based on their estimated counterparts. The most prominent example is the equally weighted ($1/N$) portfolio strategy, for which the null hypothesis of equal out-of-sample performance compared to a more sophisticated, theory based strategy often cannot be rejected at conventional significance levels (DeMiguel et al. (2009)).

In the recent literature on portfolio choice a lot of effort has been devoted to stabilizing the portfolio weight estimates by means of regularization. Among others Jagannathan and Ma (2003) propose to impose a norm-constraint directly to the portfolio optimization for stabilizing the weight estimates in small samples; Ledoit and Wolf (2003), Ledoit and Wolf (2004a), Ledoit and Wolf (2014) wrote a series of papers focusing on the improved covariance matrix estimation, which is a key ingredient in the portfolio optimization; Kourtis et al. (2012) proposed a shrinkage approach for the inverse of the covariance matrix, which can be directly used in the most of portfolio weight estimates; DeMiguel et al. (2014) introduced a VAR model to capture the serial dependence in stock returns resulting in better out-of-sample portfolio performance.

Despite all the effort no general statement can be made whether any of the approaches can outperform the equally weighted portfolio uniformly for both low- and high-dimensional setups. Notably, for the regularization-based approaches the choice of tuning parameter plays a crucial role, however there is often no clear guidance on the optimal shrinkage (regularization) intensity choice, which would at the same time be easily implemented by an empirical researcher. We address the problem from a different perspective. Instead of working with the weight estimation directly we develop a flexible algorithm which optimally combines a given set of weight estimates in a data-driven way with respect to a certain portfolio performance measure. In particular, we use the pretest estimator as a statistical tool to choose an optimal strategy with respect to the out-of-sample Certainty Equivalent and Sharpe Ratio both net of the transaction costs. The pretest estimator is based on the simple t-test commonly used in the literature (Ledoit and Wolf (2008), DeMiguel et al. (2009)), where the investor decides on how to invest his wealth for the next period based on the test outcome. Similar to shrinkage strategies which combine a given portfolio strategy with the equally weighted portfolio by some optimality criterion (DeMiguel et al. (2009), Frahm and Memmel (2010)), our pretest estimator uses the information about all underlying strategies through the outcome of the performance test. However, contrary to the shrinkage approaches, our pretest strategy can be continuously updated to incorporate the most recent information in the rolling window forecasting set-up. In the previous work Kazak and Pohlmeier (2018) show that the existing portfolio performance tests are correctly sized, but for realistic scenarios have a very low power. This implies that the pretest estimator cannot be used directly and needs to be adjusted. Our first contribution is a novel approach of optimizing the significance level for the pretest estimator.

We propose a fully data-driven and time-adaptive significance level choice, which optimizes a trade-off between Type I and Type II error with respect to the chosen portfolio performance measure. The second contribution is introducing machine learning in the pretest estimation. To the best of our knowledge this paper is the first one combining bagging with pretest estimation in the portfolio context. We modify the classical pretest estimator replacing the indicator functions with the bootstrapped probabilities, which helps to stabilize the pretest estimator and reduces portfolio turnover. In an extensive empirical study we show that our proposed bagged pretest estimator outperforms the underlying weight estimation strategies and other competitors and is robust to different parameter constellations.

This paper is organized as follows. In Section 2 we use a simple motivating example which illustrates the problem of an optimal strategy choice and propose the novel bagged pretest estimator. Section 3 provides the reader with an empirical illustration of the proposed method. Section 4 summarizes the main findings and gives an outlook on future research.

2 Pretest Estimator

Consider a standard portfolio choice set-up with N risky assets. Let r_t be an excess return vector at time t with mean vector $\mathrm{E}[r_t] = \mu$ and variance-covariance matrix $\mathrm{V}[r_t] = \Sigma$. Moreover, let $\omega(s) = \omega_{(s)}(\mu, \Sigma)$ be the $N \times 1$ vector of portfolio weights for strategy s, e.g. $\omega(g) = \frac{\Sigma^{-1}\iota}{\iota \Sigma^{-1}\iota}$ for the global minimum variance portfolio (GMVP) minimizing the portfolio variance, $\omega(e) = \frac{1}{N}\iota$ for the equally weighted portfolio and $\omega(tn) = \frac{\Sigma^{-1}\mu}{\iota' \Sigma^{-1}\mu}$ for the tangency portfolio, maximizing the Sharpe Ratio. For strategy s the portfolio return at time t is given by $r_t^p(s) = \omega(s)' r_t$ with mean $\mu_p(s) = \mathrm{E}[r_t^p(s)] = \omega(s)'\mu$ and variance $\sigma_p^2(s) = \mathrm{V}[r_t^p(s)] = \omega(s)'\Sigma\omega(s)$.

Consider a portfolio performance measure for the strategy s, $\mathcal{P}(s)$, which could be for example the Certainty Equivalent (CE) given by $CE(\omega(s)) = \mu_p(s) - \frac{\gamma}{2}\sigma_p^2(s)$ with γ being the risk aversion coefficient of the investor; or a Sharpe Ratio $SR(\omega(s)) = \frac{\mu_p(s)}{\sigma_p(s)}$. The CE is the return which makes the investor indifferent between investing into the risky portfolio or receiving the certainty equivalent return $U(CE(\omega(s))) = \mathrm{E}[U(r^p(s))]$, with $U(\cdot)$ being the utility function of the investor (Merton and Samuelson, 1992). Our analysis below concentrates on the CE as performance measure, because of its simple return interpretation. However, our proposed pretest strategies can be easily generalized for the Sharpe Ratio or any other popular portfolio evaluation criterion, whenever there exists an appropriate statistical performance test. In the following strategy s is said to outperform strategy $\tilde{s}$ if the difference in certainty equivalents is non-negative $\Delta_0(s, \tilde{s}) = CE(\omega(s)) - CE(\omega(\tilde{s})) \geq 0$.

Given the population parameters (μ, Σ) the dominant strategy according to a chosen portfolio evaluation criteria is known. However, in empirical applications the first two moments of the return process have to be estimated and the estimation risk has to be taken into account. Furthermore, investment decision is a dynamic process, therefore the financial or forecasting risk has to be accounted for as well. In empirical work the actual performance of the competing portfolio allocation strategies is often evaluated based on the out-of-sample Certainty Equivalent, which takes into account both estimation and forecasting risk (Kazak and Pohlmeier, 2018). In the following we consider a typical rolling window set-up, where for the period $t+1$ the out-of-sample portfolio return $\hat{r}_{t+1}^p(s)$ is based on a one-step forecast of the portfolio weights $\hat{\omega}_{t+1|t}(s)$ with period $\{t-T, \ldots, t\}$ as an estimation window. We adopt the standard assumption for static models that the last available estimate $\hat{\omega}_t(s)$ is used to compute the out-of-sample return for the next period: $\hat{r}^p_{t+1}(s) = \hat{\omega}_{t+1|t}(s)'r_{t+1} = \hat{\omega}_t(s)'r_{t+1}$. The estimation window is shifted one period ahead H times resulting in the $H \times 1$ vector of the out-of-sample portfolio returns $\hat{r}^p(s)$. Different portfolio strategies are then evaluated based on the out-of-sample Certainty Equivalent $\widehat{CE}_{op}(\hat{\omega}(s))$ given by:

$$\begin{aligned}
\widehat{CE}_{op}(\hat{\omega}(s)) &= \hat{\mu}_{op}(s) - \frac{\gamma}{2}\hat{\sigma}^2_{op}(s), \qquad (1)\\
\text{where:} \qquad \hat{\mu}_{op}(s) &= \frac{1}{H}\sum_{h=1}^{H}\hat{r}^p_{t+h}(s) = \frac{1}{H}\sum_{h=1}^{H}\hat{\omega}_{t+h-1}(s)'r_{t+h},\\
\hat{\sigma}^2_{op}(s) &= \frac{1}{H-1}\sum_{h=1}^{H}\left(\hat{r}^p_{t+h}(s) - \hat{\mu}_{op}(s)\right)^2.
\end{aligned}$$

The driving force of the portfolio performance based on the out-of-sample CE is the estimation noise, i.e. theoretically superior strategies usually do not perform well in practice as the estimation error dominates the theoretical gain. Depending on the size of the portfolio N, length of the estimation window T portfolio performance varies dramatically.

2.1 Motivating Example

As an illustrative example consider an investor who chooses among three different strategies and wants to rebalance his portfolio monthly. Based on the monthly excess returns of 100 industry portfolios from K.R.French database[1] he estimates the weights of the following strategies:

1. GMVP based on the plug-in covariance matrix estimator

$$\hat{\omega}(g) = \frac{\hat{\Sigma}^{-1}\iota}{\iota\hat{\Sigma}^{-1}\iota}, \qquad (2)$$

   where $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})(r_t - \bar{r})'$ and ι is an $N \times 1$ vector of ones.

2. Tangency portfolio with a shrunken covariance matrix

$$\hat{\omega}(tn) = \frac{\hat{\Sigma}^{-1}(\lambda)\hat{\mu}}{\iota\hat{\Sigma}^{-1}(\lambda)\hat{\mu}}, \qquad (3)$$

   where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t$ and $\hat{\Sigma}(\lambda) = \hat{\Sigma} + \lambda I_N$ with $\lambda = 0.05N$ and I_N - identity matrix of size N.

[1] The data is taken from K.R.French website and contains monthly excess returns from 01/1953 till 12/2015.

3. Equally weighted portfolio

$$\hat{\omega}(e) = \omega(e) = \frac{1}{N}\iota. \tag{4}$$

For the portfolio evaluation the length of the out-of-sample period is set to $H = 500$ and the risk aversion parameter to $\gamma = 1$. Transaction costs at period t which an investor has to pay every month after portfolio rabalancing are computed as follows:

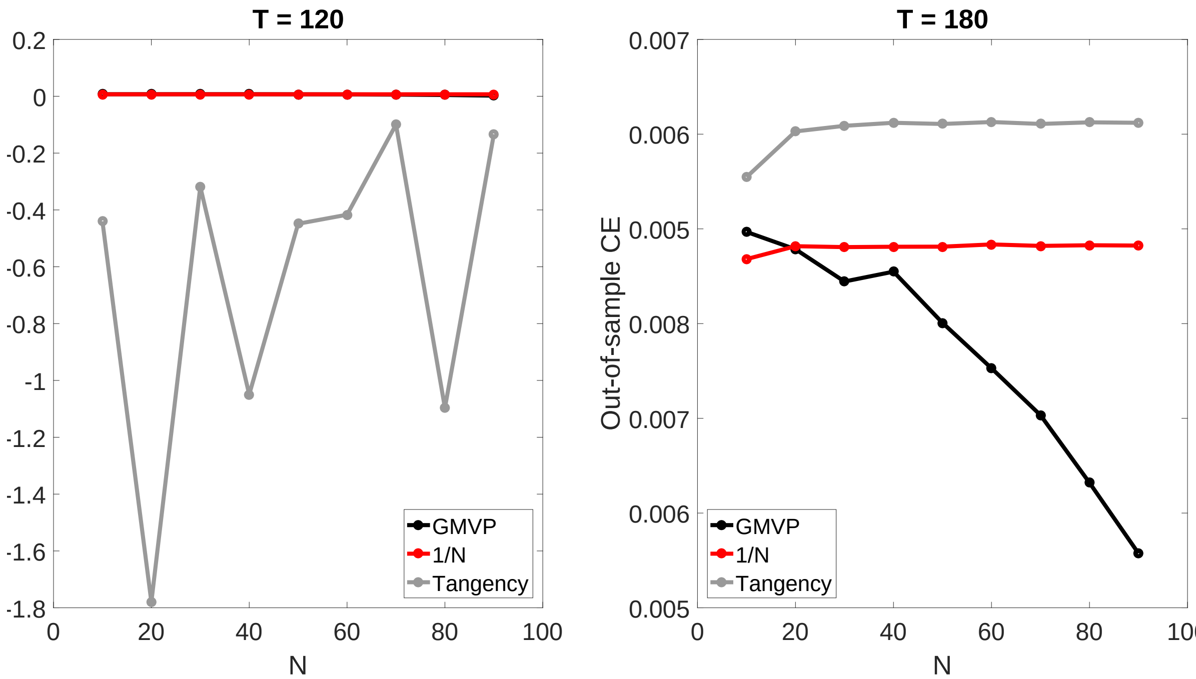
$$TC_t(s) = c \cdot \sum_{j=1}^{N} |\hat{\omega}_{j,t+1}(s) - \hat{\omega}_{j,t^+}(s)|, \tag{5}$$

where $TC_t(s)$ denotes transaction costs at period t, $\hat{\omega}_{j,t^+}$ - portfolio weight before rebalancing at $t+1$ and c - cost per transaction (5 basis points, DeMiguel et al. (2009)). The out-of-sample CE is then computed based on the net portfolio returns $\hat{r}_t^{p,net}(s) = \hat{r}_t^p(s) - TC_t(s)$ in the very similar way to (1). For a more general comparison for a given pair of (N, T) we randomly draw N assets from the available asset space and compute the out-of-sample CE for each of the 500 random draws.

The right panel of Figure 1 depicts the average out-of-sample CE for a grid of portfolio sizes and the in-sample estimation window length T of 15 years (180 months). The equally weighted portfolio (in red) does not require weight estimation and is therefore stable across the portfolio sizes N. The out-of-sample CE of GMVP (black line) is deteriorating with the increase in N as the empirical variance-covariance matrix of the returns adds more and more estimation noise. The tangency portfolio (gray line) with shrunken covariance matrix performs very well for all N. However, on the left panel with a smaller estimation window length $T = 120$ the performance of the tangency portfolio worsens dramatically (note the difference in scaling of the y-axes). Tangency portfolio is known to produce extremely unstable weight estimates for the smaller estimation windows, while the performance of the GMVP and the equally weighted portfolio is considerably less sensitive (Okhrin and Schmid, 2006). Moreover, in some circumstances it is better to keep the estimation window length smaller, e.g. due to structural breaks in the financial markets only the recent information should be included in the weight estimation.

Summing up: No general statement can be done with respect to the optimal strategy for a given portfolio space and the estimation window length. In this paper we develop a data driven procedure for an optimal portfolio allocation strategy choice, which outperforms the underlying

Figure 1: Out-of-sample CE for underlying strategies



Lines on the plots represent the average out-of-sample CE computed on net portfolio returns over a 500 randomly drawn portfolios of size N (x-axis). For each randomly drawn portfolio the out-of-sample CE is computed over an out-of-sample evaluation horizon of $H = 500$ observations, risk aversion parameter $\gamma = 1$. The left panel depicts the average out-of-sample CE with the weights of GMVP (in black) from eq.(2), tangency portfolio (in gray) from eq.(3), and equally weighted portfolio (in red) from eq.(4) computed over an estimation window length of $T = 120$ observations. The right panel plots the average out-of-sample CE for $T = 180$.

weight estimation strategies regardless of the parameter constellations.

2.2 Pretest as a Decision Rule

The first part of our algorithm is developing an appropriate decision rule for choosing a portfolio allocation strategy. We use the pretesting as a statistical tool helping the investor to decide between different strategies in a data-driven way. Assume for the sake of simplicity that the investor has to decide between two alternative strategies s and $\tilde{s}$ and maximizes the expected utility[2]. The difference in the out-of-sample CE's between the two strategies is defined as

$$\Delta_{op}(s, \tilde{s}) = CE_{op}(\hat{\omega}_t(s)) - CE_{op}(\hat{\omega}_t(\tilde{s})),$$
$$CE_{op}(\hat{\omega}_t(s)) = \mu_{op}(s) - \frac{\gamma}{2}\sigma^2_{op}(s), \quad \text{with}$$

[2]Note that all the conclusions of this Section also hold for testing Sharpe Ratio difference or difference in portfolio variance.

$$
\begin{aligned}
\mu_{op}(s) =& \,\mathrm{E}\left[\hat{r}^p_{t+1}(s)\right] = \mathrm{E}\left[\hat{\omega}_t(s)\right]' \mu, \\
\sigma^2_{op}(s) =& \,\mathrm{V}\left[\hat{r}^p_{t+1}(s)\right] = \mathrm{E}\left[\hat{\omega}_t(s)'\Sigma\,\hat{\omega}_t(s)\right] + \mu'\,\mathrm{V}\left[\hat{\omega}_t(s)\right]\mu.
\end{aligned}
$$

The goal is to choose either strategy s or strategy $\tilde{s}$ depending on the test outcome. Null and alternative hypotheses take the usual one-sided form:

$$
H_0: \quad \Delta_{op}(s,\tilde{s}) \leq 0 \qquad \text{and} \qquad H_1: \quad \Delta_{op}(s,\tilde{s}) > 0. \tag{6}
$$

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

$$
\omega_t(s,\tilde{s},\alpha) = \mathbb{1}\left(\hat{\Delta}_{op}(s,\tilde{s}) > \Delta^*(\alpha)\right)\left(\omega_t(s) - \omega_t(\tilde{s})\right) + \omega_t(\tilde{s}), \tag{7}
$$

with the estimated CE difference $\hat{\Delta}_{op}(s,\tilde{s}) = \widehat{CE}_{op}(s) - \widehat{CE}_{op}(\tilde{s})$ and the critical value $\Delta^*(\alpha)$ for significance level α. In other words the pretest estimator chooses the strategy s if it is significantly better than the alternative and the sensitivity of the pretest estimator depends on α: the lower is the chosen nominal level, the stricter is the pretest rule and the greater should be the difference $\hat{\Delta}_{op}(s,\tilde{s})$ for choosing the strategy s over $\tilde{s}$.

There are two main difficulties arising from applying the pretest estimator defined in (7) in practice. First of all, at time period t the investor does not know the out-of-sample CE difference $\hat{\Delta}_{op}(s,\tilde{s})$ and therefore the test decision from (6) is unknown. Secondly, in empirical applications an investor has to decide on α, which influences the performance of the pretest estimator, e.g. for $\alpha = 50\%$ the CE of strategy s has to be just slightly larger than the CE of $\tilde{s}$ in order to be chosen, whereas for the commonly used levels of significance of 1% and 5% the difference $\hat{\Delta}_{op}(s,\tilde{s})$ has to be fairly large for the null rejection. Kazak and Pohlmeier (2018) show that in realistic scenarios the empirical power of the portfolio performance tests is very low, which implies that even if the strategy s is truly superior, the pretest estimator is not able to choose the dominating strategy. Therefore going for the conservative α-level is not a reasonable choice, as in the presence of low power it will force the pretest estimator to choose $\tilde{s}$ even in cases when s is dominating. In particular, the problem of the low power calls for an optimal trade-off between Type I and Type II error. In this paper we propose a feasible pretest estimator with a data-driven and time adaptive significance level choice.

2.2.1 Within-Sample Pretest Strategy

We solve the first issue of unfeasible testing by constructing the pretest weight estimator based on the in-sample CE difference which is available at time t. First, at time t the weights $\hat{\omega}_t(s)$ are estimated based on the sample $\{t-T,\ldots,t\}$. The estimated within sample CE for the strategy s is computed as

$$\widehat{CE}_{in,t}(s) = \widehat{CE}_{in}(s|t-T,...,t) = \hat{\omega}_t(s)'\bar{r}_t - \frac{\gamma}{2}\hat{\omega}_t(s)'\hat{\Sigma}_t\hat{\omega}_t(s), \tag{8}$$

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 deciding between s and the benchmark $\tilde{s}$ is defined as

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

where the standard error for the CE difference is computed via Delta method (DeMiguel et al., 2009) or an appropriate block bootstrap (Ledoit and Wolf, 2008). For the multivariate comparison where the investor chooses between M alternatives $s_1,...,s_M$ and a benchmark strategy $\tilde{s}$ and for a given significance level α the pretest estimator based on the in-sample CE is defined as

$$\begin{aligned}
\mathbb{1}(s_i,\alpha) &= \mathbb{1}\left[\max\left[t_{in,t}(s_1,\tilde{s}),...,t_{in,t}(s_M,\tilde{s}),t^*(\alpha)\right] = t_{in,t}(s_i,\tilde{s})\right], \quad i=1,...,M,\\
\hat{\omega}_{in,t}(\mathcal{S},\alpha) = \hat{\omega}_{in,t}(s_1,..,s_M,\tilde{s},\alpha) &= \sum_{i=1}^{M}\mathbb{1}(s_i,\alpha)\hat{\omega}_t(s_i) + \left(1-\sum_{i=1}^{M}\mathbb{1}(s_i,\alpha)\right)\hat{\omega}_t(\tilde{s}),
\end{aligned} \tag{9}$$

where $\mathcal{S}=\{s_1,..,s_M,\tilde{s}\}$, $t^*(\alpha)$ is the corresponding critical value for the nominal level α: the $(1-\alpha)$ quantile of the standard normal distribution. In other words, the pretest estimator chooses the strategy with the largest standardized difference from the benchmark $\tilde{s}$ and which at the same time crosses the threshold $t^*(\alpha)$.

We now address the problem of the significance level choice. From the purely statistical perspective, increasing the nominal level in order to improve on the power of the test is not meaningful. From the investors perspective, however, there is a well-defined portfolio performance measure, such that the trade-off between the Type I and Type II error of the pretest estimator can be optimized with respect to it. We propose to choose a significance level α, which maximizes the difference in portfolio performance measure. In particular, at each period t the pretest weight estimates are computed according to (9) on the grid of (0,1) α-values of length J. For every $\hat{\omega}_{in,t}(\mathcal{S}, \alpha_j), \quad j = 1, .., J$ the in-sample CE of the pretest estimator is computed similarly to eq.(8):

$$\widehat{CE}_{in,t}(\mathcal{S}, \alpha_j) = \hat{\omega}_{in,t}(\mathcal{S}, \alpha_j)' \bar{r}_t - \frac{\gamma}{2} \hat{\omega}_{in,t}(\mathcal{S}, \alpha_j)' \hat{\Sigma}_t \hat{\omega}_{in,t}(\mathcal{S}, \alpha_j). \tag{10}$$

Finally, the in-sample CE optimizing significance level α^*_{t+1} is chosen for the test, determining the strategy for the next period $t+1$:

$$\alpha^*_{in,t+1} = \arg\max_{\alpha} \widehat{CE}_{in,t}(\mathcal{S}, \alpha_j). \tag{11}$$

The above procedure is repeated with every shift of the estimation window. In practice this results in a very unstable series of significance level choices $\{\alpha^*_{in,t+1}, ...\alpha^*_{in,t+H}\}$, as the choice of $\alpha^*_{in,t+1}$ is data driven and also depends on the instable estimates of the portfolio weights. On the other hand, the sequence of α^*_{in}'s along the rolling estimation window takes into account changes of the return process across time, e.g. volatility regimes. In order to mitigate the instability problem we suggest to adaptively smooth the α^*_{in}-series according to

$$\alpha^s_{t+1} = (1 - \lambda)\alpha^*_{in,t+1} + \lambda \alpha^s_t, \tag{12}$$

where the tuning parameter λ is chosen to control the degree of smoothness. The adaptive smoothing takes into account not only the latest optimal choice $\alpha^*_{in,t+1}$ but also the previous estimates with geometrically decaying weights. The smoothing parameter λ is chosen via a grid search in the similar fashion as the $\alpha^*_{in,t+1}$: for a given couple $(\alpha^*_{in,t+1}, \alpha^s_t)$ the in-sample CE is computed on the grid of λ's and the optimal λ is the one maximizing the in-sample CE of the pretest estimator $\widehat{CE}_{in,t}(\mathcal{S}, \alpha^s_{t+1})$.

2.2.2 Out-of-Sample Pretest Strategy

Another feasible pretest estimator may be obtained by performing a pseudo-out-of-sample exercise which is commonly used for parameter training in machine learning. The goal of the pretest estimator is to choose an optimal strategy in a data-driven way which results in the highest out-of-sample CE. Choosing the strategy based on the in-sample comparison does not necessarily provide a good out-of-sample choice: the pretest estimator as defined in eq.(9) does not take into account transaction costs and the forecasting risk. A feasible out-of-sample pretest estimator might be obtained by dividing the within-sample period into two parts, where the first part is used for the weight estimation and the second part is used for the pseudo-out-of-sample return computation. The optimal strategy is then defined as the one having the largest pseudo-out-of-sample CE net of the transaction costs. In particular, the weights for the strategy s are computed based on the sample of length $T/2$: $\{t-T,..,t-T/2\}$. The out-of-sample portfolio returns are computed in the rolling window of length $T/2$ and the transaction costs are subtracted from the out-of-sample returns at each time point t as in eq.(5). The resulting pseudo-out-of-sample CE is similar to eq.(1):

$$\widehat{CE}^{*}_{op,t}(\hat{\omega}(s)) = \hat{\mu}^{*}_{op,t}(s) - \frac{\gamma}{2}\hat{\sigma}^{*2}_{op,t}(s), \tag{13}$$

$$\text{where:} \quad \hat{\mu}^{*}_{op,t}(s) = \frac{1}{T/2}\sum_{h=-T/2+1}^{-T} \hat{r}^{p}_{t+h}(s) = \frac{1}{T/2}\sum_{h=-T/2+1}^{-T} \hat{\omega}_{t+h-1}(s)' r_{t+h},$$

$$\hat{\sigma}^{*2}_{op,t}(s) = \frac{1}{T/2-1}\sum_{h=-T/2+1}^{-T} \left(\hat{r}^{p}_{t+h}(s) - \hat{\mu}_{op,t}(s)\right)^2.$$

The out-of-sample pretest estimates of portfolio weights are computed based on the difference in the pseudo-out-of-sample CE:

$$t_{op,t}(s_i,\tilde{s}) = \frac{\widehat{CE}^{*}_{op,t}(s) - \widehat{CE}^{*}_{op,t}(\tilde{s})}{S.E.\left[\widehat{CE}^{*}_{op,t}(s) - \widehat{CE}^{*}_{op,t}(\tilde{s})\right]}, \tag{14}$$

$$\mathbb{1}_{op}(s_i,\alpha) = \mathbb{1}\left[\max\left[t_{op,t}(s_1,\tilde{s}),...,t_{op,t}(s_M,\tilde{s}),t^{*}(\alpha)\right] = t_{op,t}(s_i,\tilde{s})\right], \quad i=1,...,M, \tag{15}$$

$$\hat{\omega}_{op,t}(\mathcal{S},\alpha) = \hat{\omega}_{op,t}(s_1,..,s_M,\tilde{s},\alpha) = \sum_{i=1}^{M}\mathbb{1}_{op}(s_i,\alpha)\hat{\omega}_t(s_i) + \left(1-\sum_{i=1}^{M}\mathbb{1}_{op}(s_i,\alpha)\right)\hat{\omega}_t(\tilde{s}). \tag{16}$$

The pseudo-out-of-sample CE from eq.(13) is used only for the computations of the indicator functions, and the out-of-sample pretest estimator chooses the weight estimates computed on the whole in-sample estimation window. The optimal significance level choice for the out-of-sample pretest estimator can be done in the very same way as in eq.(12) using a grid search. Note that the out-of-sample portfolio returns used in the pseudo-out-of-sample CE are net of the transaction costs and therefore by construction the pretest estimator takes into account the amount of rebalancing or turnover. Also note, that the estimation window used in eq.(13) is reduced to half the size of the initial estimation window, which results in adding more estimation noise to the problem and making the weight estimates very unstable. Moreover, this pretest estimator is a sharp thresholding strategy choosing only one strategy for investing in the next period. In the rolling window scenario this may lead to large turnover costs, if the pretest estimator switches between different strategies frequently with the rolling estimation window.

2.2.3 Optimal Portfolio Allocation with Statistical Learning

As a solution we propose to stabilize the pretest estimator using bagging. In order to do so, the in-sample estimation window is bootstrapped B times smoothing the indicator functions of the pretest estimator by the bootstrapped probabilities. The computation of the bagged out-of-sample pretest estimator can be summarized in a following algorithm[3]:

Bagging the pretest estimator

1. At the period t define the estimation window of length T: $\{r_{t-T}, ..., r_t\}$.
2. Divide the sample into two parts and compute the pseudo-out-of-sample CE net of the transaction costs according to eq.(13) and the out-of-sample test statistic according to eq.(14) for each pair of strategies $(s_i, \tilde{s})$, $i = 1, ..., M$.
3. For a grid of J α values compute the out-of-sample pretest weight estimates according to eq.(16) and choose $\alpha^*_{op,t+1}$ which results in the largest out-of-sample CE on the grid.
4. For a grid of λ values compute the smoothed $\alpha^s_{op,t+1}$ using the previous optimal significance level and choose the one maximizing the out-of-sample CE of the pretest estimator.

[3]Note, that monthly returns do not posses any significant SACF or SPACF patterns, therefore the i.i.d. bootstrap of Efron (1992) is appropriate to use. For the returns of higher frequencies one should use the circular block bootstrap by Politis and Romano (1992).

For every bootstrap iteration $b = 1, ..., B$ compute (5) and (6):

5. Randomly sample the rows of the in-sample $T \times N$ data with replacement and repeat step (2) for the out-of-sample test statistic computation keeping the weight estimates fixed.
6. For a significance level chosen in step 4 compute the indicator functions $\mathbb{1}^{b}_{op}(s_i, \alpha^s_{op,t+1})$ for every strategy i and bootstrap iteration b.
7. The bagged probability of the strategy s_i is then defined as $\hat{p}(s_i, \alpha^s_{op,t+1}) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}^{b}_{op}(s_i, \alpha^s_{op,t+1})$ using eq.(15).
8. Finally the bagged out-of-sample 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:
$$\hat{\omega}^{B}_{op,t}(\mathcal{S}, \alpha^s_{op,t+1}) = \sum_{i=1}^{M+1} \hat{p}(s_i, \alpha^s_{op,t+1})\hat{\omega}_t(s_i). \tag{17}$$

The proposed out-of-sample bagged pretest estimator is novel by providing the investor with a fully data driven way of an optimal portfolio allocation strategy choice according to a specific performance measure. For instance, if the investor is looking for a strategy with the highest out-of-sample Sharpe Ratio, the proposed bagging algorithm can be easily adapted. Moreover, the algorithm takes into account the amount of transaction costs and is time adaptive through the significance level choice.

Bagging is a powerful way of variance reduction for the unstable estimators, e.g. it is widely used for stabilizing classification and regression trees which are based on sharp thresholding. Bühlmann et al. (2002) show that bagging reduces the variance of the plug-in pretest estimator and the same intuition holds in the portfolio context. For a fixed strategy s_i let $F_{s_i}(\cdot)$ denote the cumulative distribution function of the out-of-sample test statistic from (14). The mean and the variance of the pretest indicator function is:

$$\begin{aligned}
&\mathrm{E}\left[\mathbb{1}\left(t_{op}(s_i, \tilde{s}) > t^*(\alpha)\right)\right] = 1 - \mathrm{E}\left[\left(t_{op}(s_i, \tilde{s}) \leq t^*(\alpha)\right)\right] = 1 - F_{s_i}(t^*(\alpha)),\\
&\mathrm{V}\left[\mathbb{1}\left(t_{op}(s_i, \tilde{s}) > t^*(\alpha)\right)\right] = \left(1 - F_{s_i}(t^*(\alpha))\right) F_{s_i}(t^*(\alpha)),
\end{aligned}$$

where $t^*(\alpha)$ denotes the threshold corresponding to the significance level α, i.e. for a fixed significance level $\alpha = 50\%$ the threshold is exactly zero and the variance of the pretest indicator is $1/4$ for a symmetric F_{s_i}. Note that the pretest estimator is unstable in the sense that it assumes values 0 or 1 with positive probability. Following Corollary 2.1 of Bühlmann et al. (2002) the variance of the pretest bagged pretest indicator for $\alpha = 50\%$ is

$$\mathrm{V}\left[\frac{1}{B}\sum_{b=1}^{B}\mathbb{1}_{op}^{b}(s_i, 50\%)\right] = \mathrm{V}\left[F_{s_i}(t_{op}(s_i, \tilde{s}))\right] \to \mathrm{V}\left[U\right],$$

where U denotes a $[0,1]$-uniformly distributed random variable and its variance equals $1/12$, which is *three times smaller* than the variance of the single pretest indicator function. The variance of the indicators in the portfolio context translates into the amount of portfolio rebalancing imposed by pretesting: smaller variance of the indicator functions implies less turnover. In other words using bagging helps to stabilize the problem and reduce transaction costs, as the bootstrapped probabilities smooth the transition between the strategies along the rolling window.

2.2.4 Sequential Performance Weighting Strategy

As a competitor for the bagged pretest estimator we consider a sequential relative performance weighting inspired by the approach of Shan and Yang (2009) used in the forecast combinations. The idea behind this approach is very similar to boosting. First, at each period t the relative performance of different strategies is measured by the exponential function of the in-sample CE. Then the time-adaptive weight $d_{t,i}$ for the strategy i is computed according to (18), where the initial values for the weights are fixed to $d_{0,i} = \frac{1}{M+1}$. The resulting weight estimator is denoted as $\hat{\omega}_t^{SP}(\mathcal{S})$:

$$d_{t,i} = \frac{d_{t-1,i}\exp(\widehat{CE}_{in,t}(s_i))}{\sum_{j=1}^{M+1} d_{t-1,j}\exp(\widehat{CE}_{in,t}(s_j))}, \tag{18}$$

$$\hat{\omega}_t^{SP}(\mathcal{S}) = \sum_{j=1}^{M+1} d_{t,i}\hat{\omega}_t(s_i), \quad i = 1, ..., M+1 \tag{19}$$

where $\mathcal{S} = \{s_1, .., s_M, \tilde{s}\}$ is a space of the underlying strategies and $d_{t,i}$ are the time-adaptive coefficients used to weight the relative performance of the underlying strategies. This approach is computationally very easy and is adapting the weights with the every shift of the estimation window. It is also a smooth combination of the underlying strategies, potentially requiring less rebalancing and reducing transaction costs. Using the exponential loss function for reweighing the input of unstable classifiers is used in boosting and has been widely applied in machine learning literature. In the portfolio allocation context the boosting idea is applied to portfolio strategies s_i, which are reweighed with shifting estimation window depending on their past performance.

3 Empirical Evidence

In this section we provide empirical evidence on the performance of our pretested portfolio strategies introduced in the previous section. We first continue with the example from Section 2.1 and consider the GMVP, tangency and the equally weighted portfolios as the underlying competing strategies. We use data on monthly excess returns from K.R.French database[4] of a 100 industry portfolios. In the analysis below for a given portfolio size N we randomly draw N out of 100 unique assets and report the average portfolio performance over 500 random draws. For bagging the number of bootstrap iterations is set to $B = 200$, the out-of-sample evaluation window is fixed to $H = 500$ observations and the risk aversion parameter is set to $\gamma = 1$.

Table 1: Out-of-sample CE for $T = 120$.

	N = 10	N = 20	N = 30	N = 40	N = 50	N = 60	N = 70	N = 80	N = 90
GMVP	0.0069	0.0070	0.0070	0.0072	0.0069	0.0064	0.0055	0.0045	0.0023
Tangency	-0.0545	-0.8500	-0.6386	-0.1405	-1.0088	-1.2433	-0.1927	-43.9858	-0.2068
1/N	0.0068	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069
$\alpha = 5\%$	**0.0073**	0.0073	0.0041	0.0071	0.0071	0.0067	-0.0085	-0.0130	0.0008
α^s_{in}	0.0070	0.0070	-0.0087	0.0044	-0.4950	-0.2427	-0.0104	-0.3138	0.0000
α^s_{in} B	0.0072	0.0073	0.0073	0.0061	0.0072	0.0072	0.0071	0.0069	-0.0017
α^s_{op}	0.0041	-0.1294	-0.0190	-0.0002	-0.0483	-0.1552	-0.0124	-0.4365	-0.0121
α^s_{op} B	**0.0073**	**0.0074**	**0.0074**	**0.0075**	0.0074	**0.0075**	**0.0074**	**0.0074**	**0.0074**
Seq. Perf.	**0.0073**	**0.0074**	**0.0074**	0.0067	**0.0076**	**0.0075**	0.0073	0.0072	0.0066

Figures in the table correspond to the average out-of-sample CE computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample CE is computed over an evaluation horizon of $H = 500$ observations, risk aversion parameter $\gamma = 1$, in-sample estimation window length $T = 120$. $\alpha = 5\%$ denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, α^s_{in} allows for a flexible significance level choice according to eq. (12) and α^s_{in} B is the bagged version of α^s_{in}. α^s_{op} and α^s_{op} B denote the out-of-sample significance level choice and the one combined with bagging as in eq.(17). Seq. Perf. corresponds to eq.(19). Numbers in bold correspond to the largest CE for a given portfolio size N.

Table 1 reports the average out-of-sample Certainty Equivalent net of the transaction costs for the underlying strategies (first block), pretest estimators (second block) and sequential performance weighting (last row) for the in-sample estimation window length of 10 years (120 monthly observations) with different portfolio sizes N in the corresponding columns. As before, with the increase in N the out-of-sample CE of the GMVP decreases, CE of the tangency portfolio is negative resulting from the extreme weight estimates and the CE of the equally weighted portfolio is stable across N. The forth row corresponds to the CE of the pretest

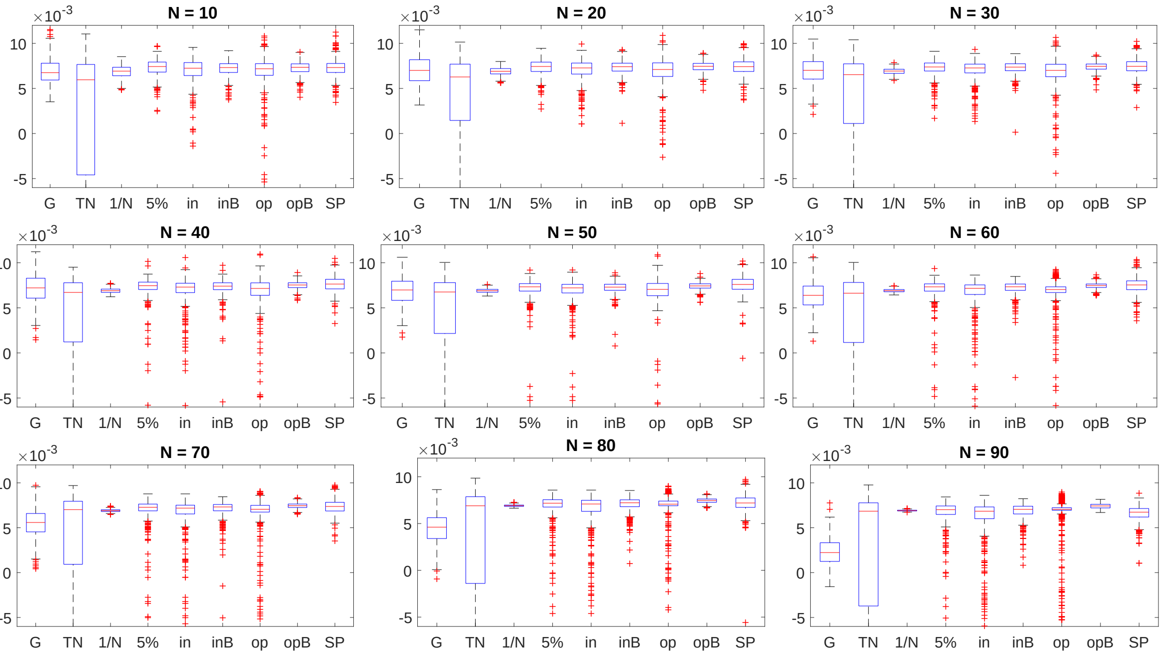
[4]The data is taken from K.R.French website and contains monthly excess returns from 01/1953 till 12/2015.

estimator based on the in-sample testing as in eq. (9) with a fixed conventional significance level of 5%. Here the pretest estimator chooses the strategy which standardized in-sample CE difference from the $1/N$ is greater than the 95% quantile of the standard normal distribution. This pretest estimator is working quite well for smaller asset spaces, however for $N \geq 30$ its performance deteriorates and the CE of the pretest estimator is no longer greater than the values of the underlying strategies, potentially suffering from the low power of the underlying portfolio performance tests. Allowing for a flexible significance level choice according to eq.(12) based on the in-sample CE is quite unstable across N and is not performing well, i.e. the pretest estimator is based on $T = 120$ observations only, which is not enough to determine the best strategy. However, combining this pretest estimator with bagging ($\alpha^s_{in}B$) outperforms the underlying strategies up to $N = 80$ having a greater out-of-sample CE than the best of the stand-alone strategies. The next row (α^s_{op}) reports the average out-of-sample CE for the pretest estimator with adaptively smoothed out-of-sample significance level choice, which takes into account transaction costs, but suffers from the short evaluation period of $T/2$. Out-of-sample pretest estimator is very unstable across N and does not outperform the underlying strategies, however it always outperforms the worst underlying strategy. Notably, the bagged out-of-sample pretest estimator ($\alpha^s_{op}B$) performs extremely well, producing the largest out-of-sample CE for all portfolio sizes N always greater than the best underlying strategy. The sequential performance weighting results in the average out-of-sample CE which is as good as the bagged out-of-sample pretest estimator, except for $N = 90$, where it is not able to beat the equally weighted portfolio benchmark.

Note, that our portfolios of different size are designed such that the asset spaces for the smaller portfolios are true subsets of the larger ones. This guarantees that the theoretical CE of a larger portfolio always has to dominate the CE of a smaller portfolio. For the empirical portfolios this dominance frequently does not hold even for regularized portfolio estimates, as the increase in estimation noise due to the increase in portfolio dimension dominates the increase in the theoretical gains. A typical example is the sequence of CEs for the GMVP (first row of Table 1). For our bagged pretest estimator, however, we do not find any decrease in performance with increasing portfolio dimension and estimation noise.

The proposed bagged out-of-sample pretest estimator is performing very well even in the

Figure 2: Boxplots of the out-of-sample CE for $T = 120$.



Each boxplot corresponds to the distribution of the out-of-sample CE computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample CE is computed over an evaluation horizon of $H = 500$ observations, risk aversion parameter $\gamma = 1$, in-sample estimation window length $T = 120$. X-axes denote different ways of computing portfolio weights: G from eq.(2), TN from eq.(3), 1/N from eq.(4), 5% denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, *in* allows for a flexible significance level choice according to eq.(12), *inB* is the bagged version of *in*. *op* and *opB* denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). SP corresponds to eq.(19).

presence of the estimation noise and is preferred to the sequential relative performance weighting. Figure 2 depicts the boxplots of the out-of-sample CE's of the considered strategies across the 500 random draws of portfolios of a given size N. The boxplot of the bagged out-of-sample pretest estimator denoted by opB shows how stable this strategy is and how few negative outliers does it have in comparison to the other ones. For instance, for $N \geq 40$ the out-of-sample CE's of the pretest estimators despite of having very similar medians are very different from each other in terms of the number of negative outliers. Comparing boxplots in with inB and op with opB shows that bagging always helps to stabilize the out-of-sample performance of the pretest estimators and works exceptionally well in the combination with the time-adaptive out-of-sample significance level choice. In particular, the out-of-sample CE of the proposed estimator is almost as stable as the equally weighted portfolio and at the same time it outperforms the $1/N$ benchmark in mean and has very few outliers in comparison with the sequential performance weighting from (19).

Table 2: Turnover for $T = 120$.

	N = 10	N = 20	N = 30	N = 40	N = 50	N = 60	N = 70	N = 80	N = 90
GMVP	0.0749	0.1413	0.2036	0.2680	0.3365	0.4160	0.5127	0.6415	0.8293
Tangency	0.9084	1.4565	1.1571	0.7946	1.7806	1.5813	1.1478	6.4067	1.0036
1/N	0.0217	0.0222	0.0225	0.0225	0.0226	0.0226	0.0227	0.0227	0.0227
$\alpha = 5\%$	0.0367	0.0437	0.0789	0.0641	0.0609	0.0848	0.1673	0.2719	0.1914
α_{in}^{s}	0.0485	0.0559	0.1685	0.1184	0.4787	0.5698	0.2326	0.6230	0.2454
α_{in}^{s} B	0.0301	0.0362	0.0401	0.0689	0.0499	0.0546	0.0642	0.0712	0.1585
α_{op}^{s}	0.1241	0.3858	0.2446	0.2049	0.2674	0.5261	0.2833	0.6939	0.2216
α_{op}^{s} B	0.0264	0.0305	0.0319	0.0331	0.0336	0.0332	0.0341	0.0353	0.0382
Seq. Perf.	0.0364	0.0502	0.0669	0.1007	0.0982	0.1157	0.1352	0.1575	0.1872

Figures in the table correspond to the average turnover over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample CE is computed over an evaluation horizon of $H = 500$ observations, in-sample estimation window length $T = 120$. $\alpha = 5\%$ denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, α_{in}^{s} allows for a flexible significance level choice according to eq.(12) and α_{in}^{s} B is the bagged version of α_{in}^{s}. α_{op}^{s} and α_{op}^{s} B denote the out-of-sample significance level choice and the one combined with bagging as in eq.(17). Seq. Perf. corresponds to eq.(19).

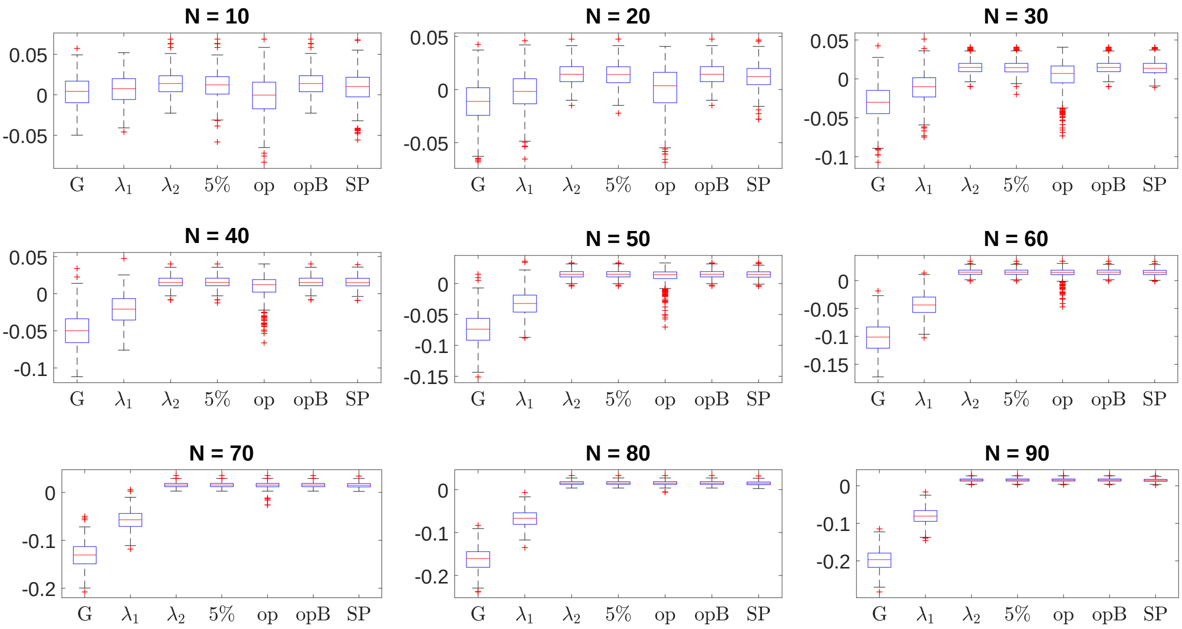
The source of the superior performance of the bagged pretest estimator is the reduction in transaction cost. Table 2 reports the average turnover[5] of the considered strategies. As expected, the equally weighted portfolio produces the smallest turnover among the considered strategies, the turnover of the GMVP and the tangency portfolio increases with the increase in N. Bagging the pretest estimators reduces turnover and the average turnover of the proposed bagged out-of-sample pretest estimator (α_{op}^{s} B) is as small as the one of the equally weighted portfolio. The average turnover of sequential performance weighting turns out to be considerably larger and steeply increases with the portfolio size N. Moreover, our novel approach outperforms the sequential relative performance weighting also in terms of the out-of-sample Sharpe Ratio (SR) based on the net returns. Figure 4 in Appendix depicts the boxplots of the Sharpe Ratios across the 500 random portfolios of size N. Similarly to the out-of-sample CE, bagging stabilizes the SR and reduces the number of outliers. Furthermore, the results are robust to the in-sample estimation window length. Tables 3 and 4 report the average out-of-sample CE and turnover of the considered strategies for $T = 180$. In this case, the tangency portfolio and the GMVP perform well and the proposed bagged pretest estimator is performing as good as the best underlying strategy. Figures 6 and 5 report the boxplots of the out-of-sample CE and SR for different portfolio sizes N and $T = 180$, where again, the most stable strategy with less outliers

[5] Turnover of a strategy s is computed as $TO(s) = \frac{1}{H}\sum_{t=1}^{H}\left(\sum_{j=1}^{N}|\hat{\omega}_{j,t+1}(s) - \hat{\omega}_{j,t^{+}}(s)|\right)$

is the bagged out-of-sample pretest estimator, performing very well for all randomly drawn portfolios.

The proposed bagging algorithm is not only useful for choosing the optimal portfolio allocation strategy, it can also be adapted to choosing an optimal tuning parameter. In the second example we consider daily returns of S&P500 constituents from January 2014 until end of December 2014 ($T = 251$ days) for estimation and January 2015 - December 16 ($H = 503$ days) for evaluation periods. We consider 3 different versions of the covariance matrix shrinkage for the GMVP allocation in eq.(2): (1) the plug-in sample covariance estimator; (2) shrinkage to market estimator by Ledoit and Wolf (2004b) denoted by λ_1; (3) shrinkage proportional to the portfolio dimension $\lambda_2 = 0.05 \cdot N$. Bagging is adjusted for the daily data, such that we now use circular block bootstrap.

Figure 3: Boxplots of the out-of-sample SR for $T = 120$.



Each boxplot corresponds to the distribution of the out-of-sample SR computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample SR is computed over an evaluation horizon of $H = 503$ observations, in-sample estimation window length $T = 251$. X-axes denote different ways of computing portfolio weights: G from eq. (2), λ_1 for GMVP with Ledoit and Wolf (2004b) shrinkage, $\lambda_2 = 0.05 \cdot N$ for Shrunken GMVP similar to eq.(3), 5% denotes the pretest estimator based on the out-of-sample test statistic as in with a fixed significance level of 5%, *op* and *opB* denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). SP corresponds to eq.(19).

Figure 3 depicts the boxplots of the out-of-sample Sharpe Ratios of the considered strategies across the 500 randomly drawn portfolio of size N. On every graph the first three boxplots correspond to the underlying versions of the GMVP, where λ_2-type of covariance matrix shrinkage converges to the equally weighted allocation with an increase in N. Not surprisingly for the noisy daily data the larger the asset space, the more pronounced is the dominance of the $GMVP(\lambda_2)$, e.g. for $N = 90$ the intense shrinkage has the highest Sharpe Ratio among the underlying strategies for all 500 randomly drawn portfolios. Therefore, all of the proposed estimators consistently choose the $GMVP(\lambda_2)$ strategy in every period for every portfolio. This feature of the portfolio choice algorithm guarantees that whenever there is a clear dominating strategy it will always be selected. On the contrary, for the smaller portfolio dimensions of up to $N = 30$ assets, pretest-based estimator and sequential performance weighting have some room for improvement over the underlying strategies. In this case the bagged strategy is as stable as the intense shrinkage and has less negative outliers than the sequential performance weighting. Tables 5 and 6 report the average out-of-sample Sharpe Ratio and average turnover for the considered strategies. It follows from the reported results that the bagged pretest estimator has the lowest turnover and is therefore suggested to be the preferred way of automatizing portfolio allocation. Furthermore, whenever the underlying strategies are heterogeneous enough, in the sense that there are less profitable & less risky allocations versus more profitable & more risky strategies, combining them in the proposed data driven way yields better performance compared to the underlying strategies. Whenever there is a clear dominating strategy, i.e. more profitable & less risky, the bagged pretest estimator would chose this strategy independent of portfolio size and parameter constellations.

4 Conclusions

This paper introduces a novel pretest estimator with a data-driven and time adaptive significance level choice. Our pretest estimator is designed to choose an optimal portfolio allocation strategy among M alternatives and a fixed benchmark strategy. The natural candidates for the benchmark strategy is the buy-and-hold strategy or the equally weighted portfolio, but can be easily adjusted to the problem at hand. Equally weighted portfolio is known as one of the toughest benchmarks in the empirical finance, i.e. it is very difficult for a more sophisticated theory based strategy to outperform the naive $1/N$ in both in-sample and out-of-sample comparisons. However, for certain datasets and parameter constellations it is possible for other strategies to be better than the equally weighted portfolio according to a certain portfolio evaluation criteria. Our proposed pretest estimator uses the weight estimates of the competing strategies and chooses the one, maximizing the predetermined criterion. We have constructed an estimator which results in the highest out-of-sample CE and SR based on the net portfolio returns, the algorithm however can be easily adjusted to other performance measures.

Our estimator takes into account transaction costs and with the help of statistical learning techniques computes the optimal portfolio weights. It first evaluates the pseudo-out-of-sample performance measure and chooses an optimal significance level in a time adaptive way. With bagging it smooths the indicator functions of the pretest estimator, which stabilizes the performance and reduces the turnover costs. We show that our estimator is robust to different portfolio sizes and in-sample estimation window lengths. We also provide evidence that the proposed estimator outperforms the underlying strategies with respect to the out-of-sample Certainty Equivalent and Sharpe Ratio.

Despite its very promising performance of the pretest bagging strategy we see still potential to further improve this approach. For example, one path of further improvement can be the development of more powerful performance tests in the pretest stage. Moreover, the boosting-based approach shows a suboptimal but still very good performance. In future work it needs to be shown whether combining the two statistical learners can yield an improvement in portfolio allocation.

References

Bühlmann, P., B. Yu, et al. (2002): "Analyzing bagging," *The Annals of Statistics*, 30, 927–961.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953.

DeMiguel, V., F. J. Nogales, and R. Uppal (2014): "Stock Return Serial Dependence and Out-of-Sample Portfolio Performance," *Review of Financial Studies*, 27, 1031–1073.

Efron, B. (1992): *Bootstrap methods: another look at the jackknife*, Springer.

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

Jagannathan, R. and T. Ma (2003): "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *The Journal of Finance*, 58, 1651–1684.

Kazak, E. and W. Pohlmeier (2018): "Testing Out-of-Sample Portfolio Performance," Tech. rep., forthcoming in International Journal of Forecasting.

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

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

——— (2004a): "Honey, I Shrunk the Sample Covariance Matrix," *The Journal of Portfolio Management*, 30, 110–119.

——— (2004b): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88, 365–411.

——— (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

——— (2014): "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks," Working paper, Department of Economics, University of Zurich.

Markowitz, H. (1952): "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

Merton, R. C. and P. A. Samuelson (1992): *Continuous-time finance*, Blackwell Boston.

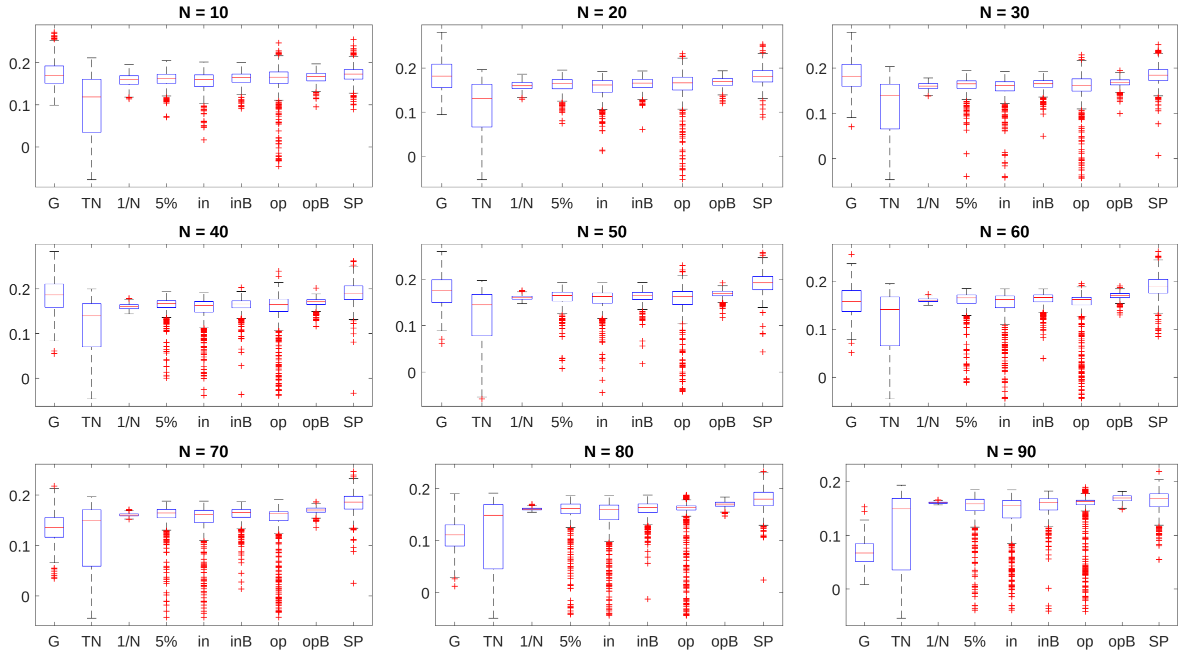
Okhrin, Y. and W. Schmid (2006): "Distributional properties of portfolio weights," *Journal of Econometrics*, 134, 235–256.

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

Shan, K. and Y. Yang (2009): "Combining regression quantile estimators," *Statistica Sinica*, 1171–1191.

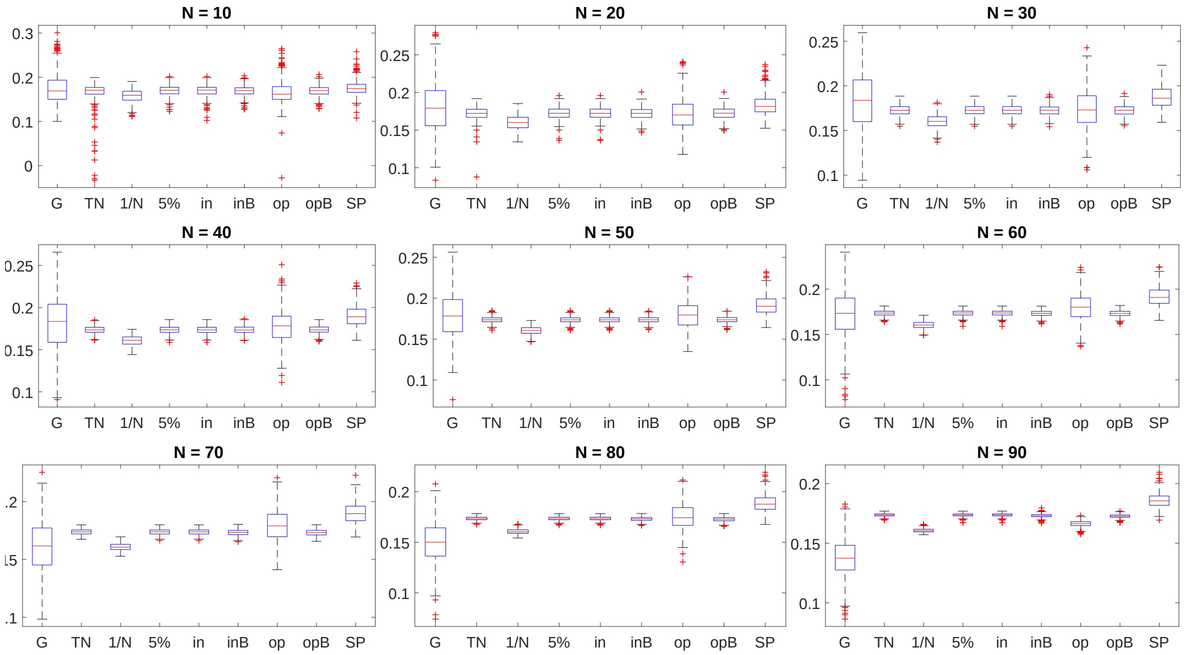
A Appendix

Figure 4: Boxplots of the out-of-sample SR for $T = 120$.



Each boxplot corresponds to the distribution of the out-of-sample SR computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample CE is computed over an out-of-sample evaluation horizon of $H = 500$ observations, risk aversion parameter is $\gamma = 1$, in-sample estimation window length $T = 120$. X-axes denote different ways of computing portfolio weights: G from eq.(2), TN from eq.(3), 1/N from eq.(4), 5% denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, *in* allows for a flexible significance level choice according to eq.(12), *inB* is the bagged version of *in*. *op* and *opB* denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). SP corresponds to eq.(19).

Figure 5: Boxplots of the out-of-sample SR for $T = 180$.



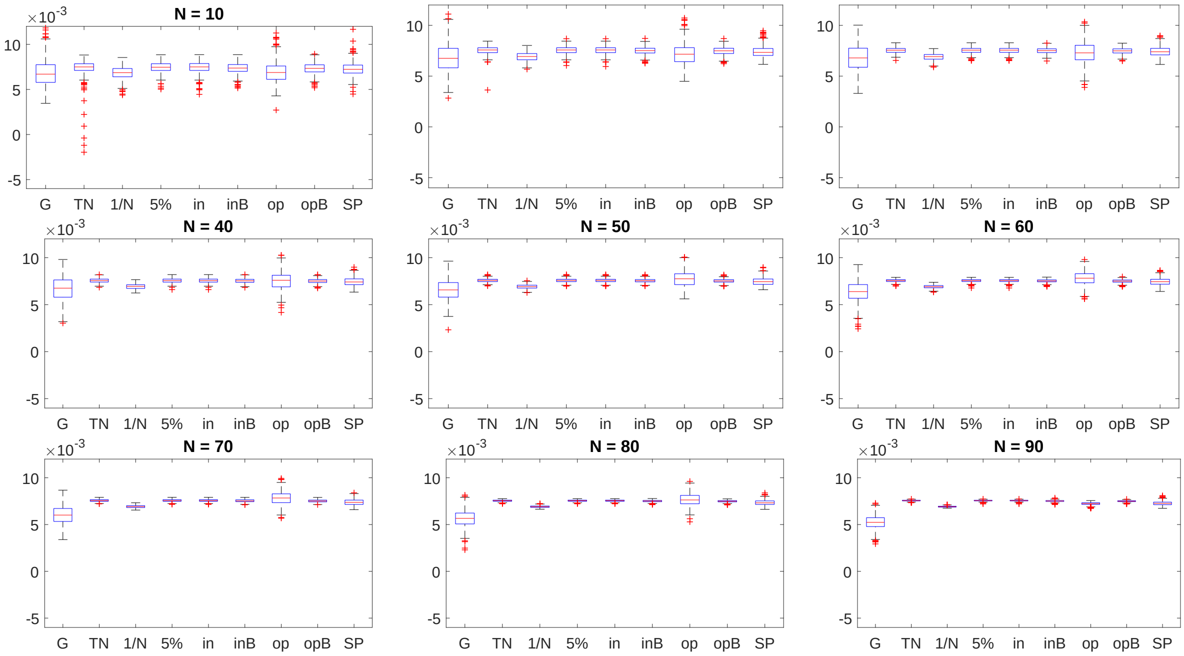
Each boxplot corresponds to the distribution of the out-of-sample SR computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample CE is computed over an out-of-sample evaluation horizon of $H = 500$ observations, risk aversion parameter is $\gamma = 1$, in-sample estimation window length $T = 180$. X-axes denote different ways of computing portfolio weights: G from eq.(2), TN from eq.(3), 1/N from eq.(4), 5% denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, *in* allows for a flexible significance level choice according to eq.(12), *inB* is the bagged version of *in*. *op* and *opB* denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). SP corresponds to eq.(19).

Table 3: Out-of-sample CE for $T = 180$.

	N = 10	N = 20	N = 30	N = 40	N = 50	N = 60	N = 70	N = 80	N = 90
GMVP	0.0069	0.0069	0.0068	0.0067	0.0066	0.0064	0.0060	0.0056	0.0052
Tangency	0.0060	0.0075	0.0075	0.0076	0.0076	0.0076	0.0076	0.0075	0.0076
1/N	0.0068	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069
$\alpha = 5\%$	0.0074	0.0075	0.0075	0.0076	0.0076	0.0076	0.0076	0.0075	0.0076
α^s_{in}	0.0074	0.0075	0.0075	0.0076	0.0076	0.0076	0.0076	0.0075	0.0076
α^s_{in} B	0.0074	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075
α^s_{op}	0.0067	0.0072	0.0073	0.0075	0.0077	0.0078	0.0078	0.0077	0.0072
α^s_{op}	0.0073	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075
Seq. Perf.	0.0073	0.0074	0.0074	0.0075	0.0075	0.0075	0.0074	0.0073	0.0073

Figures in the table correspond to the average out-of-sample CE computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample CE is computed over an out-of-sample evaluation horizon of $H = 500$ observations, risk aversion parameter is $\gamma = 1$, in-sample estimation window length $T = 180$. $\alpha = 5\%$ denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, α^s_{in} allows for a flexible significance level choice according to eq.(12) and α^s_{in} B is the bagged version of α^s_{in}. α^s_{op} and α^s_{op} B denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). Seq. Perf. corresponds to eq.(19).

Figure 6: Boxplots of the out-of-sample CE for $T = 180$.



Each boxplot corresponds to the distribution of the out-of-sample CE computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample CE is computed over an out-of-sample evaluation horizon of $H = 500$ observations, risk aversion parameter is $\gamma = 1$, in-sample estimation window length $T = 180$. X-axes denote different ways of computing portfolio weights: G from eq.(2), TN from eq.(3), 1/N from eq.(4), 5% denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, *in* allows for a flexible significance level choice according to eq.(12), *inB* is the bagged version of *in*. *op* and *opB* denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). SP corresponds to eq.(19).

Table 4: Turnover for $T = 180$.

	N = 10	N = 20	N = 30	N = 40	N = 50	N = 60	N = 70	N = 80	N = 90
GMVP	0.0673	0.1238	0.1733	0.2227	0.2718	0.3213	0.3757	0.4334	0.4995
Tangency	0.0529	0.0232	0.0225	0.0224	0.0225	0.0226	0.0226	0.0226	0.0226
1/N	0.0218	0.0223	0.0225	0.0225	0.0226	0.0226	0.0226	0.0227	0.0227
$\alpha = 5\%$	0.0226	0.0226	0.0227	0.0226	0.0226	0.0227	0.0226	0.0227	0.0227
α^s_{in}	0.0235	0.0226	0.0227	0.0226	0.0226	0.0227	0.0226	0.0227	0.0227
α^s_{in} B	0.0220	0.0225	0.0228	0.0228	0.0230	0.0231	0.0231	0.0232	0.0235
α^s_{op}	0.0524	0.0626	0.0750	0.0847	0.0877	0.0869	0.0786	0.0574	0.0227
α^s_{op}	0.0220	0.0226	0.0228	0.0228	0.0229	0.0230	0.0230	0.0230	0.0231
Seq. Perf.	0.0244	0.0316	0.0396	0.0476	0.0557	0.0627	0.0675	0.0733	0.0758

Figures in the table correspond to the average turnover over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the average turnover is computed over an out-of-sample evaluation horizon of $H = 500$ observations, in-sample estimation window length $T = 180$. $\alpha = 5\%$ denotes the pretest estimator based on the in-sample test statistic as in eq.(9) with a fixed significance level of 5%, α^s_{in} allows for a flexible significance level choice according to eq.(12) and α^s_{in} B is the bagged version of α^s_{in}. α^s_{op} and α^s_{op} B denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). Seq. Perf. corresponds to eq.(19).

Table 5: Out-of-sample SR for $T = 251$ and daily data.

	N = 10	N = 20	N = 30	N = 40	N = 50	N = 60	N = 70	N = 80	N = 90
GMVP	0.0034	-0.0112	-0.0297	-0.0498	-0.0738	-0.1016	-0.1312	-0.1615	-0.1967
λ_1	0.0066	-0.0019	-0.0111	-0.0209	-0.0322	-0.0440	-0.0576	-0.0673	-0.0810
λ_2	0.0140	0.0149	0.0150	0.0157	0.0150	0.0152	0.0153	0.0152	0.0154
$\alpha = 5\%$	0.0117	0.0142	0.0148	0.0156	0.0151	0.0152	0.0153	0.0152	0.0154
α^s_{op}	-0.0015	0.0005	0.0032	0.0083	0.0108	0.0132	0.0149	0.0151	0.0154
α^s_{op}	0.0140	0.0149	0.0151	0.0157	0.0151	0.0152	0.0153	0.0152	0.0154
Seq. Perf.	0.0092	0.0123	0.0142	0.0153	0.0147	0.0148	0.0148	0.0145	0.0145

Figures in the table correspond to the average out-of-sample SR computed on net portfolio returns over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the out-of-sample SR is computed over an out-of-sample evaluation horizon of $H = 503$ observations, in-sample estimation window length $T = 251$. Table rows correspond to different ways of computing portfolio weights: G from eq. (2), λ_1 for GMVP with Ledoit and Wolf (2004b) shrinkage, $\lambda_2 = 0.05 \cdot N$ for Shrunken GMVP similar to eq.(3), 5% denotes the pretest estimator based on the out-of-sample test statistic as in with a fixed significance level of 5%, *op* and *opB* denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). SP corresponds to eq.(19).

Table 6: Turnover for $T = 251$ and daily data.

	N = 10	N = 20	N = 30	N = 40	N = 50	N = 60	N = 70	N = 80	N = 90
GMVP	0.0341	0.0701	0.1112	0.1560	0.2042	0.2595	0.3173	0.3834	0.4546
λ_1	0.0279	0.0527	0.0783	0.1041	0.1291	0.1557	0.1812	0.2053	0.2305
λ_2	0.0085	0.0087	0.0087	0.0087	0.0088	0.0087	0.0087	0.0087	0.0087
$\alpha = 5\%$	0.0123	0.0097	0.0090	0.0088	0.0088	0.0087	0.0087	0.0087	0.0087
α^s_{op}	0.0383	0.0330	0.0280	0.0213	0.0162	0.0122	0.0095	0.0089	0.0087
α^s_{op}	0.0085	0.0087	0.0087	0.0087	0.0088	0.0087	0.0087	0.0087	0.0087
Seq. Perf.	0.0149	0.0144	0.0117	0.0110	0.0109	0.0110	0.0112	0.0113	0.0114

Figures in the table correspond to the average turnover over a 500 randomly drawn portfolios of size N. For each randomly drawn portfolio the average turnover is computed over an out-of-sample evaluation horizon of $H = 503$ observations, in-sample estimation window length $T = 251$. Table rows correspond to different ways of computing portfolio weights: G from eq. (2), λ_1 for GMVP with Ledoit and Wolf (2004b) shrinkage, $\lambda_2 = 0.05 \cdot N$ for Shrunken GMVP similar to eq.(3), 5% denotes the pretest estimator based on the out-of-sample test statistic as in with a fixed significance level of 5%, *op* and *opB* denote the out-of-sample significance level choice and the same pretest estimator combined with bagging as in eq.(17). SP corresponds to eq.(19).