Improving portfolio performance using recent and new techniques

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Econometrics and Management Science

Name student: Menno Boor
Student ID number: 456065

Supervisor: dr. W. Wang
Second assessor: dr. O. Kleen

Date final version: July 7, 2021

Abstract

We investigate whether recent and new techniques are able to improve portfolio performance. We hereby develop new methods to reduce the variance of estimated portfolio weights that can also be used in a setup that is able to capture empirical properties of asset returns. The first method is a modified version of a trimming method of Radchenko et al. (2020) and estimates mean-variance portfolio weights while restricting the sum of absolute weights. The second method is a simulation-based shrinkage rule which shrinks the estimated weights towards equal weights when the sum of the sample variances of the estimated weights is high compared with previous periods. We show that these methods are indeed effective; especially our modified trimming rule performs well and it significantly outperforms the $1/N$ strategy in some of the data sets whereas it is never significantly outperformed.

Contents

1 Introduction

The performance of an investing portfolio is often determined by its Sharpe ratio, defined as the expected excess portfolio return divided by the standard deviation of the excess portfolio returns. Theoretically, we can use the framework of Markowitz (1952) to determine the optimal portfolio weights, known as the mean-variance allocation. However, the portfolio weights in the mean-variance allocation are a function of the theoretical mean vector and covariance matrix of the returns, which are unknown, and hence these weights need to be estimated. Among others Kan and Zhou (2007) and DeMiguel et al. (2009) show that estimating these weights often leads to large estimation errors, which results in low Sharpe ratios out-of-sample. DeMiguel et al. (2009) additionally report many different methods that can reduce estimation error by reducing the variance of the estimated weights, but show that none of these methods systematically outperforms the $1/N$ portfolio. However, these methods all assume implicitly that the means and variances of the asset returns are the same in the out-of-sample period as in the estimation period, whereas empirical studies of Cont (2001) and Stoyanov et al. (2011) show that this is not observed in practice. Low et al. (2016) construct portfolios in a way that incorporates changes of the means and variances along with other empirical properties. They show that their method improves the performance relative to the mean-variance portfolio that uses sample moments to estimate the portfolio weights, but they also find that their portfolios are not able to beat the $1/N$ portfolio after correcting for transaction costs. We therefore investigate new methods in this paper that can both specifically model empirical properties of asset returns and reduce the variance of the estimated weights in order to obtain a better portfolio performance.

This research is consequently relevant from both an academic and a practical point of view. It namely provides new ways to handle estimation error, which is a well-known problem in the portfolio optimisation literature. From a practical point of view this research is relevant for investors, as it provides new ways to get better expected returns for the amount of risk they are willing to take. Furthermore, this research also analyses the performance in the presence of transaction costs, thereby making the methods applicable for investing in practice.

In this paper we investigate whether recent and new techniques are able to improve portfolio performance. To answer this research question we explore three main approaches that have the potential to improve the performance of the mean-variance allocation. We first investigate whether we can improve the portfolio performance by specifically modelling the empirical properties of asset returns, because this method has the potential to reduce the bias of the estimated weights. We then evaluate whether we can improve portfolio performance by trimming large (negative) weights, as DeMiguel et al. (2009) show that methods that constrain portfolio weights perform relatively good and Radchenko et al. (2020) show that trimming portfolio weights performs exceptionally well for combining forecasts. We finally explore whether shrinking portfolio weights towards equal weights when the variance of the estimated weights is high can improve the portfolio performance, as a high variance of the estimated weights decreases the out-of-sample performance and DeMiguel et al. (2009) show that the $1/N$ portfolio is hard to outperform.

In order to answer these questions we make use of the same empirical data sets as used in DeMiguel et al. (2009) because many different portfolio strategies have already been evaluated on these data sets. To specifically incorporate empirical properties of asset returns reported by Cont (2001) and Stoyanov et al. (2011) we first estimate a dynamic model for the individual asset returns using a similar setup as Low et al. (2016) that allows for volatility clustering, negative correlation of returns with volatility and both skewness and excess kurtosis of the return distribution. We then estimate the dependence structure of the different asset returns with copula models; we hereby deviate from the setup of Low et al. (2016) in order to explore more flexible copula models, as studies of Okhrin et al. (2013), Longin and Solnik (2001) and Ang and Chen (2002) show that asset returns have a complex dependence structure. We subsequently use our estimated models to simulate asset returns that take empirical properties of asset returns into account. These simulated returns are then used to estimate the portfolio weights in the coming period.

To evaluate whether trimming large (negative) weights can improve portfolio performance we modify the methods in Radchenko et al. (2020), who show that in the context of forecast combinations trimming of negative weights can lead to better performance than taking an equally weighted combination of forecasts. We modify these methods so that they can both take the expected portfolio returns into account and are able to trim positive portfolio weights if the sum of the portfolio weights is negative. This way these methods are applicable in our portfolio optimisation setup.

We lastly develop a new method that can be used to reduce the variance of the estimated portfolio weights in a setup that uses simulation of asset returns in order to incorporate empirical properties of asset returns into the estimation problem. This method can be used for any model that simulates asset returns and it shrinks the estimated portfolio weights towards equal weights if the sum of the variances of the estimated weights is high compared with previous periods. This method differs from other shrinkage methods in the literature as it specifically shrinks the estimated weights of the assets (instead of shrinking the sample moments used to estimate these weights) and it does so in a way that does not rely on the normality assumption (or any other assumption) for the distribution of the asset returns. We finally compare the three aforementioned methods to investigate their usefulness in isolation and in combination with each other.

We find that some of the recent and new techniques are able to improve portfolio performance. Although dynamic modelling of specific asset return properties is not able to improve performance as a result of an increase in the variance of the estimated weights, we find that our methods that specifically reduce the variance of the estimated weights are able to do so. Especially our modified trimming rule is beneficial and even significantly outperforms the $1/N$ benchmark in some data sets whereas it is never significantly outperformed. Moreover, these results also hold when transaction costs are taken into account.

Our research therefore extends the current literature on portfolio optimisation, because we develop new methods to reduce the estimation error and we show that these methods are effective. Our modified trimming method works both for trimming weights in a mean-variance portfolio and for trimming positive weights when the sum of the portfolio weights is negative (indicating that we have a short position in the portfolio), whereas previous trimming methods are only applied to a minimum variance portfolio with a long position in the portfolio. Moreover, our new simulation-based shrinkage method is a robust and flexible method that be used in any portfolio setting that simulates returns irrespective of the distribution of the simulated returns.

The rest of this paper is organised as follows. We describe in Section 2 which data we use to evaluate our methods. We continue in Section 3 with a description of how portfolios are constructed and evaluated and how their performance can be improved. After laying out these foundations, we describe in the subsequent sections in detail which methods we use in our research; Section 4 discusses the estimation of weights using dynamic models, Section 5 describes how trimming portfolio weights works and Section 6 explains our simulation-based shrinkage rule. We then describe how these methods can be combined in Section 7 after which we discuss the results of applying these methods in Section 8. We finally conclude our research in Section 9.

2 Data

To evaluate our portfolio strategies, we use the same data sets of excess asset returns as DeMiguel et al. (2009) because many different portfolio strategies have already been evaluated on these data sets. These data sets can be downloaded from the website of Lorenzo Garlappi [1]. A clear overview of the seven empirical data sets that are used by DeMiguel et al. (2009) can be found in Table 2 of their paper and a more elaborate description of these data sets can be found in Appendix A of their paper.

In short, the first data set contains ten sector portfolios of the S&P 500 and the US equity market portfolio over a period from January 1981 until December 2002 (264 observations). The second data set contains ten industry portfolios and the US equity market portfolio from July 1926 until November 2004 (941 observations). The third data set contains eight country indexes and the World Index from January 1970 until July 2001 (379 observations). The fourth data set contains the SMB and HML portfolios as well as the US equity market portfolio from July 1926 until November 2004 (941 observations). The fifth, sixth and seventh data set then all contain twenty size- and book-to-market portfolios and are augmented with a US equity market portfolio (fifth data set), a US equity market portfolio, an SMB and an HML portfolio (sixth data set), or a US equity market portfolio, an SMB, an HML and a UMD (momentum) factor portfolio (seventh data set), all containing observations from July 1926 until November 2004 (941 in total). We therefore observe that the seven empirical data sets have quite some overlap. Although the total number of return series in our data sets is equal to 102, the total number of unique return series only equals 54 (where we treat the US equity market returns in the first data set as different from the US equity market returns in the other data sets as they are taken at different dates in each month, so that the resulting returns are different).

From five out of seven of the data sets (namely the ones that have observations starting in 1926), only a part of these data sets is used by DeMiguel et al. (2009); they only use the observations in these data sets from July 1963 until November 2004 (497 observations in total). We therefore also use the observations from July 1963 until November 2004 to be able to compare our results with their results. However, we use the earlier observations from January 1927 until June 1963 (438 in total, because we omit the first observations in 1926 as then the momentum factor is not available) as ‘experimental data sets’ to evaluate certain methods or model choices. We do this in order to limit the number of different methods that we compare (in combination with other methods) on the same data as used by DeMiguel et al. (2009).

[1] Source: https://www.dropbox.com/sh/vsc9r9i5rts1giw/AAAKe8LX-vewGjXYQJq–xQYa?dl=0

3 Portfolio optimisation framework

In this section we describe the portfolio optimisation framework based on the setup of DeMiguel et al. (2009). In Section 3.1 we describe how portfolios are constructed and in Section 3.2 we explain how portfolios are evaluated. We then discuss in Section 3.3 how portfolio performance can be improved.

3.1 Portfolio construction

We consider the construction of a portfolio of N risky assets, whose returns in month t (in excess of the risk-free rate) are given by the $N \times 1$ vector $\boldsymbol{r}_t$. We denote the $N \times 1$ vector of expected excess returns in month t by $\boldsymbol{\mu}_t$ and the $N \times N$ covariance matrix of the excess returns in month t by $\boldsymbol{\Sigma}_t$. Following the setup of DeMiguel et al. (2009), we let $\boldsymbol{x}_t$ denote the vector of portfolio weights invested in the N risky assets at the beginning of month t, leaving a fraction of $1 - \boldsymbol{\iota}'\boldsymbol{x}_t$ to be invested in the risk-free asset, where $\boldsymbol{\iota}$ denotes an $N \times 1$ vector of ones.

To obtain the portfolio weights that maximise the expected utility of a mean-variance investor, we solve the problem

$$\max_{\boldsymbol{x}_t} \; \boldsymbol{x}_t'\boldsymbol{\mu}_t - \frac{\gamma}{2}\boldsymbol{x}_t'\boldsymbol{\Sigma}_t\boldsymbol{x}_t, \tag{1}$$

where the parameter $\gamma > 0$ denotes the investor's risk aversion. Solving the first-order condition yields the solution $\boldsymbol{x}_t = \frac{1}{\gamma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t$. Similar to DeMiguel et al. (2009), we focus on the relative portfolio weights in a portfolio with only risky assets. These relative weights at time t are obtained as

$$\boldsymbol{w}_t = \frac{\boldsymbol{x}_t}{|\boldsymbol{\iota}'\boldsymbol{x}_t|} = \frac{\frac{1}{\gamma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t}{|\boldsymbol{\iota}'\frac{1}{\gamma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t|} = \frac{\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t}{|\boldsymbol{\iota}'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t|}, \tag{2}$$

where we used that $\gamma > 0$ by assumption. The risky portfolio weights are scaled by their absolute sum in order to ensure that the investor goes long or short in each asset in the same proportion as is optimal in a portfolio that also includes a risk-free asset. It is therefore possible that the relative weights sum to -1, indicating a short position in the portfolio of risky assets.

Equation (2) shows the optimal relative portfolio weights as a function of the mean vector $\boldsymbol{\mu}_t$ and the inverse covariance matrix $\boldsymbol{\Sigma}_t^{-1}$. However, both are unknown in practice and therefore have to be estimated using past return observations. We denote the number of time series observations in each data set by T and the number of time series observations in our rolling estimation window by M, where we take $M = 120$ like DeMiguel et al. (2009). We then estimate the relative portfolio weights $\hat{\boldsymbol{w}}_{t^*}$ for time $t^* \in \{M+1, ..., T\}$ based on the previous M excess returns.

A well-known approach is to estimate the mean vector and covariance matrix with their unbiased sample versions and to plug them into Equation (2). This gives

$$\hat{\boldsymbol{w}}_{t^*} = \frac{\hat{\boldsymbol{\Sigma}}_{t^*}^{-1}\hat{\boldsymbol{\mu}}_{t^*}}{\left|\boldsymbol{\iota}'\hat{\boldsymbol{\Sigma}}_{t^*}^{-1}\hat{\boldsymbol{\mu}}_{t^*}\right|},$$

where we take

$$\hat{\boldsymbol{\mu}}_{t^*} = \frac{1}{M}\sum_{t=t^*-M}^{t^*-1}\boldsymbol{r}_t,$$

and

$$\hat{\boldsymbol{\Sigma}}_{t^*} = \frac{1}{M-1}\sum_{t=t^*-M}^{t^*-1}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_{t^*})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_{t^*})'.$$

It is also possible to estimate these weights in different ways, as we discuss in Sections 4, 5 and 6.

3.2 Evaluating portfolio performance

In order to evaluate the constructed portfolio of assets we use the out-of-sample Sharpe ratio as is done in DeMiguel et al. (2009). The out-of-sample Sharpe ratio of strategy k is given by

$$\widehat{\text{SR}}_k = \frac{\bar{r}_{k,t}}{\sqrt{\frac{1}{T-M-1}\sum_{t=M+1}^{T}(r_{k,t} - \bar{r}_{k,t})^2}},$$

where

$$\bar{r}_{k,t} = \frac{1}{T-M}\sum_{t=M+1}^{T} r_{k,t},$$

with $r_{k,t} = \hat{\boldsymbol{w}}'_{k,t}\boldsymbol{r}_t = \sum_{j=1}^{N}\hat{w}_{k,j,t}r_{j,t}$ the resulting excess return of strategy k in month t. As we additionally want to evaluate the usefulness of our portfolio strategies for application in practice, we also compute the Sharpe ratios of the different portfolio strategies in the presence of transaction costs. We hereby make use of the setup of DeMiguel et al. (2009), who make use of proportional transaction costs (denoted by a) of 50 basis points per transaction, which they base on several studies of the transaction costs of individual stocks that trade on the NYSE. Following their notation, we let $\hat{w}_{k,j,t^+}$ denote the relative portfolio weight of strategy k invested in asset j at the end of month t, so just before rebalancing which happens at the begin of month $t+1$. We can compute it, for both $\boldsymbol{\iota}'\hat{\boldsymbol{w}}_{k,t} = 1$ and $\boldsymbol{\iota}'\hat{\boldsymbol{w}}_{k,t} = -1$, as

$$\hat{w}_{k,j,t^+} = \frac{\hat{w}_{k,j,t}(1 + r_{j,t})}{\left|\sum_{l=1}^{N}\hat{w}_{k,l,t}(1 + r_{l,t})\right|}.$$

Rebalancing the portfolio at the begin of month $t+1$ then gives rise to transaction costs (expressed as a fraction of total wealth at the end of month t) of $a \cdot \sum_{j=1}^{N} |\hat{w}_{k,j,t+1} - \hat{w}_{k,j,t^+}|$. DeMiguel et al. (2009) then compute the evolution of wealth of strategy k at the beginning of each month as

$$W_{k,t+1} = W_{k,t}(1 + r_{k,t})\left(1 - a \cdot \sum_{j=1}^{N} \left|\hat{w}_{k,j,t+1} - \hat{w}_{k,j,t^+}\right|\right), \tag{3}$$

so that the return net of transaction costs can consequently be computed as

$$r_{k,t}^{\text{net}} = \frac{W_{k,t+1}}{W_{k,t}} - 1. \tag{4}$$

However, our method of computing returns net of transaction costs deviates (slightly) from the method above, as computing the wealth net of transaction costs using Equation (3) works well under reasonable allocations and assumptions, but possibly produces incorrect results in extreme cases. In particular, this method can fail if the gross returns $1 + r_{k,t}$ are negative and the transaction costs $a \cdot \sum_{j=1}^{N} |\hat{w}_{k,j,t+1} - \hat{w}_{k,j,t^+}|$ are higher than 100% of the wealth of the investor, which can happen in some of the data sets for the methods that take extreme positions due to large estimation errors. To be precise, the latter happens if the turnover from rebalancing at the beginning of month $t+1$ is larger than $\frac{1}{a} = \frac{1}{\text{50 basis points}} = 200$. In the case of negative gross returns and transaction costs that exceed 100%, you incur enormous losses and have to pay very high transaction costs, but in this formula the two negative numbers cancel each other so that the resulting wealth at time $t+1$ is positive. Moreover, if gross returns are negative enough and transaction costs are high enough, the wealth at time $t+1$ obtained with this formula even exceeds the wealth at time t, leading to a (very large) positive return. To prevent this from happening, we check for each estimated vector of portfolio weights whether the resulting gross return is still positive and whether the transaction costs are smaller than 100%. If one of these two conditions is violated at some point in time for a given strategy in a given data set, we will not report the Sharpe ratio in the presence of transaction costs for that portfolio strategy in that data set as it is not reliable.

3.3 Improving portfolio performance

DeMiguel et al. (2009) show that optimal diversification is very beneficial in-sample, whereas out-of-sample the gains from optimal diversification are more than offset by estimation error. We therefore need to reduce estimation error in order to improve the out-of-sample performance. We can measure estimation error using the mean squared error (MSE) of the estimated vector of portfolio weights.

We define the MSE of the estimated weight vector $\hat{\boldsymbol{w}}_{t^*}$ at time t^* as the sum of the MSEs of each individual weight. That is,

$$\begin{aligned}\text{MSE}(\hat{\boldsymbol{w}}_{t^*}) &= \text{E}\big[(\hat{\boldsymbol{w}}_{t^*} - \boldsymbol{w}_{t^*})'(\hat{\boldsymbol{w}}_{t^*} - \boldsymbol{w}_{t^*})\big] = \text{E}\left[\sum_{j=1}^{N}(\hat{w}_{j,t^*} - w_{j,t^*})^2\right] \\ &= \sum_{j=1}^{N}\text{E}\big[(\hat{w}_{j,t^*} - w_{j,t^*})^2\big] = \sum_{j=1}^{N}\text{MSE}(\hat{w}_{j,t^*}).\end{aligned} \tag{5}$$

The MSE of the individual estimated weights can be decomposed as the sum of the squared bias and the variance, so that we obtain from Equation (5) that

$$\text{MSE}(\hat{\boldsymbol{w}}_{t^*}) = \sum_{j=1}^{N}\text{MSE}(\hat{w}_{j,t^*}) = \sum_{j=1}^{N}\left(\text{E}\big[(\hat{w}_{j,t^*} - w_{j,t^*})\big]^2 + \text{E}\Big[\big(\hat{w}_{j,t^*} - \text{E}[\hat{w}_{j,t^*}]\big)^2\Big]\right). \tag{6}$$

We see from Equation (6) that we can decrease estimation error by reducing the variance or reducing the (squared) bias of the estimated weights. If the reduction in variance (respectively squared bias) outweighs the corresponding increase in squared bias (respectively variance), then the MSE of the estimated weights decreases. This in turn leads to an estimated portfolio that is closer to the true (but unknown) optimal mean-variance portfolio that achieves the highest Sharpe ratio.

DeMiguel et al. (2009) consider different portfolio allocations that aim to reduce estimation error. These methods impose constraints (either on the estimated mean and covariance matrix or directly on the estimated weights) or they shrink the sample mean towards a certain target during the estimation process. Other papers, such as Ledoit and Wolf (2003) and Ledoit and Wolf (2004), also develop methods that aim to reduce the estimation error by shrinking the estimated covariance matrix; these methods are not separately evaluated by DeMiguel et al. (2009) as the Jagannathan and Ma (2003) explain that imposing short sale constraints in a minimum variance portfolio is equivalent to shrinking the elements in the covariance matrix and yields a similar performance. All these methods intend to reduce the variance of the estimated weights (at the cost of an increase in squared bias) in order to reduce estimation error. We describe the methods that we use for reducing estimation error of the estimated weights in Sections 4, 5 and 6.

4 Estimating weights based on dynamic forecasting using copulas

To obtain a good portfolio performance in the out-of-sample period, we need good estimates of the sample moments in the out-of-sample period. As mentioned in Section 3.1, it is a well-known approach to obtain the estimated portfolio weights by plugging the unbiased sample estimates of the mean vector and covariance matrix of the excess returns into Equation (2). Such an approach implicitly assumes that the means and variances of each return series will be the same in the next month as they were in the previous M months that are used to estimate them. However, empirical studies of Cont (2001) and Stoyanov et al. (2011) indicate that the volatility of asset returns is not constant over time, so that this approach is likely to produce biased portfolio weights. We therefore use a dynamic model that captures changes in return volatility in order to decrease the bias of the estimated portfolio weights. This method differs from the portfolio methods in DeMiguel et al. (2009) as it aims to reduce estimation error by decreasing the (squared) bias of the estimated weights as opposed to decreasing the variance of the estimated weights. We describe in Section 4.1 and 4.2 how we model the joint distribution of all asset returns. We then explain in Section 4.3 how the parameters in our models can be estimated and in Section 4.4 how we can subsequently obtain the resulting portfolio weights.

4.1 Modelling the dependence structure of the returns

To model the joint distribution of all asset returns in a particular data set we make use of copulas, as this provides a very flexible framework for generating multivariate models. An n-dimensional copula is a distribution function on $[0,1]^n$ with standard uniform marginal distributions. Sklar (1973) proves that copulas can be used to link marginal distributions to a joint distribution. In particular, if F is an n-dimensional distribution function with marginal distributions $F_1, ..., F_n$, then there exists an n-dimensional copula C such that

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)), \tag{7}$$

for all real-valued numbers $\{x_1, ..., x_n\}$. Since $F_j(X_j) \sim U(0,1)$ when the random variable X_j has marginal distribution function F_j, we can first model the marginal distributions of the random variables X_j for $j \in \{1, ..., n\}$, and subsequently model the distribution of $(U_1, ..., U_n)$ where $U_j = F_j(X_j)$ using a copula, which by Equation (7) is equivalent to modelling the joint distribution of $(X_1, ..., X_n)$.

To model the dependence structure of the N asset returns in each data set, we use a method that is able to model a complex dependence structure. This is necessary because Okhrin et al. (2013) mention that it is rarely a feasible assumption in practical applications to assume that all asset return pairs have the same dependence structure and because Longin and Solnik (2001) and Ang and Chen (2002) show that the correlations of equity returns are greater for low returns than for high returns, known as asymmetric tail dependence. To incorporate such a complex dependence structure we make use of vines. Bedford and Cooke (2001) introduce regular vines (R-vines) as graphical models that can be used to model dependent variables and Aas et al. (2009) use two special cases of R-vines to construct a multivariate copula density from pairwise copulas that act on several different conditional probability distributions. As there exist many different pair copula constructions for high-dimensional distributions (Morales Napoles et al. (2010) show that the number of ways in which the N-variate copula density can be constructed is equal to $\frac{N!}{2} \cdot 2^{\binom{N-2}{2}}$), we need a way to organise these constructions, which can be done by vines. We use the approach described by Dißmann et al. (2013) to build high-dimensional models from pair copulas using R-vines, as it is an automated technique for model selection and parameter estimation that allows us to capture a rich pattern of dependencies between asset returns. In particular, each pairwise copula in this approach can be chosen (using the Akaike Information Criterion (AIC) as selection criterion) from a large set of copulas so that we can capture asymmetric tail dependence. A detailed explanation of how modelling with R-vines works can be found in Appendix A.

We evaluate four different copula models for the dependence structure of the asset returns. The first copula model is an R-vine copula with unrestricted bivariate copulas (meaning that we select bivariate copulas from a wide range of copula families using the AIC), corresponding with the setup in Dißmann et al. (2013). Our second copula model is an R-vine copula with normal bivariate copulas; we use this model as Low et al. (2016) use a normal copula model combined with modelling (marginal) asymmetries of asset returns and they show that this is beneficial for portfolio construction. The third copula model is a canonical vine (C-vine) copula with unrestricted bivariate copulas. Aas et al. (2009) explain that a C-vine is a special type of R-vine that can be useful when a particular variable is known to be a key variable that governs interactions in a data set. Our last copula model is a C-vine copula with Clayton bivariate copulas, where we also allow possible rotations of the Clayton copula in order to generate a richer dependence structure. We include this model as Low et al. (2013) make use of a Clayton C-vine copula for modelling returns and they show that this is beneficial for portfolio construction with ten assets or more compared to using a multivariate Clayton copula. We use the experimental data sets (containing earlier observations for five out of our seven data sets) to choose which method we subsequently use to model the dependence structure in the regular data sets. We do this in combination with choosing the setup for modelling the marginal returns, as explained at the end of Section 4.2.

4.2 Modelling marginal returns

To model the marginal distribution of each asset return series we use a model that is able to capture changes in volatility in order to reduce the bias of estimated portfolio weights. In addition to modelling time-varying volatility we also model other empirical properties of asset returns, which is motivated by the findings of Fantazzini (2009). He uses different models for returns that exhibit empirical properties of asset returns (by construction) and shows that misspecified marginal models that do not capture these properties lead to substantial increases in bias of estimates of the correlation between different assets (compared with correctly specified marginal models). Using marginal models that do not allow us to capture these empirical properties would therefore be detrimental for reducing the (squared) bias of the estimated portfolio weights.

Stoyanov et al. (2011) conclude based on an extensive body of empirical research that asset returns can be characterised by several stylised facts and they confirm that a realistic model for asset returns should take these stylised facts into account. Furthermore, Stoyanov et al. (2011) advise to use an extended model that can capture the stylised facts and contains the normal distribution as a special case. The stylised facts reported in their paper include the clustering of return volatility, the autoregressive behaviour of the returns, the skewness of returns and the fat-tails (compared with a normal distribution) of the return distribution. An earlier empirical study of Cont (2001) also investigates the empirical properties of asset returns that are common to a large set of assets and markets and his stylised facts agree with the stylised facts of Stoyanov et al. (2011) except for the autoregressive behaviour of returns. Additionally, Cont (2001) also reports conditionally heavy tails and a leverage effect as stylised facts, indicating that even after correcting returns for volatility clustering the residual time series still exhibits heavy tails and that the volatility of the asset returns is negatively correlated with the (level of the) asset returns.

In order to take these stylised facts into account but to have one setup that allows for autoregressive behaviour of asset returns and another setup that does not, we consider two setups. The first setup is similar to the one used in Low et al. (2016). This setup allows for autoregressive behaviour of asset returns by modelling the asset returns as an AR(2) process, as Low et al. (2016) mention that this is shown to have a suitably parsimonious fit for US stock returns. To model the volatility clustering of returns and to incorporate the leverage effect, we model the conditional variance of the stock returns with a GARCH-GJR(1,1) model. We further allow for (negative) skewness and excess kurtosis of the returns (even after correcting for volatility clustering) by modelling the error terms that appear in both the mean equation and the conditional variance equation with a standardised skewed Student t distribution. We thus have for each asset $j \in \{1, ..., N\}$

$$r_{j,t} = \mu_j + \phi_{j,1} \cdot (r_{j,t-1} - \mu_j) + \phi_{j,2} \cdot (r_{j,t-2} - \mu_j) + \varepsilon_{j,t}, \tag{8}$$

$$\varepsilon_{j,t} = \sigma_{j,t} \cdot z_{j,t}, \tag{9}$$

$$\sigma^2_{j,t} = \omega_j + (\alpha_j + \gamma_j \cdot I_{(-\infty,0)}(\varepsilon_{j,t-1})) \cdot \varepsilon^2_{j,t-1} + \beta_j \cdot \sigma^2_{j,t-1}, \tag{10}$$

$$z_{j,t} \sim \text{standardised skewed Student } t(\nu_j, \lambda_j), \tag{11}$$

with $I_{(-\infty,0)}(\varepsilon_{j,t-1})$ denoting an indicator function that equals 1 if $\varepsilon_{j,t-1} < 0$ and 0 otherwise. The parameter $\lambda_j > 0$ determines how skewed the distribution of $z_{j,t}$ is. In particular, for $\lambda_j = 1$ we have a symmetric distribution, for $\lambda_j \in (0, 1)$ we have a negatively skewed distribution and for $\lambda_j > 1$ we have a positively skewed distribution. Furthermore, our setup contains the normal distribution as a special case because for $\lambda_j = 1$ and $\nu_j \to \infty$ we have a normal distribution, and for $\phi_{j,1} = \phi_{j,2} = \alpha_j = \beta_j = \gamma_j = 0$ it even has a constant mean and variance. The functional form of the density of the standardised skewed Student t distribution is stated and derived in Appendix B.

The second setup is similar to the first setup, with as only difference that we do not allow autoregressive behaviour of asset returns. In particular, Equation (8) is replaced by

$$r_{j,t} = \mu_j + \varepsilon_{j,t}, \tag{12}$$

and Equations (9)-(11) are the same for the second setup. To determine which setup for modelling the marginal returns is more appropriate (in combination with a given method for modelling the dependence structure of the returns), we evaluate the eight different combinations described in Table 1 using the data in our experimental data sets. We evaluate these models using the test described in Vuong (1989) to compare the log-likelihoods of non-nested models (that can be corrected for the number of parameters) and we compare the resulting portfolio performance (using the Sharpe ratio) of using a given model. In particular, we apply the Vuong test for each month in the out-of-sample period (319 in total) and take for each model pair the average of these 319 test statistics. Using the central limit theorem (CLT) it follows that

$$\frac{\bar{V}_t - \mathrm{E}[V_t]}{\sqrt{\mathrm{var}(V_t)/319}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\bar{V}_t = \frac{V_1 + \cdots + V_{319}}{319}$ with V_t denoting the Vuong test statistic for month t. We then obtain the p-values under the null hypothesis that $\mathrm{E}[V_t] = 0$ by standardising the mean Vuong test statistic using its estimated sample variance.

Table 1: Different dynamic model setups

model number	allow for autoregressive behaviour of marginal returns?	copula model
1	yes	(i)
2	yes	(ii)
3	yes	(iii)
4	yes	(iv)
5	no	(i)
6	no	(ii)
7	no	(iii)
8	no	(iv)

Note: Copula model (i) is an R-vine copula with unrestricted bivariate copulas (meaning that we can choose our bivariate copulas from a wide range of copula families). Copula model (ii) is an R-vine copula with normal bivariate copulas. Copula model (iii) is a C-vine copula with unrestricted bivariate copulas and copula model (iv) is a C-vine copula with (possibly rotated) Clayton bivariate copulas.

4.3 Statistical inference for our models

We use the past $M = 120$ return observations to perform statistical inference for our models. The statistical inference for our models proceeds in two steps. In the first step we estimate the parameters in the marginal densities of each asset return by maximum likelihood. To guarantee that the conditional variance in Equation (10) is always positive, we restrict that the parameters $\omega_j > 0$ and $\alpha_j, \beta_j, \gamma_j \geq 0$ for $j \in \{1, ..., N\}$. In the second step we fit a copula model to the normalised ranks of the standardised residuals, which we first need to obtain.

To do so we use the estimated parameters of the marginal densities to filter out the previous M standardised residuals $\hat{z}_{j,t}$, which are estimates of the standardised errors $z_{j,t}$ in Equation (9). We then transform these standardised residuals to standard uniformly distributed variables by using the normalised ranks of the standardised residuals. Aas et al. (2009) explain that the resulting likelihood function is a pseudo-likelihood function, because these normalised ranks are only approximately uniformly distributed. We use this approach as Kim et al. (2007) show that the maximum pseudo-likelihood method performs better than the (regular) maximum likelihood method when the marginal distribution functions are unknown (which is the case as we only assume a given distribution). As a slight modification to Aas et al. (2009), we obtain the normalised ranks for asset

$j \in \{1, ..., N\}$ at time $t \in \{t^* - M, ..., t^* - 1\}$ as

$$\hat{u}_{j,t} = \frac{\text{rank}(\hat{z}_{j,t})}{M+1}.$$

We divide by $M + 1$ instead of M to ensure that the normalised ranks never become exactly equal to 1, which would give infinitely large observations if we would transform them back to standardised residuals.

We then use the observations $\hat{u}_{j,t}$ for $j \in \{1, ..., N\}$ and $t \in \{t^* - M, ..., t^* - 1\}$ to estimate our copula specifications. As explained by Dißmann et al. (2013), this involves multiple steps:

(i) Determining the tree-structure of the vine, meaning that we have to select which pairs of (un)conditional variables we use for pairwise modelling in each tree.

(ii) Choosing which copula family we use for each pair of variables selected in (i).

(iii) Estimating the parameters of each pairwise copula selected in (ii).

As mentioned in Section 4.1, the number of ways in which the N-variate copula density can be constructed grows very rapidly, so that performing steps (ii) and (iii) for all possible constructions and choosing the best combination of structure, copulas and parameters is not feasible for increasing values of N. Instead, we use the method of Dißmann et al. (2013) to select the copula specification. This is a sequential method, as step (i) uses the selected copulas (with estimated parameters) of a given tree to determine the structure of the subsequent tree. The structure of each tree is chosen so that the selected pairwise copulas model the strongest pairwise dependencies, as measured by the absolute value of the empirical Kendall's tau. For a given tree structure, the pairwise copulas are selected using the AIC and the corresponding parameters of the chosen copula are estimated using maximum likelihood.

4.4 Estimating weights using dynamic models

According to Low et al. (2013), portfolio management is a two-stage process of (1) forecasting asset returns and (2) determining the weights of the assets in the portfolio. We forecast returns by using our estimated dynamic models to simulate a sample of 1000 returns for each asset and we subsequently use these simulated returns to estimate our portfolio weights. This way the estimated portfolio weights capture the changes in volatility and other empirical properties of the returns with the goal of reducing the bias of the estimated weights. To simulate returns we first use our estimated copula specification to simulate a sample of 1000 dependent uniform observations $\{u_{j,t^*,i}\}_{i=1}^{1000}$ for asset $j \in \{1, ..., N\}$ for the coming month t^*. We do this by using Algorithm 2.2 from Dißmann et al. (2013) that is implemented in the 'VineCopula' package in the statistical software R. We then use the inverse of the estimated distribution functions of the standardised errors $z_{j,t}$ to transform the sample of simulated dependent uniform observations $\{u_{j,t^*,i}\}_{i=1}^{1000}$ to a simulated sample of standardised error terms $\{z_{j,t^*,i}\}_{i=1}^{1000}$ for $j \in \{1, ..., N\}$ for the coming month t^*. We subsequently use the estimated parameters and Equations (8) - (10) (or Equation (12) instead of Equation (8) depending on the setup) to transform this simulated sample of standardised error terms to a simulated sample of returns $\{r_{j,t^*,i}\}_{i=1}^{1000}$ for asset $j \in \{1, ..., N\}$ for the coming month t^*. We then estimate the mean vector and covariance matrix as the unbiased sample estimates of these simulated returns as

$$\hat{\boldsymbol{\mu}}_{t^*}^{\text{sim}} = \frac{1}{1000} \sum_{i=1}^{1000} \boldsymbol{r}_{t^*,i},$$

and

$$\hat{\boldsymbol{\Sigma}}_{t^*}^{\text{sim}} = \frac{1}{999} \sum_{i=1}^{1000} (\boldsymbol{r}_{t^*,i} - \hat{\boldsymbol{\mu}}_{t^*}^{\text{sim}})(\boldsymbol{r}_{t^*,i} - \hat{\boldsymbol{\mu}}_{t^*}^{\text{sim}})',$$

so that the estimated weights for the coming month t^* can be obtained as

$$\hat{\boldsymbol{w}}_{t^*} = \frac{\left(\hat{\boldsymbol{\Sigma}}_{t^*}^{\text{sim}}\right)^{-1} \hat{\boldsymbol{\mu}}_{t^*}^{\text{sim}}}{\left| \boldsymbol{\iota}' \left(\hat{\boldsymbol{\Sigma}}_{t^*}^{\text{sim}}\right)^{-1} \hat{\boldsymbol{\mu}}_{t^*}^{\text{sim}} \right|},$$

where $\boldsymbol{\iota}$ again denotes an $N \times 1$ vector of ones.

5 Trimming portfolio weights

The method discussed in Section 4 differs from the methods discussed by DeMiguel et al. (2009) as these methods all aim to reduce estimation error by reducing the variance of the estimated weights instead of reducing the (squared) bias. The portfolio allocations that perform best in their analysis apply short sale constraints on the portfolio weights, which shows the potential of applying constraints on the portfolio weights. However, Fan et al. (2012) show that the optimal portfolio that allows no short sales is not diversified enough and that this portfolio can be improved by allowing some short positions. Furthermore, Radchenko et al. (2020) demonstrate how trimming negative weights can be used for combining forecasts in a setup that allows some negative weights and they show that their method performs exceptionally well. We therefore modify the trimming methods discussed by Radchenko et al. (2020) for our portfolio optimisation setup in order to prevent very large portfolio weights. We first discuss in Section 5.1 how negative portfolio weights can arise when solving a mean-variance optimisation problem. We then explain in Section 5.2 what the trimming rules look like in our portfolio setting for a given trimming threshold $c \geq 0$ and we elaborate in Section 5.3 on how we choose this threshold.

5.1 How do negative portfolio weights arise?

Jagannathan and Ma (2003) explain how negative weights arise in a minimum variance portfolio, which minimises the variance of a portfolio that only invests in risky assets and does not look at the expected return. By deriving the first-order condition they deduce that all assets have the same marginal contribution to the portfolio variance at the optimum, meaning that assets that have larger covariances with other assets receive smaller weights. In particular, if the covariances of a certain asset with other assets are very large compared with the covariances of other pairs of assets, then the weight of that certain asset in the minimum variance portfolio can be negative.

We explore in a similar way how negative portfolio weights arise in a mean-variance portfolio that also incorporates the expected excess returns of the assets. We recall from Section 3.1 that the mean-variance portfolio can be obtained by solving the problem in Equation (1). The first-order condition of this problem reads

$$\boldsymbol{\mu}_t - \gamma \boldsymbol{\Sigma}_t \boldsymbol{x}_t = \boldsymbol{0},$$

which we can rewrite in equation-by-equation form as

$$\sum_{j=1}^{N} \sigma_{i,j,t} x_{j,t} = \frac{\mu_{i,t}}{\gamma}, \text{ for } i \in \{1, ..., N\}, \tag{13}$$

with $\sigma_{i,j,t}$ the (i,j)-th element of $\boldsymbol{\Sigma}_t$. We see from Equation (13) that the marginal contribution of each asset $i \in \{1, ..., N\}$ to the portfolio variance is now proportional to the expected excess returns of that asset, and is therefore not (necessarily) the same for all assets anymore. Furthermore, when two assets have a strong positive correlation (and a similar variance to each other) but one asset has a higher expected excess return, then the asset with the lower expected excess return tends to receive a smaller (possibly negative) weight and the asset with the higher expected excess return tends to receive a larger weight. We illustrate in Appendix C with an example how small differences in the expected excess returns of different assets can lead to large (negative) portfolio weights in case of positively correlated assets. Since the expected excess returns $\boldsymbol{\mu}_t$ and the covariance matrix $\boldsymbol{\Sigma}_t$ are unknown and need to be estimated, it follows that small estimation errors can lead to relatively large estimated weights (in absolute sense). The idea of trimming portfolio weights is to limit the range of values that the estimated weights for a certain asset can take on with the goal of reducing the estimation error in the estimated weights.

5.2 Trimming portfolio weights for a given threshold

We evaluate (modified versions of) the five trimming rules in Radchenko et al. (2020). We first evaluate all trimming methods on our five experimental data sets (based on the resulting portfolio performance) in order to determine which trimming method we will use in our seven regular data sets. The first trimming rule, referred to as TR1, is defined by letting

$$\hat{w}_{j,t^*}^{\text{TR1}} = \begin{cases} \alpha_1 \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} \geq -c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = 1 \\ \alpha_1 \cdot (-c), & \text{if } \hat{w}_{j,t^*} < -c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = 1 \\ \alpha_1^* \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} \leq c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = -1 \\ \alpha_1^* \cdot c, & \text{if } \hat{w}_{j,t^*} > c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = -1 \end{cases},$$

for $j \in \{1, ..., N\}$. This means that we first estimate the weights and then trim negative weights if the weights sum to 1 or trim positive weights if the weights sum to -1, after which we normalise all weights with either a scaling factor α_1 or α_1^* to ensure that $\boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*}^{\text{TR1}} = \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*}$. We prove in Appendix D that using TR1 decreases the absolute value of the weight of each asset (provided that at least one weight is trimmed), which is beneficial to reduce the variance of the estimated weights.

The second trimming rule, referred to as TR2, is defined by letting

$$\hat{w}_{j,t^*}^{\text{TR2}} = \begin{cases} \alpha_2 \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} \geq -c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = 1 \\ -c, & \text{if } \hat{w}_{j,t^*} < -c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = 1 \\ \alpha_2^* \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} \leq c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = -1 \\ c, & \text{if } \hat{w}_{j,t^*} > c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = -1 \end{cases},$$

for $j \in \{1, ..., N\}$. This means that we first estimate the weights and then trim negative weights if the weights sum to 1 or trim positive weights if the weights sum to -1, after which we rescale the weights that were not trimmed using either a scaling factor α_2 or α_2^* to ensure that $\boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*}^{\text{TR2}} = \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*}$. The main difference with TR1 is that the trimmed negative weights are now equal to $-c$ and that the trimmed positive weights are now equal to c, instead of a smaller value in absolute sense that depends on how much trimming occurs. We prove in Appendix E that using TR2 decreases the absolute value of the weight of each asset (provided that at least one weight is trimmed), which is beneficial to reduce the variance of the estimated weights.

The third trimming rule, referred to as TR3, differs from TR1 and TR2 by treating weights that exceed the trimming threshold differently so that the information in the magnitude of those weights is still captured. The rule is defined by letting

$$\hat{w}_{j,t^*}^{\text{TR3}} = \begin{cases} \alpha_3 \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} \geq -c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = 1 \\ \frac{-c}{\min\limits_{1\leq j\leq N} \hat{w}_{j,t^*}} \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} < -c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = 1 \\ \alpha_3^* \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} \leq c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = -1 \\ \frac{c}{\max\limits_{1\leq j\leq N} \hat{w}_{j,t^*}} \cdot \hat{w}_{j,t^*}, & \text{if } \hat{w}_{j,t^*} > c \text{ and } \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*} = -1 \end{cases},$$

for $j \in \{1, ..., N\}$. This means that we first estimate the weights and then trim negative weights if the weights sum to 1 or trim positive weights if the weights sum to -1, after which we rescale the weights that were not trimmed using either a scaling factor α_3 or α_3^* to ensure that $\boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*}^{\text{TR3}} = \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*}$. The difference with TR2 is that now only the most negative trimmed negative weight is set to $-c$ whereas other trimmed negative weights are set larger than $-c$ to keep the information about the original magnitudes of the weights preserved. Similarly, only the most positive trimmed positive weight is set to c whereas other trimmed positive weights are set smaller than c to keep the information about the original magnitudes of the weights preserved. We prove in Appendix F that using TR3 decreases the absolute value of the weight of each asset (provided that at least one weight is trimmed), which is beneficial to reduce the variance of the estimated weights.

The fourth trimming rule, referred to as TR4, differs from the previous trimming rules because the trimming of negative weights is now directly embedded into the optimisation of the weights, instead of applied afterwards. The optimisation problem solved in Radchenko et al. (2020) is given by

$$\begin{aligned} \boldsymbol{w}_t^{\text{TR4}} = \arg\min_{\boldsymbol{w}} \ & \boldsymbol{w}'\boldsymbol{\Sigma}_t\boldsymbol{w} \\ \text{s.t. } & \boldsymbol{\iota}'\boldsymbol{w} = 1 \\ & w_j \geq -c, \text{ for } j \in \{1, ..., N\}. \end{aligned} \tag{14}$$

These weights can be estimated as $\hat{\boldsymbol{w}}_{t^*}^{\text{TR4}}$ by replacing $\boldsymbol{\Sigma}_t$ by an estimator $\hat{\boldsymbol{\Sigma}}_{t^*}$ in Problem (14). Another difference with the aforementioned trimming methods is that now the weights are chosen to minimise the portfolio variance (subject to the trimming constraint), but that the expected portfolio return is not taken into account. In order to incorporate the expected portfolio return into the optimisation problem and to allow the portfolio weights to sum to -1, we can alternatively solve the problem

$$\begin{aligned}\max_{\boldsymbol{x}_t}\ & \boldsymbol{x}_t'\boldsymbol{\mu}_t - \frac{1}{2}\boldsymbol{x}_t'\boldsymbol{\Sigma}_t\boldsymbol{x}_t \\ \text{s.t. } & x_{jt} \geq (-c)\cdot\boldsymbol{\iota}'\boldsymbol{x}_t - z_2\cdot R,\ \text{ for } j\in\{1,...,N\} \\ & x_{jt} \leq (-c)\cdot\boldsymbol{\iota}'\boldsymbol{x}_t + z_1\cdot R,\ \text{ for } j\in\{1,...,N\} \\ & z_1, z_2 \in \mathbb{Z}_{\leq 1} \\ & z_1 + z_2 = 1,\end{aligned} \tag{15}$$

with $\mathbb{Z}_{\leq 1}$ denoting the set of integers smaller than or equal to 1 and R a very large number such that we only have the first constraint set if $z_1 = 1$ and only have the second constraint set if $z_2 = 1$. This formulation ensures that we trim negative weights when $\boldsymbol{\iota}'\boldsymbol{x}_t > 0$ (corresponding with $z_1 = 1$) or that we trim positive weights when $\boldsymbol{\iota}'\boldsymbol{x}_t < 0$ (corresponding with $z_2 = 1$). We then take $\boldsymbol{w}_t = \frac{\boldsymbol{x}_t}{|\boldsymbol{\iota}'\boldsymbol{x}_t|}$ afterwards so that $|\boldsymbol{\iota}'\boldsymbol{w}_t| = 1$. Estimated weights $\hat{\boldsymbol{w}}_{t^*}^{\text{TR4}^{**}}$ for this approach can be obtained by plugging the relevant estimates $\hat{\boldsymbol{\mu}}_{t^*}$ and $\hat{\boldsymbol{\Sigma}}_{t^*}$ of the mean vector and covariance matrix into Problem (15) and normalising the outcome by the absolute sum.

The fifth trimming rule, referred to as TR5, also embeds the trimming of negative weights directly into the optimisation of the weights, but does this by using the gross-exposure constraint discussed by Fan et al. (2012). The resulting optimisation problem solved in Radchenko et al. (2020) is given by

$$\begin{aligned}\boldsymbol{w}_t^{\text{TR5}} = \arg\min_{\boldsymbol{w}}\ & \boldsymbol{w}'\boldsymbol{\Sigma}_t\boldsymbol{w} \\ \text{s.t. } & \boldsymbol{\iota}'\boldsymbol{w} = 1 \\ & \|\boldsymbol{w}\|_1 \leq 1 + c.\end{aligned} \tag{16}$$

The hyperparameter $c \geq 0$ determines how stringent the trimming of negative weights is, where $c = 0$ is equivalent to a no-short-sale constraint and $c = \infty$ is equivalent to imposing no additional constraint. With this constraint the weights cannot become smaller than $-c/2$ (or larger than $1 + c/2$), as shown in Appendix G. Moreover, it follows by the same arguments that the sum of the negative weights cannot become smaller than $-c/2$ and that the sum of the positive weights cannot become larger than $1 + c/2$. For TR4 the smallest possible sum of the negative weights equals $-c\cdot(N-1)$ and the largest possible sum of the positive weights equals $1 + c\cdot(N-1)$, which can be obtained by letting all weights except for one be equal to the lower bound of $-c$. This illustrates that trimming rule TR5 restricts the range of values that the estimated weights can take on more than trimming rule TR4 whenever $c > 0$. The trimmed weights can be estimated as $\hat{\boldsymbol{w}}_{t^*}^{\text{TR5}}$

by replacing $\boldsymbol{\Sigma}_t$ by an estimator $\hat{\boldsymbol{\Sigma}}_{t^*}$ in Problem (16). As with TR4, the weights are chosen to minimise the portfolio variance, whereas the expected portfolio return is not taken into account. To incorporate the expected portfolio return into the optimisation problem and to allow the portfolio weights to sum to -1, we can solve

$$\begin{aligned}
&\max_{\boldsymbol{x}_t} \boldsymbol{x}_t' \boldsymbol{\mu}_t - \frac{1}{2} \boldsymbol{x}_t' \boldsymbol{\Sigma}_t \boldsymbol{x}_t \\
&\text{s.t. } \|\boldsymbol{x}_t\|_1 \leq (1+c) \cdot \boldsymbol{\iota}' \boldsymbol{x}_t + z_2 \cdot R \\
&\quad\;\; \|\boldsymbol{x}_t\|_1 \leq -(1+c) \cdot \boldsymbol{\iota}' \boldsymbol{x}_t + z_1 \cdot R \\
&\quad\;\; z_1, z_2 \in \mathbb{Z}_{\leq 1} \\
&\quad\;\; z_1 + z_2 = 1,
\end{aligned} \tag{17}$$

with $\mathbb{Z}_{\leq 1}$ denoting the set of integers smaller than or equal to 1. We then take $\boldsymbol{w}_t = \frac{\boldsymbol{x}_t}{|\boldsymbol{\iota}' \boldsymbol{x}_t|}$ afterwards so that $|\boldsymbol{\iota}' \boldsymbol{w}_t| = 1$. Estimated weights $\hat{\boldsymbol{w}}_{t^*}^{\text{TR5}^{**}}$ for this approach can be obtained by plugging the relevant estimates $\hat{\boldsymbol{\mu}}_{t^*}$ and $\hat{\boldsymbol{\Sigma}}_{t^*}$ of the mean vector and covariance matrix into Problem (17) and normalising the outcome by the absolute sum. Similar to our comparison between TR5 and TR4, we can see that trimming rule TR5** restricts the range of values that the estimated weights can take on more than the other trimming rules that can have both positive and negative sums of weights (so TR1, TR2, TR3 and TR4**) whenever $c > 0$. This means that applying the fifth trimming rule generally leads to a larger variance reduction of the estimated weights, possibly at the cost of a larger increase in bias.

5.3 Selecting the threshold for trimming portfolio weights

We now discuss how we select the trimming threshold, for which we use two different methods. The first method simply sets the trimming threshold c equal to 0, leading to the largest amount of trimming and thereby the strongest reduction in the variance of the estimated portfolio weights. The second method is our own version of the data-driven approaches used by Fan et al. (2012) and Radchenko et al. (2020) to choose the value of c for each point in time. Radchenko et al. (2020) report that the performance of a data-driven threshold is almost as good as the ex ante best fixed threshold and they argue that a data-driven threshold therefore offers a feasible way to estimate the (unknown) best threshold in practice. The data-driven approach of Fan et al. (2012) minimises the average squared return over a validation set containing (pseudo) out-of-sample observations and the data-driven approach of Radchenko et al. (2020) minimises the MSE over such a validation set. The approach of Fan et al. (2012) thus ignores the expected portfolio return, whereas the approach of Radchenko et al. (2020) is not feasible in a portfolio setting as we cannot compute the MSE of the optimal weights because the true optimal weights are unknown. However, our approach is a feasible alternative that incorporates both the expected portfolio return and variance as we select the trimming threshold from a set of possible values in order to achieve the highest Sharpe ratio in the validation set containing (pseudo) out-of-sample observations.

We evaluate two sets of possible trimming thresholds, so that we choose $c \in \mathcal{C}_1$ or $c \in \mathcal{C}_2$ in our data-driven approach, where

$$\mathcal{C}_1 = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$$

and

$$\mathcal{C}_2 = \{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}.$$

For a given $c_{\max} \in \{1, 2\}$ the most negative weights cannot be smaller than $-c_{max}$ if the weights sum to 1 and the most positive weights cannot be larger than $c_{\max}$ if the weights sum to -1. We choose these relatively low values of $c_{\max}$ as Radchenko et al. (2020) conclude that allowing for a larger set of possible thresholds gains limited consistency but introduces more uncertainty.

As before, we use a rolling window of $M = 120$ return observations to estimate the weights for the next period. In addition to this, we use the previous $M/6 = 20$ return observations to compute for the considered thresholds the (pseudo) out-of-sample Sharpe ratio, and we pick the threshold that maximises this Sharpe ratio. This means that we can estimate the trimmed weights $\hat{\boldsymbol{w}}_{t^*}^{\text{TR}}$ for $t^* \in \{141, ..., T\}$. To do so, we compute for a given threshold c the Sharpe ratio of the portfolio returns that are obtained with the trimmed weights that use this threshold. That is, we compute for a given c

$$\widehat{\text{SR}}(c, t^*, M) = \frac{\bar{r}_{p,t}(c, t^*, M)}{\sqrt{\frac{1}{M/6-1} \sum_{t=t^*-M/6}^{t^*-1} (r_{p,t}(c) - \bar{r}_{p,t}(c, t^*, M))^2}},$$

where

$$\bar{r}_{p,t}(c, t^*, M) = \frac{1}{M/6} \sum_{t=t^*-M/6}^{t^*-1} r_{p,t}(c),$$

with

$$r_{p,t}(c) = \hat{\boldsymbol{w}}_t^{\text{TR}}(c)' \boldsymbol{r}_t \tag{18}$$

the resulting portfolio return in month t. Here the weights $\hat{\boldsymbol{w}}_t^{\text{TR}}(c)$ are estimated using the returns from month $t - M$ until month $t - 1$. The optimal threshold c^* is chosen as

$$c^* = \arg\max_c \widehat{\text{SR}}(c, t^*, M),$$

and the trimmed weights obtained with this data-driven approach are given by $\hat{\boldsymbol{w}}_{t^*}^{\mathrm{TR}} = \hat{\boldsymbol{w}}_{t^*}^{\mathrm{TR}}(c^*)$. In a similar way we can repeat the selection of the optimal threshold and estimation of the resulting trimmed weights in a setting where we use returns net of transaction costs. In particular, we then first compute for a given c the portfolio returns in Equation (18) and transform these to net portfolio returns $r_{p,t}^{\mathrm{net}}(c)$ using Equations (3) and (4). We can then use these portfolio returns net of transaction costs to compute

$$\widehat{\mathrm{SR}}^{\mathrm{net}}(c, t^*, M) = \frac{\bar{r}_{p,t}^{\mathrm{net}}(c, t^*, M)}{\sqrt{\frac{1}{M/6-1}\sum_{t=t^*-M/6}^{t^*-1}(r_{p,t}^{\mathrm{net}}(c) - \bar{r}_{p,t}^{\mathrm{net}}(c, t^*, M))^2}},$$

where

$$\bar{r}_{p,t}^{\mathrm{net}}(c, t^*, M) = \frac{1}{M/6}\sum_{t=t^*-M/6}^{t^*-1} r_{p,t}^{\mathrm{net}}(c).$$

Using this method potentially leads to a different optimal threshold c^*, as we now specifically incorporate transaction costs into the selection procedure, which will favour portfolio weights that do not exhibit very large changes over time.

6 Simulation-based shrinkage rule

As mentioned before, the method discussed in Section 4 aims to reduce estimation error by reducing the (squared) bias of the estimated portfolio weights. In doing so we have to estimate more parameters because we estimate the parameters in the models for the distribution of the asset returns in addition to estimating the mean vector and covariance matrix of the simulated returns. This is likely to increase the variance of the resulting estimated weight vector. To prevent that the increase in variance of the estimated weights outweighs the benefits of the bias reduction we develop a method that shrinks the estimated weights when the variance of the estimated weights is high.

Shrinkage has been applied to the estimated mean vector by Jorion (1986) and to the estimated covariance matrix by Ledoit and Wolf (2003) and Ledoit and Wolf (2004). Later Golosnoy and Okhrin (2007) have applied shrinkage directly to the estimated weights. They show that portfolio selection can be improved by means of a multivariate shrinkage estimator for portfolio weights that shrinks the estimated mean-variance weights to the vector of current portfolio weights for increasing estimation uncertainty. Additionally, they derive expressions for the optimal shrinkage intensity. However, this derivation depends on the assumptions that the asset returns are independent and follow a normal distribution, both of which are not confirmed by the empirical studies of Cont (2001) and Stoyanov et al. (2011). Moreover, the optimal shrinkage intensity derived by Golosnoy and Okhrin (2007) depends on the mean and covariance matrix of the asset returns and therefore has to be estimated as well. The method that we develop is more robust and flexible as it does not make any distributional assumption and it can be used in combination with other methods (as discussed in Section 7). It can be applied for any setup that simulates asset returns to determine portfolio weights.

The idea behind our method is to reduce the estimation error of the estimated portfolio weights by shrinking the estimated weights towards equal weights if the sum of the sample variances of the estimated weights is high compared to the previous periods. We hereby use the $1/N$ portfolio as a shrinkage target as the variance of the estimated weights with this allocation is zero and our goal is to reduce the variance of the estimated weights. Moreover, DeMiguel et al. (2009) show that it is very hard to outperform the $1/N$ portfolio rule, making it a reasonable shrinkage target for our method. We further make the choice to compare the sum of the variances of the simulated weights over all assets, as comparing the variance of the weights of each asset individually with previous periods would neglect the fact that other weights might have a very high variance so that it would be preferred to take equal weights.

We now explain in detail how our method works. We use the past $M = 120$ return observations to estimate the portfolio weights in a similar way as described in Section 4.4. The difference is that we now simulate 10 times as much return observations $\{r_{j,t^*,i}\}_{i=1}^{10000}$ for asset $j \in \{1, ..., N\}$ for the coming month t^*. We then separate these 10000 simulated returns into 10 samples of 1000 simulated returns and use these 10 samples of returns to generate 10 different weight vectors $\{\hat{\boldsymbol{w}}_{t^*,s}\}_{s=1}^{10}$. Since the estimated weights $\{\hat{\boldsymbol{w}}_{t^*,s}\}_{s=1}^{10}$ can either sum to 1 or to -1, we define the sets $\mathcal{W}_{t^*,1} = \{\hat{\boldsymbol{w}}_{t^*,s} : \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*,s} = 1, s = 1, ..., 10\}$ and $\mathcal{W}_{t^*,-1} = \{\hat{\boldsymbol{w}}_{t^*,s} : \boldsymbol{\iota}'\hat{\boldsymbol{w}}_{t^*,s} = -1, s = 1, ..., 10\}$ and we proceed the analysis with the set of weights that contains the most elements (choosing $\mathcal{W}_{t^*,1}$ if both would contain the same amount of elements).

In addition we use the past $M/6 = 20$ observations of estimated portfolio weights (that are determined prior to setting some of them equal to $1/N$ weights) to compare the sample variance of the estimated weights for $j \in \{1, ..., N\}$ and $t \in \{t^* - M/6, ..., t^* - 1\}$, which is given by

$$\widehat{\text{var}}(\hat{w}_{j,t}) = \begin{cases} \frac{1}{|\mathcal{W}_{t,1}|-1} \sum_{s\in\mathcal{W}_{t,1}} \left(\hat{w}_{j,t,s} - \frac{1}{|\mathcal{W}_{t,1}|} \sum_{s\in\mathcal{W}_{t,1}} \hat{w}_{j,t,s}\right)^2, & \text{if } |\mathcal{W}_{t,1}| \geq |\mathcal{W}_{t,-1}| \\ \frac{1}{|\mathcal{W}_{t,-1}|-1} \sum_{s\in\mathcal{W}_{t,-1}} \left(\hat{w}_{j,t,s} - \frac{1}{|\mathcal{W}_{t,-1}|} \sum_{s\in\mathcal{W}_{t,-1}} \hat{w}_{j,t,s}\right)^2, & \text{otherwise} \end{cases}.$$

As a decision rule, we determine the weights at time t^* as

$$\hat{\boldsymbol{w}}_{t^*} = \begin{cases} \hat{\boldsymbol{w}}_{t^*}^{\text{avg}}, & \text{if } \sum_{j=1}^{N} \widehat{\text{var}}(\hat{w}_{j,t^*}) \leq \sum_{j=1}^{N} \widehat{\text{var}}(\hat{w}_{j,t})_{(d \cdot M/6)} \\ \frac{1}{N}\boldsymbol{\iota}, & \text{otherwise} \end{cases}, \tag{19}$$

with

$$\hat{\boldsymbol{w}}_{t^*}^{\text{avg}} = \begin{cases} \frac{1}{|\mathcal{W}_{t^*,1}|} \sum_{s \in \mathcal{W}_{t^*,1}} \hat{\boldsymbol{w}}_{t^*,s}, & \text{if } |\mathcal{W}_{t^*,1}| \geq |\mathcal{W}_{t^*,-1}| \\ \frac{1}{|\mathcal{W}_{t^*,-1}|} \sum_{s \in \mathcal{W}_{t^*,-1}} \hat{\boldsymbol{w}}_{t^*,s}, & \text{otherwise} \end{cases}, \tag{20}$$

where $\sum_{j=1}^{N} \widehat{\text{var}}(\hat{w}_{j,t})_{(d \cdot M/6)}$ denotes the $(d \cdot M/6)^{\text{th}}$ order statistic and hence the d^{th} quantile of the past $M/6$ sums of sample variances, and $\boldsymbol{\iota}$ again denotes a vector of ones. Here d denotes the shrinkage threshold that influences how often we replace the estimated weights by equal weights. We evaluate the results by picking $d \in \{0.05, 0.10, 0.15, ..., 0.90, 0.95, 1.00\}$ at each point in time using a data-driven approach similar to what we do for trimming the portfolio weights in Section 5.3. This means that we take $\hat{\boldsymbol{w}}_{t^*} = \hat{\boldsymbol{w}}_{t^*}(d^*)$, where the threshold d^* maximises the (pseudo) out-of-sample Sharpe ratio (either with or without incorporating transaction costs). Selecting a lower value of d implies more shrinkage towards equal weights. As a consequence of using the data-driven approach for selecting the shrinkage threshold d, we have to use an additional $M/6 = 20$ observations. This means that we can estimate the weights with our simulation-based shrinkage rule for $t^* \in \{161, ..., T\}$.

As specified in Equation (19), we take the average simulated weights if the sum of sample variances is relatively low compared with the previous months. Taking the average of those simulated weights is likely to reduce the variance of the estimated weights, as possible outliers are averaged out. We therefore also evaluate the portfolio performance of choosing $\hat{\boldsymbol{w}}_{t^*}^{\text{avg}}$ given in Equation (20) in each period to investigate whether it is the shrinkage of portfolio weights or the averaging of portfolio weights that leads to a difference in performance.

7 Combining multiple methods

We can also combine the aforementioned methods. It is particularly interesting to see whether trimming the weights obtained using dynamic forecasting and then applying our simulation-based shrinkage rule afterwards has benefits over applying the methods separately. To evaluate the added benefit of a certain method, both on its own and in combination with other methods, we evaluate all possible combinations of methods based on five decision variables. These are (i) whether we estimate weights based on dynamic forecasting of returns using copulas (or whether we use sample estimates based on previous returns), (ii) whether we simulate multiple return samples and take the average weights over those samples (or whether we obtain weights based on one simulated return sample), (iii) whether we use trimming of portfolio weights (or not), (iv) whether we use a data-driven approach for selecting the trimming threshold c (or whether we always take $c = 0$) and (v) whether we use our simulation-based shrinkage rule (or not). This gives a total of 12 combinations to compare, which are listed in Table 2.

Table 2: Combinations of methods

combination	(i)	(ii)	(iii)	(iv)	(v)
1	no	-	no	-	-
2	no	-	yes	no	-
3	no	-	yes	yes	-
4	yes	no	no	-	no
5	yes	yes	no	-	no
6	yes	no	yes	no	no
7	yes	no	yes	yes	no
8	yes	yes	yes	no	no
9	yes	yes	yes	yes	no
10	yes	yes	no	-	yes
11	yes	yes	yes	no	yes
12	yes	yes	yes	yes	yes

Note: Decision variable (i) indicates whether we use dynamic modelling. Decision variable (ii) indicates whether we simulate multiple samples and take the average weights over these samples. Decision variable (iii) indicates whether we use trimming of portfolio weights and decision variable (iv) indicates whether this is done by a data-driven approach (as opposed to using a fixed threshold of 0). The last decision variable (v) indicates whether we use our simulation-based shrinkage method. If a certain decision variable is not applicable it is denoted with -.

Combination 12 first estimates trimmed portfolio weights with a data-driven approach to select the trimming threshold, after which the simulation-based shrinkage rule is used. Therefore, portfolio weights $\hat{\boldsymbol{w}}_{t^*}$ can be obtained with this combination for $t^* \in \{181, ..., T\}$ and to make a fair comparison between the combinations we evaluate all combinations for $t^* \in \{181, ..., T\}$. We do so by comparing the Sharpe ratios (both with and without incorporating transaction costs) and we determine the significance of differences in Sharpe ratio using HAC inference with a prewhitened Parzen kernel as discussed in Ledoit and Wolf (2008).

8 Results

In this section we discuss the results of our research. In Sections 8.1 and 8.2 we evaluate using the experimental data sets which dynamic model setup and which trimming method we continue our analysis on the regular data sets with. We then compare all different combinations of portfolio strategies with the $1/N$ strategy in Section 8.3 to draw general conclusions about which combinations perform best. Thereafter we analyse the performance of the three methods described in Sections 4, 5 and 6 in more detail in Sections 8.4, 8.5 and 8.6.

8.1 Determining the dynamic model setup

We first analyse the fit of each of the dynamic models by means of the likelihood ratio test of Vuong (1989). Table 3 reports the averages (over all estimation windows) of these likelihood ratio test statistics for the first experimental data set. This table shows that model 5 and 7 obtain the highest likelihoods and that their likelihoods are similar to each other. It also shows that model 1 and model 3 have similar likelihoods and that their performance is closest to the performance of model 5 and model 7. Furthermore, it follows from comparing the marginal models with and without modelling an autoregressive model for the returns (for a given copula model, so e.g. model 1 versus model 5 and model 2 versus model 6), that not using an autoregressive model leads to a better log-likelihood.

Table 3: Average results of Vuong test for experimental data set 1 without parameter correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	5.55	0.08	8.22	-1.62	3.44	-1.60	5.62
	(0.00)	(0.15)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-5.44	3.03	-6.10	-1.76	-6.09	1.61
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			8.20	-1.65	3.36	-1.67	5.57
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-8.50	-3.61	-8.52	-1.44
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					5.73	-0.04	8.27
					(0.00)	(0.43)	(0.00)
model 6						-5.68	2.77
						(0.00)	(0.00)
model 7							8.39
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the first experimental data set) without a correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the *p*-values of the average Vuong test statistics.

These three main findings also hold when applying Vuong tests with AIC and SIC corrections for the number of parameters and when fitting the dynamic models for the other experimental data sets, the results of which can be found in Appendix H. We conclude from these findings that it is preferred for modelling the dependence structure to choose pairwise copulas from a large set of possible copulas rather than limiting ourselves to a certain type of pairwise copulas. This confirms the finding of Okhrin et al. (2013) that it is rarely a feasible assumption in practical applications to assume that all asset return pairs have the same dependence structure. Furthermore, the results imply that we do not favour a regular vine structure or a canonical vine structure, as the difference between these approaches is small and not consistent across the experimental data sets. Since the canonical vine structure is a special case of the regular vine structure that performs well when a certain variable is a key variable that governs interactions in a data set, this suggests that such a key variable is present. This makes sense as all data sets contain (an equivalent of) a market portfolio that can function as such a key variable. Lastly, the Vuong tests indicate that the model fit improves if we do not allow an autoregressive specification for the returns, which seems to confirm the stylised facts of Cont (2001) that autocorrelations of asset returns are often insignificant.

This last finding might seem to be surprising, because if omitting the autoregressive parameters would be optimal then the larger model could simply set them to zero. However, we only include these parameters in marginal modelling, and subsequently the dependence of the normalised ranks of the standardised residuals is modelled with copulas. As the residuals are different for different marginal models, the log-likelihood can differ for the copula models. Since fitting marginal returns and fitting the dependence structure happens separately, it is possible that a better marginal fit yields residuals that lead to a worse fit for modelling the dependence structure. Indeed, a closer inspection learns us that despite a better fit for the marginal returns, the R-vine copula model for model 1 has only 24/319 times a better fit than R-vine copula model for model 5.

We now look at the performance of the portfolios constructed using simulated returns from each of the different dynamic models, as described in Section 4.4. Table 4 shows the Sharpe ratios that do not incorporate transaction costs for the five experimental data sets.

Table 4: Sharpe ratios that do not incorporate transaction costs for the different dynamic models

dynamic model	data set 1	data set 2	data set 3	data set 4	data set 5
model 1	0.0563	0.0544	0.0588	0.0440	0.0951
model 2	0.0861	0.0466	0.0632	0.0755	0.0801
model 3	-0.0100	0.0544	0.0151	-0.0059	0.1334
model 4	-0.0152	0.0135	0.0468	0.0903	0.0329
model 5	0.0017	0.0926	-0.0204	0.1161	0.0210
model 6	-0.0543	-0.0543	0.0902	0.0489	0.0977
model 7	-0.1105	0.0926	0.0593	0.0180	-0.0200
model 8	0.0240	0.0730	-0.0111	0.0578	0.1075

Note: This table reports the monthly Sharpe ratios that do not incorporate transaction costs for the experimental data sets. The different dynamic models are described in Table 1.

Model 1 and model 2 perform best in terms of Sharpe ratio without taking transaction costs into account, followed (at some distance) by model 5 and model 8. This implies that allowing for autoregressive behaviour of asset returns is beneficial for portfolio performance, although differences are small. We do not report the Sharpe ratios that incorporate transaction costs because all resulting portfolios take such extreme positions that returns net of transaction costs cannot be computed correctly, as discussed in Section 3.2.

The combined results of the Vuong tests with the portfolio performance over the experimental data sets indicate that either model 1 or model 5 is preferred, implying that it is beneficial to model the dependence structure using a regular vine copula model and to determine the pairwise copulas from a wide range of possibilities (using the AIC as a selection criterion). As model 1 leads to a better out-of-sample performance, which is what we are ultimately interested in, and theoretically model 5 is a submodel of model 1, we choose to use model 1 as our dynamic model. This means that we allow autoregressive behaviour when modelling the marginal returns.

8.2 Determining the trimming method

We now evaluate the performance of the different trimming methods described in Section 5 on the experimental data sets. Table 5 and Table 6 show the Sharpe ratios that incorporate transaction costs for the different trimming methods over the five experimental data sets, where the trimming threshold c is selected from the set $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. This threshold is selected using the data-driven approach that maximises the Sharpe ratio that either incorporates transaction costs (Table 5) or does not incorporate transaction costs (Table 6). Table 5 shows that the trimming methods that directly estimate trimmed weights (TR4, TR4**, TR5 and TR5**) generally perform better than the trimming methods that trim the estimated weights (TR1, TR2 and TR3). The difference between TR4** and TR5** on the one hand and TR1, TR2 and TR3 on the other can be explained by noting that large (negative) weights in a mean-variance portfolio often arise for assets that have a strong positive correlation but different expected returns, as explained in Section 5.1. When large (negative) weights are trimmed, it might not be beneficial anymore to have a large position (with opposite sign) in the assets that have a large positive correlation with the assets whose weights are trimmed. The methods that directly estimate the trimmed weights can capture this when estimating the weights of other assets, whereas the trimming methods that trim the estimated weights rescale the positions of all non-trimmed weights by the same factor. We illustrate this explanation with an example in Appendix I. The relatively good performance of TR4 and TR5 can be explained by a similar argument, as large (negative) weights in the minimum variance portfolio arise due to large positive correlations between different assets. Additionally, minimum variance portfolios are often found to have a better out-of-sample performance than mean-variance portfolios (by among others DeMiguel et al. (2009)) as they do not suffer from estimation error in the expected return vector.

Among these direct trimming methods we see from Table 5 that the trimming methods that maximise a trade-off between the mean and variance (TR4** and TR5**) perform better than the trimming methods that only minimise the variance (TR4 and TR5). This can be explained by the fact that TR4 and TR4** respectively TR5 and TR5** restrict the range of values that the estimated weights can assume to the same extent (for the same trimming threshold c). Since the bad out-of-sample performance of the mean-variance portfolio relative to the minimum variance portfolio that is generally found in the literature (e.g. by DeMiguel et al. (2009)) is caused by a large variance of the estimated weights (as a result of estimation errors in the sample inputs), we would expect that trimming the weights is able to solve this problem. Given that the portfolio weights cannot become very large anymore, it turns out that incorporating information about the expected returns of the assets is beneficial for reducing the bias of the estimated weights and thereby increasing portfolio performance.

We further see that the trimming methods that constrain the sum of the absolute portfolio weights (TR5 and TR5**) perform better than their counterparts that constrain individual weights (TR4 and TR4**). We explained in Section 5.2 that the methods that constrain the sum of the absolute portfolio weights restrict (for a given trimming threshold c) the range of values that the estimated weights can take on more than the methods that constrain individual weights. Since the aforementioned methods perform better, this indicates that applying more stringent restrictions reduces the variance of the estimated weights to a larger extent than that it increases the (squared) bias of the estimated weights. Combining our observations we conclude that trimming method TR5** performs best for selecting the trimming threshold c from the set with $c_{\max} = 1$ with the data-driven approach that maximises the Sharpe ratio that incorporates transaction costs.

Table 6 shows that using a data-driven approach that does not incorporate transaction costs leads to somewhat different results. The main difference is that trimming method TR4** performs relatively worse. This is because this trimming method allows relatively large weights (compared with trimming methods that restrict the sum of absolute portfolio weights) for a given trimming threshold c. Since the trimming threshold is now chosen with a data-driven approach that does not incorporate transaction costs, the selected trimming thresholds will be larger so that less stringent restrictions on the portfolio weights are applied. This leads to larger position changes over time which increases transaction costs, so that the performance deteriorates relative to the other methods that have smaller position changes over time.

However, the main conclusion that trimming method TR5** performs best remains unchanged. Furthermore, comparing Table 5 with Table 6 shows that applying the data-driven approach that incorporates transaction costs leads to a better performance of the best trimming methods. This makes sense as we report the Sharpe ratios that incorporate transaction costs, such that the data-driven approach that also incorporates transaction costs is expected to perform better by design. However, applying a data-driven approach that incorporates transaction costs also leads to higher Sharpe ratios that do not incorporate transaction costs for the best-performing methods, as can be seen from Tables 33 and 34 in Appendix J. This can be explained by the fact that using a data-driven approach that incorporates transaction costs leads to smaller values of c, such that more trimming occurs. This apparently reduces the variance of the estimated weights to a larger extent than that it increases the (squared) bias of the estimated weights, resulting in a better performance even without considering transaction costs.

The same conclusions hold for selecting the best trimming method when the trimming threshold c is set to 0 or is selected with a data-driven approach with $c_{\max} = 2$ that maximises the Sharpe ratio that does (or does not) incorporate transaction costs; these results can be found in Appendix J. Additionally, we see that the results improve slightly by fixing the trimming threshold at 0 and slightly deteriorate when using a data-driven approach with $c_{\max} = 2$ instead of $c_{\max} = 1$, once again confirming that reducing the variance of the estimated weights (at the cost of potential increases in squared bias) is an effective way to improve portfolio performance. Taken together, we choose to use trimming method TR5** for trimming portfolio weights in the regular data sets; we hereby separately evaluate the case where we set $c = 0$ and the case where we apply the data-driven approach that maximises the Sharpe ratio that incorporates transaction costs to find a trimming threshold c from the set $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

Table 5: Sharpe ratios that incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 1$ that incorporates transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2425	0.2136	0.1615	-0.1611	-0.0852
TR2	0.2252	0.2168	0.0290	-0.1242	-0.0834
TR3	0.2324	0.2165	0.1097	-0.1060	-0.0394
TR4	0.1851	0.1240	0.1078	0.0787	0.2322
TR4**	0.1595	0.2316	0.0839	0.1683	0.2536
TR5	0.2488	0.1229	0.1964	0.0986	0.2395
TR5**	0.2216	0.2275	0.2059	0.1840	0.3103

Note: This table reports the monthly Sharpe ratios that incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ using the data-driven approach that maximises the Sharpe ratio that incorporates transaction costs.

Table 6: Sharpe ratios that incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 1$ that does not incorporate transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2425	0.2180	0.1396	-0.1214	-0.0833
TR2	0.2218	0.2221	-0.0307	-0.1092	-0.1021
TR3	0.2330	0.2138	0.0761	-0.0690	-0.0433
TR4	0.1849	0.1209	0.0193	0.0758	0.1704
TR4**	0.1187	0.2179	0.0348	0.0442	0.0680
TR5	0.2526	0.1222	0.1955	0.1163	0.1816
TR5**	0.2159	0.2191	0.1896	0.1609	0.2648

Note: This table reports the monthly Sharpe ratios that incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ using the data-driven approach that maximises the Sharpe ratio that does not incorporate transaction costs.

8.3 Comparing the different portfolio strategies with the $1/N$ strategy

We use the dynamic model and trimming method that perform best on our experimental data sets to compare the different combinations of portfolio strategies with the $1/N$ strategy on the regular data sets. Table 7 shows the Sharpe ratios that do not incorporate transaction costs for all investigated combinations. We see that the combinations that estimate trimmed portfolio weights generally perform best, suggesting that reducing the variance of the estimated weights by trimming portfolio weights is an effective way to improve portfolio performance. We further see that constructing portfolios based on simulated returns that satisfy empirical properties of asset returns leads to a worse performance. This suggests that the reduction in (squared) bias due to incorporating changing patterns in the return volatility is not enough to outweigh the corresponding increase in the variance of the estimated weights that arises because more parameters have to be estimated. Moreover, we see that no combination is able to consistently outperform the $1/N$ portfolio over all seven data sets. As the variance of the estimated weights in the $1/N$ portfolio is zero, this implies that the increase in (squared) bias of the estimated portfolio weights in the $1/N$ portfolio is not significantly larger than the resulting decrease in variance over all seven data sets.

Looking at the results of each combination in more detail, we see that combination 2, 3, 6, 7, 8, 9 and 11 significantly outperform the $1/N$ portfolio in two or three data sets and are never significantly outperformed by the $1/N$ portfolio, which again implies that reducing the variance of the estimated weights by trimming portfolio weights is an effective way to improve portfolio performance. Combination 12 performs slightly better than the $1/N$ portfolio, although it only significantly outperforms the $1/N$ portfolio in one data set. We further observe that combination 1 and 4 perform worse than the $1/N$ portfolio, although the difference is only significant in one data set. This suggests that a regular mean-variance portfolio and a mean-variance portfolio based on dynamic modelling of specific asset return properties do not perform well if nothing is done to reduce the variance of the estimated weights. Finally, combination 5 and 10 perform similarly to the $1/N$ portfolio as differences in Sharpe ratio are not significant.

Table 7: Sharpe ratios that do not incorporate transaction costs

portfolio rule	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
$1/N$	0.1077	0.1657	0.1802	0.2229	0.1656	0.1696	0.1775
comb. 1 (mv)	0.0739	-0.0229	0.0676	0.2176	0.0702	0.0157	0.2011
	(0.77)	(0.02)	(0.10)	(0.93)	(0.15)	(0.14)	(0.88)
comb. 2	0.1066	0.1215	0.1576	0.2335	0.2260	0.2491	0.3292
	(0.98)	(0.21)	(0.57)	(0.84)	(0.00)	(0.18)	(0.01)
comb. 3	0.1316	0.1186	0.1686	0.2111	0.3405	0.3250	0.3549
	(0.80)	(0.26)	(0.79)	(0.84)	(0.00)	(0.01)	(0.01)
comb. 4	0.1821	0.0229	0.0716	0.0593	0.0457	-0.0021	0.2228
	(0.54)	(0.10)	(0.22)	(0.75)	(0.25)	(0.04)	(0.49)
comb. 5	0.1461	0.0258	0.0653	-0.0406	0.2541	0.1054	0.2092
	(0.70)	(0.07)	(0.21)	(0.33)	(0.29)	(0.44)	(0.58)
comb. 6	0.0215	0.1494	0.1367	0.1310	0.3335	0.3742	0.3415
	(0.37)	(0.77)	(0.43)	(0.20)	(0.00)	(0.00)	(0.00)
comb. 7	0.0813	0.1688	0.1350	0.1339	0.3654	0.3689	0.3821
	(0.72)	(0.96)	(0.41)	(0.22)	(0.00)	(0.00)	(0.00)
comb. 8	0.0842	0.1939	0.1532	0.1106	0.3380	0.3563	0.3331
	(0.75)	(0.57)	(0.61)	(0.12)	(0.00)	(0.00)	(0.00)
comb. 9	0.0712	0.2071	0.1587	0.1089	0.3713	0.3747	0.3743
	(0.61)	(0.41)	(0.67)	(0.12)	(0.00)	(0.00)	(0.00)
comb. 10	0.0824	0.1660	0.1602	0.1028	0.1858	0.1750	0.1559
	(0.71)	(0.99)	(0.40)	(0.07)	(0.15)	(0.69)	(0.69)
comb. 11	0.0847	0.1685	0.2054	0.2257	0.2500	0.2490	0.2379
	(0.37)	(0.92)	(0.48)	(0.92)	(0.05)	(0.05)	(0.08)
comb. 12	0.0426	0.1723	0.2043	0.1435	0.2654	0.2345	0.2185
	(0.09)	(0.84)	(0.53)	(0.12)	(0.00)	(0.09)	(0.21)

Note: This table reports the monthly Sharpe ratios that do not incorporate transaction costs of the given portfolio rules, where the different combinations are described in Table 2. In parentheses are the p-values of the differences in Sharpe ratio with the Sharpe ratio of the $1/N$ strategy.

We now analyse the results when the Sharpe ratios are computed for the returns net of transaction costs. These Sharpe ratios are shown in Table 8. We again see that no combination is able to consistently outperform the $1/N$ portfolio over all seven data sets. Moreover, the performance relative to the $1/N$ portfolio deteriorates for all combinations; this makes sense as the $1/N$ portfolio has very low transaction costs because the target weights in each period are equal to $1/N$. However, we see that combination 2 and 3 are still able to significantly outperform the $1/N$ portfolio in one data set and that these combinations are never significantly outperformed by the $1/N$ portfolio. This implies that, relative to the $1/N$ portfolio, trimming portfolio weights in a setting that does not specifically model empirical properties of asset returns reduces the (squared) bias of the estimated weights more than that it increases the variance of the estimated weights, so that trimming remains beneficial after correcting for transaction costs.

We further see that the combinations that do not trim portfolio weights or do not reduce the variance of the estimated weights by a simulation-based shrinkage rule take such extreme positions that returns net of transaction costs cannot even be computed correctly. This holds for combination 1, 4 and 5 and clearly shows the necessity of reducing the variance of the estimated weights. However, only applying the simulation-based shrinkage rule without trimming portfolio weights makes sure that weights do not become too extreme, but it is not enough to obtain a good performance; this can be seen by the fact that combination 10 is clearly outperformed by the $1/N$ portfolio in all data sets. On the other hand, only trimming portfolio weights is also not enough in a setting that specifically models empirical properties of asset returns, as shown by the fact that combination 6, 7, 8 and 9 are significantly outperformed by the $1/N$ portfolio in four of the seven data sets. Only if the simulation-based shrinkage rule is applied to the trimmed portfolio weights the resulting portfolio becomes somewhat comparable to the $1/N$ portfolio, as combination 11 and 12 are less often significantly outperformed by the $1/N$ portfolio.

Table 8: Sharpe ratios that incorporate transaction costs

portfolio rule	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
$1/N$	0.1038	0.1631	0.1869	0.2155	0.1639	0.1676	0.1753
comb. 1 (mv)	0.0043	-	-0.0984	0.2020	-	-	-
	(0.38)	-	(0.00)	(0.82)	-	-	-
comb. 2	0.0920	0.1028	0.1440	0.2233	0.2031	0.2285	0.3114
	(0.84)	(0.09)	(0.41)	(0.88)	(0.05)	(0.30)	(0.03)
comb. 3	0.1035	0.0864	0.1445	0.1966	0.3043	0.2572	0.3022
	(1.00)	(0.07)	(0.46)	(0.75)	(0.00)	(0.16)	(0.08)
comb. 4	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
comb. 5	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
comb. 6	-0.1333	-0.0320	0.0034	-0.0369	0.1757	0.1711	0.1383
	(0.01)	(0.00)	(0.00)	(0.00)	(0.84)	(0.95)	(0.49)
comb. 7	-0.1093	-0.0508	-0.0142	-0.0377	0.1515	0.1373	0.1249
	(0.00)	(0.00)	(0.00)	(0.00)	(0.81)	(0.63)	(0.34)
comb. 8	-0.0647	0.0145	0.0188	-0.0581	0.1846	0.1546	0.1289
	(0.02)	(0.00)	(0.00)	(0.00)	(0.71)	(0.82)	(0.38)
comb. 9	-0.0904	-0.0011	0.0168	-0.0640	0.1661	0.1317	0.1162
	(0.01)	(0.00)	(0.00)	(0.00)	(0.97)	(0.55)	(0.28)
comb. 10	-0.0460	-0.0653	0.0532	-0.0103	0.0190	0.0952	0.0757
	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)
comb. 11	0.0319	0.1085	0.1516	0.1602	0.1679	0.1476	0.1249
	(0.02)	(0.08)	(0.50)	(0.08)	(0.93)	(0.59)	(0.14)
comb. 12	-0.0044	0.1079	0.1539	0.0573	0.1571	0.1343	0.1075
	(0.01)	(0.09)	(0.52)	(0.00)	(0.83)	(0.33)	(0.03)

Note: This table reports the monthly Sharpe ratios that incorporate transaction costs of the given portfolio rules, where the different combinations are described in Table 2. No values are reported if returns net of transaction costs could not be computed correctly due to extreme portfolio weights. In parentheses are the p-values of the differences in Sharpe ratio with the Sharpe ratio of the $1/N$ strategy.

8.4 Evaluating the performance of dynamic forecasting using copulas

We now compared the different combinations with the $1/N$ strategy and find that combination 2 and 3 perform best. To draw conclusions about which methods improve portfolio performance, we compare the difference in portfolio performance for different features in Table 2 by individually comparing the combinations that differ in that feature and are otherwise identical. Table 9 shows the differences in Sharpe ratio (without incorporating transaction costs) that result from using dynamic modelling. We see that dynamic modelling does not have a significant effect on portfolio performance when transaction costs are not considered, as differences are once significantly positive and once significantly negative and otherwise insignificant. This means that the reduction in (squared) bias of the estimated weights by incorporating specific properties of asset returns is of similar magnitude as the increase in variance of the estimated weights due to the additional parameters that have to be estimated to fit the models for the asset returns.

Table 9: Differences in Sharpe ratios that do not incorporate transaction costs resulting from using dynamic modelling

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
4 minus 1	0.1082	0.0458	0.0041	-0.1583	-0.0245	-0.0178	0.0218
	(0.34)	(0.52)	(0.95)	(0.03)	(0.75)	(0.83)	(0.60)
6 minus 2	-0.0851	0.0279	-0.0209	-0.1025	0.1075	0.1251	0.0123
	(0.38)	(0.61)	(0.72)	(0.10)	(0.06)	(0.02)	(0.78)
7 minus 3	-0.0503	0.0501	-0.0335	-0.0772	0.0249	0.0439	0.0272
	(0.60)	(0.34)	(0.58)	(0.22)	(0.65)	(0.41)	(0.63)

Note: This table reports the differences in monthly Sharpe ratios that do not incorporate transaction costs that are a result of using a dynamic model to simulate returns. The different combinations are described in Table 2. In parentheses are the p-values of the reported differences in Sharpe ratio.

Table 10 shows the differences in Sharpe ratio when transaction costs are incorporated. We see that dynamic modelling significantly impairs portfolio performance when incorporating transaction costs. This can be explained by the fact that dynamic modelling leads to an increase of the variance of the estimated weights, leading to larger changes in portfolio weights from one time period to the next and thus an increase in transaction costs. A second reason for the increase in transaction costs is that dynamic modelling captures changing patterns in level and volatility of the asset returns, leading to larger changes in portfolio weights from one time period to the next. These observations answer our first subquestion by showing that the portfolio performance does not improve by specifically modelling the empirical properties of asset returns due to an increased variance of the estimated weights.

Table 10: Differences in Sharpe ratios that incorporate transaction costs resulting from using dynamic modelling

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
4 minus 1	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
6 minus 2	-0.2253	-0.1348	-0.1406	-0.2602	-0.0274	-0.0574	-0.1731
	(0.02)	(0.02)	(0.02)	(0.00)	(0.65)	(0.34)	(0.00)
7 minus 3	-0.2128	-0.1372	-0.1587	-0.2344	-0.1528	-0.1199	-0.1772
	(0.02)	(0.01)	(0.01)	(0.00)	(0.01)	(0.04)	(0.00)

Note: This table reports the differences in monthly Sharpe ratios that incorporate transaction costs that are a result of using a dynamic model to simulate returns. The different combinations are described in Table 2. No values are reported if returns net of transaction costs could not be computed correctly due to extreme portfolio weights. In parentheses are the p-values of the reported differences in Sharpe ratio.

8.5 Evaluating the performance of trimming portfolio weights

One of the methods that reduce the variance of the estimated weights is to trim portfolio weights. The differences in Sharpe ratio (without incorporating transaction costs) as a consequence of trimming portfolio weights can be found in Table 11. We see that trimming portfolio weights is beneficial; in 21 out of 56 cases it significantly increases the Sharpe ratio, whereas it never decreases the Sharpe ratio significantly. This confirms our earlier finding that trimming portfolio weights reduces the variance of the estimated weights more than that it increases the (squared) bias of the estimated weights. We further see (as expected) that the benefit of trimming portfolio weights is largest when the variance of the estimated weights is larger; that is, when we also apply the simulation-based shrinkage rule to reduce the variance of the estimated weights the added benefit of trimming decreases, although it is still beneficial as overall it increases the Sharpe ratio and 2 out of 14 times this increase is significant.

Table 11: Differences in Sharpe ratios that do not incorporate transaction costs resulting from trimming portfolio weights

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
2 minus 1	0.0327	0.1444	0.0901	0.0159	0.1559	0.2334	0.1282
	(0.65)	(0.07)	(0.18)	(0.40)	(0.04)	(0.13)	(0.04)
3 minus 1	0.0578	0.1415	0.1010	-0.0066	0.2704	0.3093	0.1538
	(0.31)	(0.06)	(0.10)	(0.43)	(0.02)	(0.00)	(0.04)
6 minus 4	-0.1606	0.1265	0.0651	0.0717	0.2879	0.3763	0.1187
	(0.19)	(0.09)	(0.42)	(0.69)	(0.00)	(0.00)	(0.04)
7 minus 4	-0.1008	0.1459	0.0634	0.0746	0.3197	0.3710	0.1592
	(0.32)	(0.05)	(0.42)	(0.21)	(0.00)	(0.00)	(0.00)
8 minus 5	-0.0619	0.1681	0.0880	0.1512	0.0839	0.2509	0.1239
	(0.51)	(0.01)	(0.27)	(0.01)	(0.22)	(0.00)	(0.01)
9 minus 5	-0.0750	0.1813	0.0934	0.1495	0.1172	0.2693	0.1650
	(0.43)	(0.01)	(0.24)	(0.01)	(0.14)	(0.00)	(0.00)
11 minus 10	0.0023	0.0025	0.0452	0.1229	0.0642	0.0740	0.0820
	(0.97)	(0.95)	(0.23)	(0.03)	(0.15)	(0.09)	(0.20)
12 minus 10	-0.0398	0.0063	0.0441	0.0407	0.0796	0.0595	0.0627
	(0.52)	(0.86)	(0.31)	(0.46)	(0.01)	(0.16)	(0.31)

Note: This table reports the differences in monthly Sharpe ratios that do not incorporate transaction costs that are a result of trimming portfolio weights. The different combinations are described in Table 2. In parentheses are the p-values of the reported differences in Sharpe ratio.

Table 12 shows the differences in Sharpe ratio when transaction costs are incorporated. It further illustrates the usefulness of trimming portfolio weights. With one (very insignificant) exception, trimming portfolio weights always increases the portfolio performance after taking transaction costs into account. In many cases the returns net of transaction costs could not be correctly computed without trimming portfolio weights, showing that trimming is necessary to prevent extreme portfolio weights. In the other cases we can quantify the effect of trimming and Table 12 shows us that trimming improves the performance (significantly for 9 out of the 20 cases). Comparing this with Table 11 (in which the difference was only significant in 2 out of these 20 cases) learns us that trimming becomes even more beneficial when transaction costs are incorporated. This can be explained by the fact that trimming the portfolio weights reduces the variance of the estimated weights, so that the portfolio weights change less over time which decreases the transaction costs relative to the methods that do not trim portfolio weights. We can thus answer our second subquestion by concluding that trimming portfolio weights is an effective strategy to increase portfolio performance.

Table 12: Differences in Sharpe ratios that incorporate transaction costs resulting from trimming portfolio weights

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
2 minus 1	0.0877	-	0.2424	0.0213	-	-	-
	(0.22)	-	(0.00)	(0.26)	-	-	-
3 minus 1	0.0992	-	0.2429	-0.0054	-	-	-
	(0.08)	-	(0.00)	(0.53)	-	-	-
6 minus 4	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
7 minus 4	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
8 minus 5	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
9 minus 5	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
11 minus 10	0.0779	0.1737	0.0984	0.1705	0.1489	0.0525	0.0492
	(0.29)	(0.00)	(0.04)	(0.00)	(0.00)	(0.24)	(0.43)
12 minus 10	0.0416	0.1732	0.1007	0.0676	0.1381	0.0391	0.0319
	(0.42)	(0.00)	(0.04)	(0.20)	(0.00)	(0.38)	(0.59)

Note: This table reports the differences in monthly Sharpe ratios that incorporate transaction costs that are a result of trimming portfolio weights. The different combinations are described in Table 2. No values are reported if returns net of transaction costs could not be computed correctly due to extreme portfolio weights. In parentheses are the p-values of the reported differences in Sharpe ratio.

In order to evaluate whether there is a preference between trimming with a fixed trimming threshold of zero or a trimming threshold that is selected using a data-driven approach we look at Tables 13 and 14. Table 13 shows that using a data-driven approach is slightly preferred when transaction costs are not taken into account, whereas Table 14 shows that we do not have a clear preference when transaction costs are incorporated. This implies that (compared with a fixed trimming threshold of $c = 0$) the data-driven approach that allows values of c up to $c_{\max} = 1$ leads to a decrease in the (squared) bias of the estimated weights that is slightly larger than the increase in the variance of the estimated weights. However, this slight increase in performance disappears when transaction costs are incorporated because a trimming threshold of zero reduces the range of possible portfolio weights more than a larger trimming threshold, which decreases changes of portfolio weights over time and thus lowers transaction costs. These results show that although it is important to trim portfolio weights, there is no clear indication whether we prefer to use a fixed trimming threshold of 0 or the data-driven approach with $c_{\max} = 1$.

Table 13: Differences in Sharpe ratios that do not incorporate transaction costs resulting from using a data-driven threshold (rather than a threshold of 0) for trimming

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
3 minus 2	0.0250	-0.0029	0.0109	-0.0225	0.1145	0.0759	0.0256
	(0.56)	(0.85)	(0.44)	(0.14)	(0.00)	(0.01)	(0.63)
7 minus 6	0.0598	0.0194	-0.0017	0.0029	0.0319	-0.0053	0.0406
	(0.34)	(0.28)	(0.88)	(0.87)	(0.37)	(0.70)	(0.00)
9 minus 8	-0.0130	0.0133	0.0055	-0.0017	0.0333	0.0184	0.0411
	(0.34)	(0.45)	(0.66)	(0.92)	(0.34)	(0.47)	(0.00)
12 minus 11	-0.0421	0.0038	-0.0011	-0.0822	0.0154	-0.0145	-0.0193
	(0.42)	(0.88)	(0.95)	(0.05)	(0.72)	(0.71)	(0.48)

Note: This table reports the differences in monthly Sharpe ratios that do not incorporate transaction costs that are a result of using a data-driven trimming threshold (instead of a fixed threshold of 0). The different combinations are described in Table 2. In parentheses are the p-values of the reported differences in Sharpe ratio.

Table 14: Differences in Sharpe ratios that incorporate transaction costs resulting from using a data-driven threshold (rather than a threshold of 0) for trimming

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
3 minus 2	0.0115	-0.0164	0.0005	-0.0267	0.1012	0.0286	-0.0092
	(0.80)	(0.32)	(0.97)	(0.09)	(0.00)	(0.32)	(0.83)
7 minus 6	0.0240	-0.0187	-0.0176	-0.0008	-0.0242	-0.0338	-0.0133
	(0.68)	(0.30)	(0.14)	(0.97)	(0.50)	(0.02)	(0.23)
9 minus 8	-0.0257	-0.0157	-0.0019	-0.0059	-0.0185	-0.0229	-0.0126
	(0.05)	(0.35)	(0.88)	(0.77)	(0.60)	(0.32)	(0.31)
12 minus 11	-0.0363	-0.0006	0.0023	-0.1029	-0.0108	-0.0133	-0.0173
	(0.51)	(0.98)	(0.91)	(0.02)	(0.80)	(0.72)	(0.51)

Note: This table reports the differences in monthly Sharpe ratios that incorporate transaction costs that are a result of using a data-driven trimming threshold (instead of a fixed threshold of 0). The different combinations are described in Table 2. In parentheses are the *p*-values of the reported differences in Sharpe ratio.

8.6 Evaluating the performance of the simulation-based shrinkage rule

We finally evaluate whether shrinking portfolio weights towards equal weights when the variance of the estimated weights is high can improve the portfolio performance. Table 15 shows the differences in Sharpe ratio when transaction costs are not incorporated that are a result of using our simulation-based shrinkage rule. For the methods that do not trim portfolio weights the differences are not significant, although there appears to be a tendency to improve the portfolio performance slightly when applying the simulation-based shrinkage rule. For methods that trim portfolio weights we observe a significant decrease of the portfolio performance. This seems to suggests that using our simulation-based shrinkage rule decreases the variance of the estimated weights to a smaller extent than that it increases the (squared) bias of the estimated weights in the cases where the variance of the estimated weights is already decreased through trimming portfolio weights. This would then imply that portfolio returns are not necessarily bad when estimation uncertainty (as measured by the variance sum over the replications) is high, so that the simulation-based shrinkage method would be less beneficial. However, Table 15 reports the differences in Sharpe ratio resulting from using our simulation-based shrinkage rule that selects shrinkage thresholds with a data-driven approach that maximises the Sharpe ratio that incorporates transaction costs. In fact, we can see from Appendix K that selecting the shrinkage threshold with a data-driven approach that maximises the Sharpe ratio that does not correct for transaction costs decreases the differences in Sharpe ratio without transaction costs such that they are not significant anymore.

Table 15: Differences in Sharpe ratios that do not incorporate transaction costs resulting from using simulation-based shrinkage that incorporates transaction costs

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
10 minus 5	-0.0637	0.1402	0.0949	0.1434	-0.0683	0.0696	-0.0533
	(0.48)	(0.06)	(0.29)	(0.06)	(0.42)	(0.42)	(0.42)
11 minus 8	0.0006	-0.0253	0.0521	0.1151	-0.0880	-0.1073	-0.0952
	(0.99)	(0.52)	(0.22)	(0.04)	(0.04)	(0.02)	(0.03)
12 minus 9	-0.0285	-0.0348	0.0456	0.0346	-0.1059	-0.1402	-0.1557
	(0.70)	(0.39)	(0.29)	(0.42)	(0.01)	(0.01)	(0.00)

Note: This table reports the differences in monthly Sharpe ratios that do not incorporate transaction costs that are a result of applying the simulation-based shrinkage rule with shrinkage thresholds selected using the data-driven approach that incorporates transaction costs. The different combinations are described in Table 2. In parentheses are the p-values of the reported differences in Sharpe ratio.

Table 16 shows the differences in Sharpe ratio when transaction costs are incorporated that are a result of using our simulation-based shrinkage rule. We observe that using simulation-based shrinkage is necessary for methods that do not trim portfolio weights, as otherwise the portfolio takes such extreme positions that returns net of transaction costs cannot even be computed correctly. For methods that trim portfolio weights we see that applying the simulation-based shrinkage rule improves the Sharpe ratios that incorporate transaction costs significantly in three out of the seven data sets, whereas the performance increases insignificantly or stays similar in the other four data sets. This can be explained by noting that our simulation-based shrinkage rule reduces the variance of the estimated weights. This leads to smaller changes in portfolio weights over time, so that transaction costs decrease which increases the portfolio performance that incorporates transaction costs. We thus find as an answer to the third subquestion that applying the simulation-based shrinkage rule is beneficial, once again confirming that the increased variance of the estimated weights is an important drawback of the dynamic modelling approach.

Table 16: Differences in Sharpe ratios that incorporate transaction costs resulting from using simulation-based shrinkage that incorporates transaction costs

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
10 minus 5	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
11 minus 8	0.0967	0.0940	0.1328	0.2183	-0.0167	-0.007	-0.004
	(0.25)	(0.02)	(0.00)	(0.00)	(0.70)	(0.88)	(0.93)
12 minus 9	0.0860	0.1090	0.1370	0.1213	-0.0090	0.0026	-0.0087
	(0.26)	(0.01)	(0.00)	(0.02)	(0.82)	(0.96)	(0.87)

Note: This table reports the differences in monthly Sharpe ratios that incorporate transaction costs that are a result of applying the simulation-based shrinkage rule with shrinkage thresholds selected using the data-driven approach that incorporates transaction costs. The different combinations are described in Table 2. No values are reported if returns net of transaction costs could not be computed correctly due to extreme portfolio weights. In parentheses are the p-values of the reported differences in Sharpe ratio.

As a last remark, we note that we have compared the results of the simulation-based shrinkage rule relative to the results of the dynamic model that takes average weights over multiple replications in order to isolate the effect of the shrinkage rule. Appendix L shows the differences in performance as a result of taking average weights. It shows that taking average weights over multiple simulations slightly increases the portfolio performance. This makes sense as taking average weights is likely to reduce the variance of the estimated weights because possible outliers are averaged out, whereas taking average weights is not expected to change the bias. This confirms that applying the simulation-based shrinkage rule also outperforms a method that does not take average weights.

9 Conclusion

We investigate whether recent and new techniques are able to improve portfolio performance. We first investigate whether we can improve portfolio performance by specifically modelling the empirical properties of asset returns. To do this we estimate dynamic models that capture marginal properties of asset returns and model the dependence structure using a copula model. We hereby find that specifically modelling empirical properties of asset returns does not improve the portfolio performance as a consequence of an increased variance of the estimated weights.

After that we evaluate whether trimming portfolio weights is an effective strategy to reduce the variance of the estimated weights and thereby improve portfolio performance. We hereby find that our modified trimming method that estimates mean-variance weights while restricting the sum of absolute weights increases portfolio performance significantly. Moreover, we find that our modified trimming method becomes more useful when estimation uncertainty is larger.

As a last step of our research we evaluate whether shrinking portfolio weights towards equal weights when the variance of the estimated weights is high is able to improve portfolio performance. To do so we develop a flexible and robust simulation-based shrinkage method that can be applied when forming portfolios using simulated returns. We find that our simulation-based shrinkage rule indeed improves portfolio performance, once again confirming that the increased variance of the estimated weights is an important drawback of the dynamic modelling approach.

Based on the answers to these subquestions we can answer our research question. We find that some of the recent and new techniques are able to improve portfolio performance. Although dynamic modelling of specific asset return properties is not able to improve performance, we find that our methods that reduce the variance of the estimated weights are able to do so. Especially our modified trimming rule is beneficial and even significantly outperforms the $1/N$ benchmark in some data sets whereas it is never significantly outperformed. Moreover, these results also hold when transaction costs are taken into account. This implies that we have extended the portfolio optimisation literature with two effective ways to reduce estimation uncertainty. From a practical point of view our research implies that it is beneficial for investors to apply our modified trimming rule in order improve their portfolio performance.

We finally point out potential limitations of our research. We work with the same data sets as DeMiguel et al. (2009), which all consist of portfolios of stocks rather than individual stocks. As explained in their paper, this means that the idiosyncratic volatility of the assets in these data sets is lower than if we would use individual assets, so that the loss of the 'naive' $1/N$ allocation is relatively small. This can explain why none of our methods consistently beats the $1/N$ portfolio and might also be a reason for the good performance of our simulation-based shrinkage method. We therefore provide two directions for further research. As a first idea it might be good to also evaluate data sets consisting of individual asset returns to see whether this leads to the same conclusions as our research. Additionally, it might be interesting to evaluate other shrinkage targets than the $1/N$ portfolio (such as using a portfolio based on our modified trimming method), especially if these methods would be able to outperform the $1/N$ portfolio for other investigated data sets.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of financial Economics*, 63(3):443–494.

Bain, L. J. and Engelhardt, M. (2016). *Introduction to probability and mathematical statistics, custom edition*. Cengage.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1):245–268.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

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

Dißmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

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

Fantazzini, D. (2009). The effects of misspecified marginals and copulas on computing the value at risk: A monte carlo study. *Computational Statistics & Data Analysis*, 53(6):2168–2188.

Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the american statistical association*, 93(441):359–371.

Golosnoy, V. and Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *The European Journal of Finance*, 13(5):441–458.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

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

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

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51(6):2836–2850.

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

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

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

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The journal of finance*, 56(2):649–676.

Low, R. K. Y., Alcock, J., Faff, R., and Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it? *Journal of Banking & Finance*, 37(8):3085–3099.

Low, R. K. Y., Faff, R., and Aas, K. (2016). Enhancing mean–variance portfolio selection by modeling distributional asymmetries. *Journal of Economics and Business*, 85:49–72.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Morales Napoles, O., Cooke, R. M., and Kurowicka, D. (2010). About the number of vines and regular vines on n nodes.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013). On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics*, 173(2):189–204.

Radchenko, P., Vasnev, A. L., and Wang, W. (2020). Too similar to combine? on negative weights in forecast combination. *On Negative Weights in Forecast Combination (July 1, 2020).*

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460.

Stoyanov, S. V., Rachev, S. T., Racheva-Yotova, B., and Fabozzi, F. J. (2011). Fat-tailed models for risk estimation. *The Journal of Portfolio Management*, 37(2):107–117.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333.

Appendix

A Detailed explanation of modelling with R-vines

To explain in more detail how modelling with R-vines works, we first formalise what an R-vine exactly is. Dißmann et al. (2013) provide the following formal definition:

Definition 1. *(R-vine).* $\mathcal{V} = (T_1, ..., T_{n-1})$ *is an R-vine on* n *elements if*

(i) T_1 *is a tree with nodes* $N_1 = \{1, ..., n\}$ *and a set of edges denoted by* E_1*.*

(ii) For $i = 2, ..., n-1$*,* T_i *is a tree with nodes* $N_i = E_{i-1}$ *and edge set* E_i*.*

(iii) For $i = 2, ..., n-1$ *and* $\{a, b\} \in E_i$ *with* $a = \{a_1, a_2\}$ *and* $b = \{b_1, b_2\}$ *it must hold that* $\#(a \cap b) = 1$*, with* $\#$ *denoting the cardinality of a set.*

In words, an R-vine on n elements is a nested set of $n-1$ trees such that the edges in a given tree i become the nodes in the subsequent tree $i+1$. The last condition therefore simply states that two nodes in a given tree i can only be connected by an edge if these nodes (which are edges in the previous tree by the second condition) share a node in the previous tree.

The nodes in the first tree correspond with all the variables of which we want to model the copula density. (In our case these are the N asset returns after they are transformed to standard uniform variables; we explain in more detail how this can be done in Section 4.3.) If for example the nodes i and j, the nodes j and k and the nodes k and l are connected by an edge in the first tree, then this indicates that we use pairwise copulas to model the dependence between the variables i and j, between the variables j and k and between the variables k and l. The nodes in the second tree are the edges of the first tree, so that $\{i, j\}$, $\{j, k\}$ and $\{k, l\}$ become nodes in the second tree. We then denote the corresponding edges between nodes $\{i, j\}$ and $\{j, k\}$ and between nodes $\{j, k\}$ and $\{k, l\}$ (assuming for this example that these nodes are indeed connected by an edge) by $\{i, k|j\}$ and $\{j, l|k\}$. (We discuss in Section 4.3 how to determine which nodes are connected by an edge.) Hence, if two nodes in the second tree can be connected by an edge, this means that one node (j respectively k in this example) is connected by an edge to two other nodes in the first tree. Moreover, the edge $\{i, k|j\}$ in the second tree indicates that we use a pairwise copula to model the dependence between variable i and k, conditional on the variable j. In the same way the edge $\{j, l|k\}$ in the second tree indicates that we use a pairwise copula to model the dependence between variable j and l, conditional on the variable k. Subsequently, the edges $\{i,k|j\}$ and $\{j,l|k\}$ become nodes in the third tree, which we again assume to be connected by the edge $\{i,l|j,k\}$. This edge indicates that we use a pairwise copula to model the dependence between variable i and l, conditional on the variables j and k. This example illustrates that the edges in the first tree already determine for which variables the (unconditional) dependence is modelled with a pairwise copula and on which variables we can possibly condition when modelling dependencies in later trees.

B Derivation of the standardised skewed Student t density

The density of the standardised skewed Student t distribution is given by

$$f_Z(z;\nu,\lambda) = \frac{2\sigma}{\lambda+\frac{1}{\lambda}}\left\{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\left(\frac{\mu+\sigma z}{\lambda}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} I_{[-\mu/\sigma,\infty)}(z) + \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\lambda^2(\mu+\sigma z)^2}{\nu}\right)^{-\frac{\nu+1}{2}} I_{(-\infty,-\mu/\sigma)}(z)\right\}, \tag{21}$$

with

$$\mu = \frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)}\cdot\frac{\lambda^4-1}{\lambda^3+\lambda} \tag{22}$$

and

$$\sigma = \sqrt{\frac{\nu}{\nu-2}\cdot\frac{\lambda^6+1}{\lambda^4+\lambda^2}-\left(\frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)}\cdot\frac{\lambda^4-1}{\lambda^3+\lambda}\right)^2}, \tag{23}$$

where $\Gamma(\cdot)$ denotes a gamma function and $\mathrm{B}(\cdot,\cdot)$ denotes a beta function. This density can be obtained by using the method of Fernández and Steel (1998) to introduce skewness in a (symmetric) Student t distribution, and subsequently using the mean and standard deviation of this skewed Student t distribution to obtain the density of the standardised version of this distribution via a transformation method.

We can use the method of Fernández and Steel (1998) as the density of a Student t distribution (which we denote by $f(\cdot;\nu)$) is unimodal and symmetric around zero, meaning that $f(s;\nu) = f(|s|;\nu)$ and that $f(|s|;\nu)$ is a decreasing function in $|s|$. This method yields that a random variable X that follows a skewed Student t distribution has a density of

$$f_X(x;\nu,\lambda) = \frac{2}{\lambda+\frac{1}{\lambda}}\left\{f\left(\frac{x}{\lambda};\nu\right)I_{[0,\infty)}(x) + f(\lambda x;\nu)\,I_{(-\infty,0)}(x)\right\}.$$

To obtain the density function of a standardised skewed Student t distribution we let $Z := u(X) = \frac{X-\mu}{\sigma}$, with μ and σ the mean and standard deviation of X, and we derive the distribution of Z.

We have $X = \mu + \sigma Z = u^{-1}(Z)$ and we define $w(Z) \coloneqq u^{-1}(Z)$. Using the continuous case of the transformation method in Bain and Engelhardt (2016) gives

$$\begin{aligned}
f_Z(z;\nu,\lambda) &= f_X(w(z);\nu,\lambda)\left|\frac{d}{dz}w(z)\right| \\
&= \frac{2}{\lambda+\frac{1}{\lambda}}\left\{f\left(\frac{w(z)}{\lambda};\nu\right)I_{[0,\infty)}(w(z)) + f\left(\lambda w(z);\nu\right)I_{(-\infty,0)}(w(z))\right\}\left|\frac{d}{dz}w(z)\right| \\
&= \frac{2}{\lambda+\frac{1}{\lambda}}\left\{f\left(\frac{\mu+\sigma z}{\lambda};\nu\right)I_{[0,\infty)}(\mu+\sigma z) + f\left(\lambda(\mu+\sigma z);\nu\right)I_{(-\infty,0)}(\mu+\sigma z)\right\}|\sigma| \\
&= \frac{2\sigma}{\lambda+\frac{1}{\lambda}}\left\{f\left(\frac{\mu+\sigma z}{\lambda};\nu\right)I_{[0,\infty)}(\mu+\sigma z) + f\left(\lambda(\mu+\sigma z);\nu\right)I_{(-\infty,0)}(\mu+\sigma z)\right\} \\
&= \frac{2\sigma}{\lambda+\frac{1}{\lambda}}\Bigg\{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\left(\frac{\mu+\sigma z}{\lambda}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}I_{[-\mu/\sigma,\infty)}(z) + \\
&\qquad \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\lambda^2(\mu+\sigma z)^2}{\nu}\right)^{-\frac{\nu+1}{2}}I_{(-\infty,-\mu/\sigma)}(z)\Bigg\},
\end{aligned}$$

where we plugged in the Student t density function in the last step, which confirms Equation (21).

To derive an expression for the parameters μ and σ, we still assume that X follows a skewed Student t distribution, so that it follows from Fernández and Steel (1998) that we have for positive integers r that

$$\mathrm{E}\left[X^r;\lambda\right] = M_r\frac{\lambda^{r+1}+\frac{(-1)^r}{\lambda^{r+1}}}{\lambda+\frac{1}{\lambda}},$$

where

$$M_r = \int_0^\infty s^r 2f(s;\nu)ds,$$

with $f(\cdot;\nu)$ again denoting the density function of a Student t distribution. In order to derive expressions for $\mu = \mathrm{E}\left[X;\lambda\right]$ and $\sigma = \sqrt{\mathrm{var}\left(X;\lambda\right)}$ we need to find expressions for M_1 and M_2. For M_1 we get

$$M_1 = \int_0^\infty s2f(s;\nu)ds = 2\int_0^\infty sf(s;\nu)ds = 2\int_0^\infty s\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}}ds.$$

To work this out, we substitute $p = \frac{s^2}{\nu}$, so that $s = \sqrt{\nu p}$, $ds = \frac{1}{2}\sqrt{\frac{\nu}{p}}dp$, $p \to \infty$ when $s \to \infty$ and $p \to 0$ when $s \to 0$. This gives

$$M_1 = 2\int_0^\infty \sqrt{\nu p}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}(1+p)^{-\frac{\nu+1}{2}}\frac{1}{2}\sqrt{\frac{\nu}{p}}dp = \frac{\sqrt{\nu}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty (1+p)^{-\frac{\nu+1}{2}}dp.$$

We now substitute $u = 1+p$, so that $dp = du$, $u \to \infty$ when $p \to \infty$ and $u \to 1$ when $p \to 0$. This gives

$$\begin{aligned}M_1 &= \frac{\sqrt{\nu}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_1^\infty u^{-\frac{\nu+1}{2}}du = \frac{\sqrt{\nu}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{1}{\left(-\frac{\nu+1}{2}+1\right)}u^{-\frac{\nu+1}{2}+1}\right]_{u=1}^{u=\infty} \\ &= \frac{\sqrt{\nu}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{1}{\left(\frac{1-\nu}{2}\right)}u^{\frac{1-\nu}{2}}\right]_{u=1}^{u=\infty} = \frac{\sqrt{\nu}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\cdot\frac{2}{1-\nu}\cdot(0-1) \\ &= \frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} = \frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{\Gamma\left(\frac{1}{2}+\frac{\nu}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} = \frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)},\end{aligned} \tag{24}$$

where we used in the fourth equality that $\nu > 2$ and we used the well-known relation between the beta function and the gamma function in the last equality.

For M_2, we have

$$M_2 = \int_0^\infty s^2 2f(s;\nu)ds = 2\int_0^\infty s^2 f(s;\nu)ds = \int_{-\infty}^\infty s^2 f(s;\nu)ds = \frac{\nu}{\nu-2}, \tag{25}$$

where we used in the third equality that the function $g(s) = s^2 f(s;\nu)$ is an even function (so $g(-s) = g(s)$). This follows as $f(s;\nu)$ is a symmetric function around zero and hence an even function, so that $g(s)$ is the product of two even functions and therefore also an even function. Furthermore, we used in the last equality that the second moment of a Student t distribution is equal to $\frac{\nu}{\nu-2}$.

We can now substitute the expressions of M_1 and M_2 from Equations (24) and (25) to get

$$\mathrm{E}\left[X;\lambda\right] = M_1\frac{\lambda^2+\frac{(-1)^1}{\lambda^2}}{\lambda+\frac{1}{\lambda}} = \frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)}\cdot\frac{\lambda^2-\frac{1}{\lambda^2}}{\lambda+\frac{1}{\lambda}} = \frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)}\cdot\frac{\lambda^4-1}{\lambda^3+\lambda},$$

and

$$\mathrm{E}\left[X^2;\lambda\right] = M_2\frac{\lambda^3+\frac{(-1)^2}{\lambda^3}}{\lambda+\frac{1}{\lambda}} = \frac{\nu}{\nu-2}\cdot\frac{\lambda^3+\frac{1}{\lambda^3}}{\lambda+\frac{1}{\lambda}} = \frac{\nu}{\nu-2}\cdot\frac{\lambda^6+1}{\lambda^4+\lambda^2},$$

so that we get

$$\mu = \mathrm{E}\left[X;\lambda\right] = \frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)}\cdot\frac{\lambda^4-1}{\lambda^3+\lambda},$$

and

$$\begin{aligned}\sigma &= \sqrt{\mathrm{var}\left(X;\lambda\right)} = \sqrt{\mathrm{E}\left[X^2;\lambda\right]-\left(\mathrm{E}\left[X;\lambda\right]\right)^2} \\ &= \sqrt{\frac{\nu}{\nu-2}\cdot\frac{\lambda^6+1}{\lambda^4+\lambda^2}-\left(\frac{2\sqrt{\nu}}{\nu-1}\cdot\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)}\cdot\frac{\lambda^4-1}{\lambda^3+\lambda}\right)^2},\end{aligned}$$

which confirms Equations (22) and (23).

C Illustration of negative weights in the mean-variance portfolio

We illustrate with a simple example that small differences in the expected excess returns of highly correlated assets can lead to large (negative) portfolio weights. We assume that we have $N = 3$ risky assets at time t, with expected excess return vector $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$ given by

$$\boldsymbol{\mu}_t = \begin{pmatrix} -0.01 \\ 0.01 \\ 0.01 \end{pmatrix}, \quad \boldsymbol{\Sigma}_t = \begin{pmatrix} 0.010 & 0.009 & 0 \\ 0.009 & 0.010 & 0 \\ 0 & 0 & 0.010 \end{pmatrix}.$$

That is, all assets have the same volatility and only asset 1 and 2 have a strong positive correlation. The first-order condition in Equation (13) yields the system

$$\begin{cases} 0.010x_{1,t} + 0.009x_{2,t} = \frac{-0.01}{\gamma} \\ 0.009x_{1,t} + 0.010x_{2,t} = \frac{0.01}{\gamma} \\ 0.010x_{3,t} = \frac{0.01}{\gamma} \end{cases},$$

which is equivalent to the system

$$\begin{cases} x_{1,t} + 0.9x_{2,t} = \frac{-1}{\gamma} \\ 0.9x_{1,t} + x_{2,t} = \frac{1}{\gamma} \\ x_{3,t} = \frac{1}{\gamma} \end{cases}.$$

Solving this system yields

$$\boldsymbol{x}_t = \begin{pmatrix} -\frac{10}{\gamma} \\ \frac{10}{\gamma} \\ \frac{1}{\gamma} \end{pmatrix},$$

so that the relative portfolio weights in a portfolio that only consists of risky assets are given by

$$\boldsymbol{w}_t = \frac{\boldsymbol{x}_t}{|\boldsymbol{\iota}'\boldsymbol{x}_t|} = \gamma \begin{pmatrix} -\frac{10}{\gamma} \\ \frac{10}{\gamma} \\ \frac{1}{\gamma} \end{pmatrix} = \begin{pmatrix} -10 \\ 10 \\ 1 \end{pmatrix}.$$

We thus see that asset 1 receives a large negative weight because it has a strong positive correlation with asset 2 but a lower expected return. We further see that asset 2 has the same volatility and expected return as asset 3, but because of its strong correlation with an asset with a lower expected return asset 2 receives a much larger weight in the mean-variance portfolio.

D Proof that trimming rule TR1 decreases the absolute weights

When negative or positive weights are trimmed they trivially decrease in absolute value, after which they are rescaled by the scaling factor α_1. Hence, to prove that trimming rule TR1 reduces the absolute value of the estimated portfolio weights, it remains to show that $\alpha_1 \in (0,1]$ and $\alpha_1^* \in (0,1]$, with $\alpha_1 \in (0,1)$ and $\alpha_1^* \in (0,1)$ if at least one weight is trimmed, so that both the trimmed and non-trimmed weights decrease in absolute value.

The scaling factor α_1 has to make sure that all weights after trimming again sum to 1 and is therefore given by

$$\alpha_1 = \frac{1}{\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \geq -c\}} \hat{w}_{j,t^*} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot (-c)}.$$

As

$$\begin{aligned} \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \geq -c\}} \hat{w}_{j,t^*} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot (-c) &\geq \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \geq -c\}} \hat{w}_{j,t^*} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \hat{w}_{j,t^*} \\ &= \sum_{j=1}^{N} \hat{w}_{j,t^*} = 1, \end{aligned} \tag{26}$$

it follows that $\alpha_1 \in (0,1]$. Moreover, if at least one weight is trimmed the inequality in Equation (26) becomes a strict inequality so that $\alpha_1 \in (0,1)$.

We now show in a similar way that $\alpha_1^* \in (0,1]$ and $\alpha_1^* \in (0,1)$ if at least one weight is trimmed. The scaling factor α_1^* has to make sure that all weights after trimming again sum to -1 and is therefore given by

$$\alpha_1^* = \frac{-1}{\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \leq c\}} \hat{w}_{j,t^*} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} > c\}} c}.$$

As

$$\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \leq c\}} \hat{w}_{j,t^*} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} > c\}} c \leq \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \leq c\}} \hat{w}_{j,t^*} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} > c\}} \hat{w}_{j,t^*} = \sum_{j=1}^{N} \hat{w}_{j,t^*} = -1, \tag{27}$$

it follows that $\alpha_1^* \in (0,1]$. Moreover, if at least one weight is trimmed the inequality in Equation (27) becomes a strict inequality so that $\alpha_1^* \in (0,1)$.

E Proof that trimming rule TR2 decreases the absolute weights

When negative or positive weights are trimmed they trivially decrease in absolute value. Hence, to prove that trimming rule TR2 reduces the absolute value of the estimated portfolio weights, it remains to show that the non-trimmed weights also decrease in absolute value. We do this by showing that the scaling factors $\alpha_2 \in (0,1]$ and $\alpha_2^* \in (0,1]$, with $\alpha_2 \in (0,1)$ and $\alpha_2^* \in (0,1)$ if at least one weight is trimmed.

The scaling factor α_2 has to make sure that all untrimmed weights (that have a sum of $\sum_{j=1}^N I_{\{\hat{w}_{j,t^*} \geq -c\}} \hat{w}_{j,t^*}$) will sum together with the trimmed weights to 1. This means that the untrimmed weights should sum to

$$1 - \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} < -c\}} \cdot (-c) = 1 + c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} < -c\}}$$

after being multiplied by α_2. Therefore, the scaling factor α_2 is given by

$$\alpha_2 = \frac{1 + c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} < -c\}}}{\sum_{j=1}^N I_{\{\hat{w}_{j,t^*} \geq -c\}} \hat{w}_{j,t^*}}.$$

As $c \geq 0$ and $I_{\{\hat{w}_{j,t^*} < -c\}} \in \{0,1\}$, the numerator of α_2 is at least equal to 1. The denominator is also at least equal to 1, as the originally estimated weights sum to 1 and we now only omit negative weights (when present). These two observations combined imply that $\alpha_2 > 0$. To prove that $\alpha_2 \leq 1$ we rewrite the numerator as

$$\begin{aligned}
1 + c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} < -c\}} &= \sum_{j=1}^N \hat{w}_{j,t^*} + c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} < -c\}} \\
&= \sum_{j=1}^N \hat{w}_{j,t^*} \left(I_{\{\hat{w}_{j,t^*} \geq -c\}} + I_{\{\hat{w}_{j,t^*} < -c\}} \right) + c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} < -c\}} \\
&= \sum_{j=1}^N \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \geq -c\}} + \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} < -c\}} \left(\hat{w}_{j,t^*} + c \right) \\
&\leq \sum_{j=1}^N \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \geq -c\}},
\end{aligned} \tag{28}$$

where the inequality follows as $I_{\{\hat{w}_{j,t^*} < -c\}} \left(\hat{w}_{j,t^*} + c \right)$ is either 0 if $\hat{w}_{j,t^*} \geq -c$ or strictly negative if $\hat{w}_{j,t^*} < -c$. Equation (28) shows that the numerator is smaller than the denominator, which implies that $\alpha_2 \in (0,1]$ as both the numerator and denominator are at least equal to 1. Moreover, if at least one weight is trimmed the inequality in Equation (28) becomes a strict inequality so that $\alpha_2 \in (0,1)$.

We now show in a similar way that $\alpha_2^* \in (0,1]$ and $\alpha_2^* \in (0,1)$ if at least one weight is trimmed. The scaling factor α_2^* has to make sure that all untrimmed weights (that have a sum of $\sum_{j=1}^N I_{\{\hat{w}_{j,t^*} \leq c\}} \hat{w}_{j,t^*}$) will sum together with the trimmed weights to -1. This means that the untrimmed weights should sum to

$$-1-\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}>c\}} \cdot c = -1 - c\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}>c\}}$$

after being multiplied by α_2^*. Therefore, the scaling factor α_2^* is given by

$$\alpha_2^* = \frac{-1 - c\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}>c\}}}{\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}\leq c\}}\hat{w}_{j,t^*}}.$$

As $c \geq 0$ and $I_{\{\hat{w}_{j,t^*}<-c\}} \in \{0,1\}$, the numerator of α_2 is at most equal to -1. The denominator is also at most equal to -1, as the originally estimated weights sum to -1 and we now only omit positive weights (when present). These two observations combined imply that $\alpha_2^* > 0$. To prove that $\alpha_2^* \leq 1$ we rewrite the numerator as

$$\begin{aligned}
-1 - c\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}>c\}} &= \sum_{j=1}^{N} \hat{w}_{j,t^*} - c\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}>c\}} \\
&= \sum_{j=1}^{N} \hat{w}_{j,t^*}\left(I_{\{\hat{w}_{j,t^*}\leq c\}} + I_{\{\hat{w}_{j,t^*}>c\}}\right) - c\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}>c\}} \\
&= \sum_{j=1}^{N} \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*}\leq c\}} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*}>c\}}\left(\hat{w}_{j,t^*} - c\right) \\
&\geq \sum_{j=1}^{N} \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*}\leq c\}},
\end{aligned} \tag{29}$$

where the inequality follows as $I_{\{\hat{w}_{j,t^*}>c\}}\left(\hat{w}_{j,t^*} - c\right)$ is either 0 if $\hat{w}_{j,t^*} \leq c$ or strictly positive if $\hat{w}_{j,t^*} > c$. Equation (29) shows that the numerator is larger than the denominator, which implies that $\alpha_2^* \in (0,1]$ as both the numerator and denominator are at most equal to -1. Moreover, if at least one weight is trimmed the last inequality in Equation (29) becomes a strict inequality so that $\alpha_2^* \in (0,1)$.

F Proof that trimming rule TR3 decreases the absolute weights

When negative or positive weights are trimmed they trivially decrease in absolute value. Hence, to prove that trimming rule TR3 reduces the absolute value of the estimated portfolio weights, it remains to show that the non-trimmed weights also decrease in absolute value. We do this by showing that the scaling factors $\alpha_3 \in (0,1]$ and $\alpha_3^* \in (0,1]$, with $\alpha_3 \in (0,1)$ and $\alpha_3^* \in (0,1)$ if at least one weight is trimmed.

The scaling factor α_3 has to make sure that all untrimmed weights (that have a sum of $\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \geq -c\}} \hat{w}_{j,t^*}$) will sum together with the trimmed weights to 1. This means that the untrimmed weights should sum to

$$1 - \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \left(\frac{-c}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \cdot \hat{w}_{j,t^*} \right) = 1 + c \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}}$$

after being multiplied by α_3. Therefore, the scaling factor α_3 is given by

$$\alpha_3 = \frac{1 + c \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}}}{\sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} \geq -c\}} \hat{w}_{j,t^*}}.$$

As $c \geq 0$ and $I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \in [0, 1]$, the numerator of α_3 is at least equal to 1. The denominator is also at least equal to 1, as the originally estimated weights sum to 1 and we now only omit negative weights (when present). These two observations combined imply that $\alpha_3 > 0$. To prove that $\alpha_3 \leq 1$ we rewrite the numerator as

$$\begin{aligned}
1 + c \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} &= \sum_{j=1}^{N} \hat{w}_{j,t^*} + c \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \\
&= \sum_{j=1}^{N} \hat{w}_{j,t^*} \left(I_{\{\hat{w}_{j,t^*} \geq -c\}} + I_{\{\hat{w}_{j,t^*} < -c\}} \right) + \\
&\qquad c \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \\
&= \sum_{j=1}^{N} \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \geq -c\}} + \\
&\qquad \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \left(\hat{w}_{j,t^*} + c \cdot \frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \right) \\
&\leq \sum_{j=1}^{N} \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \geq -c\}} + \sum_{j=1}^{N} I_{\{\hat{w}_{j,t^*} < -c\}} \left(\hat{w}_{j,t^*} + c \right) \\
&\leq \sum_{j=1}^{N} \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \geq -c\}},
\end{aligned} \tag{30}$$

where the first inequality follows because $c \geq 0$ and $\frac{\hat{w}_{j,t^*}}{\min\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \in (0, 1]$ for $\hat{w}_{j,t^*} < -c$ and where the second inequality follows as $I_{\{\hat{w}_{j,t^*} < -c\}} (\hat{w}_{j,t^*} + c)$ is either 0 if $\hat{w}_{j,t^*} \geq -c$ or strictly negative if $\hat{w}_{j,t^*} < -c$. Equation (30) shows that the numerator is smaller than the denominator, which implies that $\alpha_3 \in (0, 1]$ as both the numerator and denominator are at least equal to 1. Moreover, if at least one weight is trimmed the last inequality in Equation (30) becomes a strict inequality so that $\alpha_3 \in (0, 1)$.

We now show in a similar way that $\alpha_3^* \in (0, 1]$ and $\alpha_3^* \in (0, 1)$ if at least one weight is trimmed. The scaling factor α_3^* has to make sure that all untrimmed weights (that have a sum of $\sum_{j=1}^N I_{\{\hat{w}_{j,t^*} \leq c\}} \hat{w}_{j,t^*}$) will sum together with the trimmed weights to -1. This means that the untrimmed weights should sum to

$$-1 - \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \cdot \left(\frac{c}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \cdot \hat{w}_{j,t^*} \right) = -1 - c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \cdot \frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}}$$

after being multiplied by α_3^*. Therefore, the scaling factor α_3^* is given by

$$\alpha_3^* = \frac{-1 - c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \cdot \frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}}}{\sum_{j=1}^N I_{\{\hat{w}_{j,t^*} \leq c\}} \hat{w}_{j,t^*}}.$$

As $c \geq 0$ and $I_{\{\hat{w}_{j,t^*} < -c\}} \cdot \frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \in [0, 1]$, the numerator of α_3 is at most equal to -1. The denominator is also at most equal to -1, as the originally estimated weights sum to -1 and we now only omit positive weights (when present). These two observations combined imply that $\alpha_3^* > 0$. To prove that $\alpha_3^* \leq 1$ we rewrite the numerator as

$$\begin{aligned}
-1 - c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \cdot \frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} &= \sum_{j=1}^N \hat{w}_{j,t^*} - c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \cdot \frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \\
&= \sum_{j=1}^N \hat{w}_{j,t^*} \left(I_{\{\hat{w}_{j,t^*} \leq c\}} + I_{\{\hat{w}_{j,t^*} > c\}} \right) - \\
&\qquad c \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \cdot \frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \\
&= \sum_{j=1}^N \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \leq c\}} + \\
&\qquad \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \left(\hat{w}_{j,t^*} - c \cdot \frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \right) \\
&\geq \sum_{j=1}^N \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \leq c\}} + \sum_{j=1}^N I_{\{\hat{w}_{j,t^*} > c\}} \left(\hat{w}_{j,t^*} - c \right) \\
&\geq \sum_{j=1}^N \hat{w}_{j,t^*} I_{\{\hat{w}_{j,t^*} \leq c\}},
\end{aligned} \tag{31}$$

where the first inequality follows because $c \geq 0$ and $\frac{\hat{w}_{j,t^*}}{\max\limits_{1 \leq j \leq N} \hat{w}_{j,t^*}} \in (0, 1]$ for $\hat{w}_{j,t^*} > c$ and where the second inequality follows as $I_{\{\hat{w}_{j,t^*} > c\}} \left(\hat{w}_{j,t^*} - c \right)$ is either 0 if $\hat{w}_{j,t^*} \leq c$ or strictly positive if

$\hat{w}_{j,t^*} > c$. Equation (31) shows that the numerator is larger than the denominator, which implies that $\alpha_3^* \in (0, 1]$ as both the numerator and denominator are at most equal to -1. Moreover, if at least one weight is trimmed the last inequality in Equation (31) becomes a strict inequality so that $\alpha_3^* \in (0, 1)$.

G Derivation of minimum and maximum portfolio weights for TR5

The most negative weight allowed by the gross-exposure constraint of trimming method TR5 can be achieved by letting all weights except for one (say w_{j^*}) be positive. Setting $w_{j^*} = -c/2$ then gives

$$\|\boldsymbol{w}\|_1 = \sum_{j=1}^{N} |w_j| = |w_{j^*}| + \sum_{j \neq j^*} |w_j| = -w_{j^*} + \sum_{j \neq j^*} w_j = -(-c/2) + (1 - (-c/2)) = 1 + c,$$

where we used in the fourth equality that the weights sum to one. Setting $w_{j^*} < -c/2$ would result in $\|\boldsymbol{w}\|_1 > 1 + c$, confirming that $w_{j^*} = -c/2$ is the smallest possible weight that can occur.

We have in a similar way that the most positive weight allowed by the gross-exposure constraint of trimming method TR5 can be achieved by letting all weights except for one (say w_{j^*}) be negative. Setting $w_{j^*} = 1 + c/2$ then gives

$$\|\boldsymbol{w}\|_1 = \sum_{j=1}^{N} |w_j| = |w_{j^*}| + \sum_{j \neq j^*} |w_j| = w_{j^*} + \sum_{j \neq j^*} (-w_j) = 1 + c/2 + c/2 = 1 + c,$$

where we again used in the fourth equality that the weights sum to one. Setting $w_{j^*} > 1 + c/2$ would result in $\|\boldsymbol{w}\|_1 > 1 + c$, confirming that $w_{j^*} = 1 + c/2$ is the largest possible weight that can occur.

H Remaining results Vuong tests

Table 17: Average results of Vuong test for experimental data set 1 with AIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	5.12	0.11	7.85	-2.73	2.09	-2.67	4.42
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-4.99	3.03	-6.73	-3.49	-6.70	0.75
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			7.81	-2.75	2.00	-2.77	4.36
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-9.08	-4.50	-9.08	-2.69
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					5.27	-0.02	7.86
					(0.00)	(0.56)	(0.00)
model 6						-5.21	2.77
						(0.00)	(0.00)
model 7							7.95
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the first experimental data set) with an AIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 18: Average results of Vuong test for experimental data set 1 with SIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	4.52	0.15	7.33	-4.28	0.21	-4.16	2.74
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
model 2		-4.36	3.03	-7.61	-5.91	-7.56	-0.46
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			7.27	-4.28	0.11	-4.30	2.67
			(0.00)	(0.00)	(0.13)	(0.00)	(0.00)
model 4				-9.90	-5.74	-9.87	-4.44
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					4.62	0.01	7.28
					(0.00)	(0.91)	(0.00)
model 6						-4.54	2.77
						(0.00)	(0.00)
model 7							7.34
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the first experimental data set) with an SIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 19: Average results of Vuong test for experimental data set 2 without parameter correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	1.90	0.00	1.97	0.59	1.99	0.59	1.92
	(0.00)	(1.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-1.90	-0.03	-0.91	0.70	-0.91	0.40
		(0.00)	(0.63)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			1.97	0.59	1.99	0.59	1.92
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-0.89	-0.53	-0.89	-0.56
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					1.89	0.00	1.91
					(0.00)	(1.00)	(0.00)
model 6						-1.89	-0.09
						(0.00)	(0.18)
model 7							1.91
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the second experimental data set) without a correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 20: Average results of Vuong test for experimental data set 2 with AIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	1.62	0.00	1.73	-1.24	0.52	-1.24	0.33
	(0.00)	(1.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-1.62	-0.03	-1.98	-1.48	-1.98	-0.83
		(0.00)	(0.63)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			1.73	-1.24	0.53	-1.24	0.33
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-2.07	-0.66	-2.07	-1.28
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					1.60	0.00	1.67
					(0.00)	(1.00)	(0.00)
model 6						-1.60	-0.09
						(0.00)	(0.18)
model 7							1.67
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the second experimental data set) with an AIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 21: Average results of Vuong test for experimental data set 2 with SIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	1.24	0.00	1.39	-3.79	-1.53	-3.79	-1.89
	(0.00)	(1.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-1.24	-0.03	-3.48	-4.51	-3.48	-2.53
		(0.00)	(0.63)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			1.39	-3.79	-1.53	-3.79	-1.89
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-3.71	-2.31	-3.71	-3.84
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					1.19	0.00	1.32
					(0.00)	(1.00)	(0.00)
model 6						-1.19	-0.09
						(0.00)	(0.18)
model 7							1.32
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the second experimental data set) with an SIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 22: Average results of Vuong test for experimental data set 3 without parameter correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	9.09	0.04	11.14	-1.33	6.92	-1.21	8.27
	(0.00)	(0.54)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-8.71	4.43	-9.16	-1.93	-8.75	2.23
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			11.65	-1.31	6.74	-1.28	8.57
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-11.58	-4.95	-11.67	-1.93
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					9.16	0.04	10.63
					(0.00)	(0.50)	(0.00)
model 6						-8.72	3.47
						(0.00)	(0.00)
model 7							10.97
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the third experimental data set) without a correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 23: Average results of Vuong test for experimental data set 3 with AIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	8.67	0.08	10.81	-2.62	5.30	-2.32	7.02
	(0.00)	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-8.26	4.43	-10.05	-4.30	-9.51	1.22
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			11.27	-2.58	5.11	-2.48	7.23
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-12.32	-6.02	-12.33	-3.21
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					8.77	0.14	10.31
					(0.00)	(0.00)	(0.00)
model 6						-8.23	3.47
						(0.00)	(0.00)
model 7							10.54
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the third experimental data set) with an AIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 24: Average results of Vuong test for experimental data set 3 with SIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	8.08	0.14	10.36	-4.41	3.04	-3.87	5.27
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-7.63	4.43	-11.28	-7.60	-10.55	-0.18
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)
model 3			10.75	-4.36	2.84	-4.15	5.36
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-13.35	-7.50	-13.25	-5.00
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					8.21	0.29	9.87
					(0.00)	(0.00)	(0.00)
model 6						-7.54	3.47
						(0.00)	(0.00)
model 7							9.95
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the third experimental data set) with an SIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 25: Average results of Vuong test for experimental data set 4 without parameter correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	9.24	-0.54	11.85	-1.51	6.66	-2.05	8.92
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-9.47	5.05	-9.61	-2.88	-10.08	2.71
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			13.01	-0.94	6.99	-1.64	9.62
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-12.54	-6.05	-13.53	-1.92
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					8.88	-0.74	11.29
					(0.00)	(0.00)	(0.00)
model 6						-9.42	4.43
						(0.00)	(0.00)
model 7							12.63
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the fourth experimental data set) without a correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 26: Average results of Vuong test for experimental data set 4 with AIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	8.81	-0.40	11.50	-2.77	5.02	-3.14	7.64
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-8.91	5.05	-10.44	-5.33	-10.79	1.67
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			12.52	-2.25	5.26	-2.88	8.19
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-13.25	-7.18	-14.17	-3.25
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					8.44	-0.63	10.93
					(0.00)	(0.00)	(0.00)
model 6						-8.86	4.43
						(0.00)	(0.00)
model 7							12.15
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the fourth experimental data set) with an AIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 27: Average results of Vuong test for experimental data set 4 with SIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	8.20	-0.22	11.02	-4.54	2.74	-4.64	5.85
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-8.11	5.05	-11.61	-8.74	-11.79	0.23
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
model 3			11.84	-4.07	2.85	-4.61	6.20
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-14.25	-8.76	-15.06	-5.10
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					7.82	-0.48	10.42
					(0.00)	(0.00)	(0.00)
model 6						-8.09	4.43
						(0.00)	(0.00)
model 7							11.48
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the fourth experimental data set) with an SIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 28: Average results of Vuong test for experimental data set 5 without parameter correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	9.73	-0.57	12.12	-1.38	7.25	-1.95	9.38
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-9.97	4.92	-9.97	-2.73	-10.46	2.76
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			13.21	-0.79	7.59	-1.50	10.11
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-12.73	-5.88	-13.76	-1.75
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					9.34	-0.74	11.62
					(0.00)	(0.00)	(0.00)
model 6						-9.88	4.41
						(0.00)	(0.00)
model 7							13.00
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the fifth experimental data set) without a correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 29: Average results of Vuong test for experimental data set 5 with AIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	9.32	-0.42	11.79	-2.66	5.63	-3.04	8.09
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-9.41	4.92	-10.84	-5.25	-11.21	1.69
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 3			12.74	-2.12	5.86	-2.77	8.67
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-13.47	-7.05	-14.44	-3.12
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					8.92	-0.63	11.28
					(0.00)	(0.00)	(0.00)
model 6						-9.36	4.41
						(0.00)	(0.00)
model 7							12.54
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the fifth experimental data set) with an AIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

Table 30: Average results of Vuong test for experimental data set 5 with SIC correction

	model 2	model 3	model 4	model 5	model 6	model 7	model 8
model 1	8.75	-0.21	11.33	-4.44	3.36	-4.55	6.30
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 2		-8.65	4.92	-12.06	-8.76	-12.25	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
model 3			12.07	-3.97	3.45	-4.54	6.66
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
model 4				-14.51	-8.68	-15.38	-5.03
				(0.00)	(0.00)	(0.00)	(0.00)
model 5					8.34	-0.48	10.80
					(0.00)	(0.00)	(0.00)
model 6						-8.62	4.41
						(0.00)	(0.00)
model 7							11.89
							(0.00)

Note: This table reports the average Vuong test statstics (over all estimation windows in the fifth experimental data set) with an SIC correction for the number of parameters. Positive values indicate that the model in the corresponding row is preferred over the model in the corresponding column. The different dynamic models are described in Table 1. In parentheses are the p-values of the average Vuong test statistics.

I Illustration of differences in trimming methods

We illustrate the differences between the trimming methods that directly estimate trimmed mean-variance weights (TR4** and TR5**) and the trimming methods that trim the estimated mean-variance weights (TR1, TR2 and TR3) with an example. As in the example in Appendix C, we assume that we have $N = 3$ risky assets at time t, with expected excess return vector $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$ given by

$$\boldsymbol{\mu}_t = \begin{pmatrix} -0.01 \\ 0.01 \\ 0.01 \end{pmatrix}, \quad \boldsymbol{\Sigma}_t = \begin{pmatrix} 0.010 & 0.009 & 0 \\ 0.009 & 0.010 & 0 \\ 0 & 0 & 0.010 \end{pmatrix}.$$

We show in Appendix C that the untrimmed relative portfolio weights for these inputs are given by

$$\boldsymbol{w}_t = \begin{pmatrix} -10 \\ 10 \\ 1 \end{pmatrix}.$$

We now derive the relative weights with our trimming rules. We set the trimming threshold c equal to 0, so that TR4** and TR5** respectively TR1, TR2 and TR3 are equivalent. The methods that trim the estimated weights set $w_{1,t} = 0$ so that the resulting scaling factor is equal to $\frac{1}{11}$. The trimmed weight vectors are then given by

$$\boldsymbol{w}_t^{\text{TR1}} = \boldsymbol{w}_t^{\text{TR2}} = \boldsymbol{w}_t^{\text{TR3}} = \begin{pmatrix} 0 \\ \frac{10}{11} \\ \frac{1}{11} \end{pmatrix}.$$

The methods that estimate the trimmed weights require all weights to be non-negative. We first derive that this implies that it is optimal to set $w_{1,t} = 0$. Letting $\sigma_{i,j,t}$ denote the (i, j)-th element of $\boldsymbol{\Sigma}_t$, the portfolio variance is given by

$$\sigma_p^2 = w_{1,t}^2 \sigma_{1,1,t} + w_{2,t}^2 \sigma_{2,2,t} + w_{3,t}^2 \sigma_{3,3,t} + 2w_{1,t} w_{2,t} \sigma_{1,2,t} + 2w_{1,t} w_{3,t} \sigma_{1,3,t} + 2w_{2,t} w_{3,t} \sigma_{2,3,t}.$$

Plugging in the values of $\sigma_{i,j,t}$ gives a portfolio variance of

$$\sigma_p^2 = 0.010 w_{1,t}^2 + 0.010 w_{2,t}^2 + 0.010 w_{3,t}^2 + 0.018 w_{1,t} w_{2,t}.$$

Hence, the Sharpe ratio of the portfolio is given by

$$\text{SR}_p = \frac{\mu_p}{\sqrt{\sigma_p^2}} = \frac{w_{1,t}\mu_{1,t} + w_{2,t}\mu_{2,t} + w_{3,t}\mu_{3,t}}{\sqrt{\sigma_p^2}} = \frac{-0.01 w_{1,t} + 0.01 w_{2,t} + 0.01 w_{3,t}}{\sqrt{0.010 w_{1,t}^2 + 0.010 w_{2,t}^2 + 0.010 w_{3,t}^2 + 0.018 w_{1,t} w_{2,t}}}.$$

Using that $w_{1,t} + w_{2,t} + w_{3,t} = 1$ we can rewrite this Sharpe ratio as

$$\mathrm{SR}_p = \frac{0.01 - 0.02w_{1,t}}{\sqrt{0.010w_{1,t}^2 + 0.010w_{2,t}^2 + 0.010w_{3,t}^2 + 0.018w_{1,t}w_{2,t}}},$$

which shows that the Sharpe ratio is only positive for $w_{1,t} < \frac{1}{2}$, so that the optimal trimmed weights must satisfy that $0 \leq w_{1,t} < \frac{1}{2}$. Furthermore, the derivative of the Sharpe ratio with respect to $w_{1,t}$ is given by

$$\begin{aligned}
\frac{\partial \mathrm{SR}_p}{\partial w_{1,t}} &= \frac{\sqrt{0.010w_{1,t}^2 + 0.010w_{2,t}^2 + 0.010w_{3,t}^2 + 0.018w_{1,t}w_{2,t}} \cdot (-0.02)}{0.010w_{1,t}^2 + 0.010w_{2,t}^2 + 0.010w_{3,t}^2 + 0.018w_{1,t}w_{2,t}} \\
&\quad - \frac{(0.01 - 0.02w_{1,t}) \cdot \frac{0.020w_{1,t}+0.018w_{2,t}}{2\sqrt{0.010w_{1,t}^2+0.010w_{2,t}^2+0.010w_{3,t}^2+0.018w_{1,t}w_{2,t}}}}{0.010w_{1,t}^2 + 0.010w_{2,t}^2 + 0.010w_{3,t}^2 + 0.018w_{1,t}w_{2,t}} \\
&= \frac{-0.02\left(0.010w_{1,t}^2 + 0.010w_{2,t}^2 + 0.010w_{3,t}^2 + 0.018w_{1,t}w_{2,t}\right)}{\left(0.010w_{1,t}^2 + 0.010w_{2,t}^2 + 0.010w_{3,t}^2 + 0.018w_{1,t}w_{2,t}\right)^{\frac{3}{2}}} \\
&\quad - \frac{(0.01 - 0.02w_{1,t}) \cdot (0.010w_{1,t} + 0.009w_{2,t})}{\left(0.010w_{1,t}^2 + 0.010w_{2,t}^2 + 0.010w_{3,t}^2 + 0.018w_{1,t}w_{2,t}\right)^{\frac{3}{2}}}
\end{aligned}$$

showing that for $0 \leq w_{1,t} < \frac{1}{2}$ the derivative of the Sharpe ratio with respect to $w_{1,t}$ is always negative. This implies that the Sharpe ratio is a decreasing function of $w_{1,t}$ for all values of $0 \leq w_{1,t} < \frac{1}{2}$, so that it is optimal to set $w_{1,t} = 0$. This implies that the Sharpe ratio is given by

$$\mathrm{SR}_p = \frac{0.01w_{2,t} + 0.01w_{3,t}}{\sqrt{0.01w_{2,t}^2 + 0.01w_{3,t}^2}} = \frac{0.01w_{2,t} + 0.01(1 - w_{2,t})}{\sqrt{0.01w_{2,t}^2 + 0.01(1 - w_{2,t})^2}} = \frac{0.01}{\sqrt{0.01(2w_{2,t}^2 + 1 - 2w_{2,t})}}, \tag{32}$$

which is maximised when $(2w_{2,t}^2 + 1 - 2w_{2,t})$ is minimised. As $(2w_{2,t}^2 + 1 - 2w_{2,t})$ is a convex function, it is minimised when the first derivative is equal to zero, giving $w_{2,t} = \frac{1}{2}$. We thus get

$$\boldsymbol{w}_t^{\mathrm{TR4}^{**}} = \boldsymbol{w}_t^{\mathrm{TR5}^{**}} = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

The resulting Sharpe ratios for the different trimming methods can be obtained by plugging the obtained value of $w_{2,t}$ into Equation (32), because for both methods $w_{1,t} = 0$. This yields that the trimming methods that trim the estimated weights have a Sharpe ratio of 0.1095 whereas the trimming methods that directly estimate the trimmed weights have a Sharpe ratio of 0.1414. This difference arises because the methods that trim the estimated weights still give a relatively large weight to asset 2, whereas this is only beneficial if we can take a negative position in asset 1. The methods that directly estimate the trimmed weights are able to incorporate this information, resulting in a better performance.

J Results of trimming methods for other trimming thresholds

Table 31: Sharpe ratios that do not incorporate transaction costs for the different trimming methods using a fixed trimming threshold of $c = 0$

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2480	0.2463	0.2157	-0.0577	-0.0271
TR2	0.2480	0.2463	0.2157	-0.0577	-0.0271
TR3	0.2480	0.2463	0.2157	-0.0577	-0.0271
TR4	0.2678	0.1353	0.2347	0.1362	0.2895
TR4**	0.2462	0.2456	0.2058	0.2749	0.3693
TR5	0.2677	0.1353	0.2355	0.1363	0.2897
TR5**	0.2462	0.2458	0.2058	0.2750	0.3692

Note: This table reports the monthly Sharpe ratios that do not incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5.

Table 32: Sharpe ratios that incorporate transaction costs for the different trimming methods using a fixed trimming threshold of $c = 0$

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2434	0.2295	0.1998	-0.0812	-0.0509
TR2	0.2434	0.2295	0.1998	-0.0812	-0.0509
TR3	0.2434	0.2295	0.1998	-0.0812	-0.0509
TR4	0.2621	0.1305	0.2286	0.1312	0.2808
TR4**	0.2247	0.2309	0.1902	0.2474	0.3369
TR5	0.2620	0.1305	0.2278	0.1312	0.2809
TR5**	0.2256	0.2310	0.1902	0.2475	0.3369

Note: This table reports the monthly Sharpe ratios that incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5.

Table 33: Sharpe ratios that do not incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 1$ that does not incorporate transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2485	0.2482	0.2138	0.0537	0.0869
TR2	0.2482	0.2490	0.1484	0.0523	0.0823
TR3	0.2560	0.2410	0.1858	0.0703	0.1068
TR4	0.2400	0.1294	0.1610	0.1318	0.2475
TR4**	0.2097	0.2457	0.1728	0.1857	0.2601
TR5	0.2754	0.1302	0.2248	0.1698	0.2605
TR5**	0.2581	0.2452	0.2274	0.2339	0.3414

Note: This table reports the monthly Sharpe ratios that do not incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ using the data-driven approach that maximises the Sharpe ratio that does not incorporate transaction costs.

Table 34: Sharpe ratios that do not incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 1$ that incorporates transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2482	0.2435	0.2152	-0.0308	0.0509
TR2	0.2542	0.2431	0.1545	0.0229	0.0625
TR3	0.2550	0.2438	0.1958	0.0222	0.0930
TR4	0.2354	0.1312	0.2150	0.1132	0.2709
TR4**	0.2289	0.2571	0.1906	0.2399	0.3251
TR5	0.2712	0.1305	0.2253	0.1549	0.2825
TR5**	0.2604	0.2518	0.2392	0.2480	0.3675

Note: This table reports the monthly Sharpe ratios that do not incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ using the data-driven approach that maximises the Sharpe ratio that incorporates transaction costs.

Table 35: Sharpe ratios that do not incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 2$ that does not incorporate transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2495	0.2373	0.2167	0.0845	0.1072
TR2	0.2560	0.2347	0.0585	0.0622	0.0863
TR3	0.2619	0.2298	0.1587	0.0725	0.1035
TR4	0.2435	0.1299	0.1568	0.1407	0.2511
TR4**	0.2122	0.2393	0.1411	0.1774	0.2351
TR5	0.2615	0.1301	0.2070	0.1770	0.2391
TR5**	0.2439	0.2430	0.2220	0.1981	0.3091

Note: This table reports the monthly Sharpe ratios that do not incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ using the data-driven approach that maximises the Sharpe ratio that does not incorporate transaction costs.

Table 36: Sharpe ratios that do not incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 2$ that incorporates transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2484	0.2418	0.2147	0.0056	0.0676
TR2	0.2624	0.2324	0.1396	0.0229	0.0492
TR3	0.2611	0.2317	0.1635	0.0262	0.1045
TR4	0.2382	0.1308	0.2027	0.1152	0.2771
TR4**	0.1854	0.2432	0.1988	0.2299	0.3275
TR5	0.2665	0.1312	0.2185	0.1495	0.2906
TR5**	0.2447	0.2540	0.2297	0.2392	0.3607

Note: This table reports the monthly Sharpe ratios that do not incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ using the data-driven approach that maximises the Sharpe ratio that incorporates transaction costs.

Table 37: Sharpe ratios that incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 2$ that does not incorporate transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2424	0.2006	0.0691	-0.1353	-0.1136
TR2	0.1973	0.2057	-0.1443	-0.1457	-0.1223
TR3	0.2164	0.2006	-0.0002	-0.0844	-0.0873
TR4	0.1781	0.1219	0.0350	0.0919	0.1741
TR4**	0.0591	0.2095	-0.0252	0.0064	-0.012
TR5	0.2237	0.1218	0.1535	0.1018	0.1304
TR5**	0.1907	0.2155	0.1701	0.0877	0.1935

Note: This table reports the monthly Sharpe ratios that incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ using the data-driven approach that maximises the Sharpe ratio that does not incorporate transaction costs.

Table 38: Sharpe ratios that incorporate transaction costs for the different trimming methods using the data-driven approach with $c_{\max} = 2$ that incorporates transaction costs

trimming method	data set 1	data set 2	data set 3	data set 4	data set 5
TR1	0.2416	0.2086	0.1484	-0.1456	-0.0698
TR2	0.2136	0.2023	0.0062	-0.1238	-0.1008
TR3	0.2153	0.1998	0.0564	-0.1171	-0.0567
TR4	0.1832	0.1234	0.1012	0.0846	0.2392
TR4**	0.0621	0.2139	0.1014	0.1492	0.2535
TR5	0.2325	0.1240	0.1763	0.0813	0.2496
TR5**	0.1960	0.2317	0.1869	0.1739	0.2975

Note: This table reports the monthly Sharpe ratios that incorporate transaction costs for portfolios constructed using the trimming methods described in Section 5. The trimming threshold c is selected from the set $\{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ using the data-driven approach that maximises the Sharpe ratio that incorporates transaction costs.

K Results shrinkage without incorporating costs for thresholds

Table 39: Differences in Sharpe ratios that do not incorporate transaction costs resulting from using simulation-based shrinkage that does not incorporate transaction costs

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
10 minus 5	-0.0779	0.2175	-0.0033	0.1116	-0.0417	0.0850	-0.1004
	(0.34)	(0.00)	(0.97)	(0.12)	(0.35)	(0.36)	(0.09)
11 minus 8	-0.0057	-0.0265	0.0216	0.0277	-0.0515	-0.0313	-0.0396
	(0.94)	(0.43)	(0.54)	(0.44)	(0.10)	(0.43)	(0.29)
12 minus 9	-0.0441	-0.0446	0.0163	0.0297	-0.0523	-0.0761	-0.0462
	(0.45)	(0.14)	(0.68)	(0.41)	(0.07)	(0.12)	(0.20)

Note: This table reports the differences in monthly Sharpe ratios that do not incorporate transaction costs that are a result of applying the simulation-based shrinkage rule with shrinkage thresholds selected using the data-driven approach that does not incorporate transaction costs. The different combinations are described in Table 2. In parentheses are the p-values of the reported differences in Sharpe ratio.

L Sharpe ratio differences from taking average weights

Table 40: Differences in Sharpe ratios that do not incorporate transaction costs resulting from taking the average weights over simulated samples

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
5 minus 4	-0.0360	0.0029	-0.0064	-0.1000	0.2085	0.1075	-0.0136
	(0.49)	(0.95)	(0.81)	(0.13)	(0.00)	(0.01)	(0.68)
8 minus 6	0.0627	0.0445	0.0165	-0.0205	0.0044	-0.0179	-0.0084
	(0.30)	(0.09)	(0.23)	(0.14)	(0.70)	(0.30)	(0.17)
9 minus 7	-0.0102	0.0383	0.0237	-0.0250	0.0059	0.0058	-0.0078
	(0.79)	(0.15)	(0.06)	(0.14)	(0.57)	(0.77)	(0.35)

Note: This table reports the differences in monthly Sharpe ratios that do not incorporate transaction costs that are a result of simulating multiple return samples and taking average weights. The different combinations are described in Table 2. In parentheses are the p-values of the reported differences in Sharpe ratio.

Table 41: Differences in Sharpe ratios that incorporate transaction costs resulting from taking the average weights over simulated samples

combinations	data set 1	data set 2	data set 3	data set 4	data set 5	data set 6	data set 7
5 minus 4	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
8 minus 6	0.0685	0.0466	0.0153	-0.0212	0.0089	-0.0165	-0.0094
	(0.25)	(0.08)	(0.20)	(0.13)	(0.49)	(0.28)	(0.17)
9 minus 7	0.0189	0.0496	0.0311	-0.0263	0.0146	-0.0056	-0.0087
	(0.64)	(0.06)	(0.03)	(0.14)	(0.17)	(0.78)	(0.43)

Note: This table reports the differences in monthly Sharpe ratios that incorporate transaction costs that are a result of simulating multiple return samples and taking average weights. The different combinations are described in Table 2. No values are reported if returns net of transaction costs could not be computed correctly due to extreme portfolio weights. In parentheses are the p-values of the reported differences in Sharpe ratio.