ERASMUS UNIVERSITEIT ROTTERDAM

Master Thesis Quantitative Finance

Optimizing non-zero weights in classic long-short portfolios

Author
Zef Orbons (510772)

Supervisor
Dr. S.O.R. Lonn (EUR)

Abstract

The search for signals that possess high predictive power for expected returns keeps many researchers occupied. The traditional procedure of constructing portfolios is to sort assets according to their factor scores into pre-chosen quantiles and form long-short portfolios correspondingly. Using this kind of methodology ignores any information of the covariance matrix of stock returns, which was already proven to be of interest for optimizing weights in this kind of setting back in 1952. This research paper demonstrates that using the DCC-NL estimator, replacing the unobservable true covariance matrix in order to optimize the non-zero weights within quintiles, substantially enhances the strength of cross-sectional factor tests.

February 26, 2021

Contents

1 Introduction

The search for factors that predict the variation of the cross-section of expected returns within markets generates a rich literature. Whereas the Capital Asset Pricing Model (CAPM) with its strong assumptions remained fundamental, this literature shows that additional factors are now required to explain why the cross-section of expected returns vary within markets. In addition to the extensive literature of the many available anomalies, the complementary question emerges of how to actually construct portfolios to trade on any given factor. With factors, a function of historical data that is able to explain the cross-section of subsequent stock returns is explicitly described. Currently, the factor zoo contains numerous factors, where the factors in over 300 papers discussed by Harvey, Liu, and Zhu (2016) only appear to be a small fraction within the strand of literature.

The standard methodology sorts stocks according to their factor scores into quantiles, upon which a corresponding dollar-neutral long-short portfolio is formed. The portfolio goes long in the assets that are ranked into the top quantile and shorts the assets in the bottom quantile. However, Ledoit, Wolf, and Zhao (2019) argue that this *status quo*, where the simplistic sorting is used, poses a conundrum, since Markowitz (1952) showed that the optimal weights in a mean-variance (MV) preference profile is depending on the covariance matrix, whereas the aforementioned investment strategy does not take the covariance matrix into account. In addition, Daniel, Mota, Rottke, and Santos (2020) have shown that the standard sorting strategy does not guarantee the set of portfolios to span the MV efficient frontier, meaning that the weights allocated to the stocks are not optimized in order to generate the highest average returns given a level of variance. This seems to be due to the fact that sorting on a characteristic results in portfolios that are likely to load on an unpriced source of common risk. Thus, there needs to be more to it than a simple sorting. Therefore, there is still space to investigate the weights of the stocks within these top and bottom quantiles.

In this respect, the optimization problem is investigated, whereby the factor(s) and the quantile are given. More specifically, this research paper investigates how to optimize the non-zero weights of the long-short (multi)factor portfolio, within the given quantile and specific factor(s) for high dimensional data sets. This research paper deviates from Ledoit et al. (2019) by also taking into consideration bivariate sorts instead of only looking at univariate sorts. This results in mixed conclusions. The conditional sorts on beta and size provide better returns then the single sorts. Value and momentum constitute the opposite effect. Furthermore, in order to perform the sort within the extreme quantiles, the methodology of the sequential sort is followed as is proposed by Lambert and Hübner (2013). This way of conditional sorting is preferred from a portfolio management perspective, as it is closer to the manner investors approach an asset allocation problem. In general, investors deal with one problem at a time and thus sequentially tackle the different pricing anomalies. Additionally, the signal of the specific factor(s) is implemented within the model by estimating the covariance matrix on a reduced dimensionality based upon the quantile sorts, meaning that the estimator only has to deal with the stocks that are within the specific quantiles, instead of the whole universe of assets. This contrasts with Ledoit et al. (2019), since they implement the factor signal by targeting the target exposure of the signal and in addition, run the covariance matrix estimator over the whole universe of stocks. The returns of these two strategies do not differ much as the applied strategy returned higher t-statistics then the methodology used by Ledoit et al. (2019) in half the cases, regarded in a subsample. Furthermore, empirics have shown that the global minimum variance (GMV) portfolio generates high returns, whereas the MV portfolio is deemed to produce poor results. Hence, the considered problem is optimized with respect to the minimum variance profile, where Ledoit et al. (2019) show that it is possible for their strategy to make a direct connection with the MV optimization described by Markowitz (1952). Next to that, the analysis conducted shows that the two profiles result in very similar observations and therefore, the statement of poor out-of-sample returns for the MV profile does not hold in this case. Overall, the size factor seems to be the best proxy for risk, since sorting on this signal yields the highest returns.

The research paper is organized as follows. The next section provides a literature overview with relevant articles, establishing the foundation and complementing this paper. In Section 3, the methodology used in this paper is lined out. Then, Section 4 comprises a description of the data that is used in order to conduct an empirical analysis, followed by the actual empirical analysis in Section 5. Finally, Section 6 concludes the paper.

2 Literature review

Going back to at least Fama and French (1993), the favored method for establishing the validity of factors has been to construct portfolios based on sorting, where one goes long the stocks that are in the top quintile and short the stocks in the bottom quintile according to their factor scores. Considering quintiles seems to be a common choice regarding the literature. However, instead of quintiles, some may favor among others deciles or terciles. Furthermore, within the sorting of the portfolios, Chen, Copeland, and Mayers (1987) state that multifactor models yield superior returns in well-diversified portfolios in comparison to single factor sorts. The high dimensionality of the data set used in this research paper enables the construction of well-diversified portfolios and therefore, the focus is laid upon sequential sorting. With respect to the factors, the four factor model discussed by Carhart (1997) is used, comprising beta, size, value and momentum. This model is in turn based upon the three factor model introduced by Fama and French (1993), whereby momentum was not taken into account yet. The focus in this research paper is especially laid upon the size factor, since Lambert and Hübner (2014) claim that this factor is underestimated. The size pricing anomaly was introduced by Banz (1981), stating that firms with low market capitalizations outperform firms with large capitalizations. Although it is generally recognized that anomalies proxy for risk Lo and MacKinlay (1990) and Black (1992) object empirical procedures that use the size characteristic as a proxy for risk, since they claim that an empirical anomaly is, by definition, an empirical fact that cannot be supported by the prevailing theory. Nevertheless, Berk (1995) criticizes the fact that size-related regularities in asset prices are called an anomaly in practice, because an anomaly should have a relationship that is unobservable. He provides a theoretical answer to the question why relative size measures risk by showing that there actually exists an inverse relationship between size-related regularities and returns. In addition, Berk (1995) shows that the logarithm of a firm's market value always measures the firm's discount rate. According to Berk (1995) the same logic can be applied to explain the predictive power of other factors such as earnings-to-price ratio, dividend yield, book-to-market equity or simply market capitalization itself.

Once the portfolios are sorted based on one or several signals, the topic of investigating which weights are optimal within the given quantiles becomes of interest. However, in order to optimize the weights within the quantiles, estimation risk is one of the challenges that belongs to empirical portfolio modelling. In the standard MV preference framework described by Markowitz (1952) in his Portfolio Selection paper, he proves that the optimal weights of an N-dimensional asset space do not only depend on the first moment of the underlying asset returns, but also on the covariance matrix. More specifically, the inverse of the covariance matrix. However, the true mean and covariance matrix are unobservable and therefore need to be estimated. In order to do so, the commonly used sample estimations of the N means and $N\frac{(N+1)}{2}$ covariance parameters can be used. Nevertheless, these estimates generate noisy portfolio weight estimates with large standard errors. Jobson and Korkie (1980) show that the out-of-sample results of portfolios with these estimates are very disappointing. In addition, Kan and Zhou (2007) also show among others that the standard plug-in approach that replaces the population parameters by their sample estimates generate very poor out-of-sample performances. Next to that, Michaud (1989) sums up pros and especially cons of the use of optimizers. He states that the unintuitive character of the optimized portfolios can be traced back to the fact that in a fundamental sense, optimizers are maximizing estimation errors associated with the input estimates. Without careful problem definition derived from sound investment judgment and sophisticated adjustment of the inputs, optimization may often do more harm than good according to Michaud (1989). Moreover, he concludes that these portfolios result in extreme short positions with little diversifications. Furthermore, Michaud (1989) describes the fundamental problem of parameter uncertainty as that the level of mathematical sophistication of the optimization algorithm is far greater than the level of information in the input estimations. Additionally, Nakagawa, Imamura, and Yoshida (2018) mention that portfolio managers increasingly focus their attention on risk-based portfolios only, such as done by the GMV profile. In this respect, the notoriously hard estimation of the expected returns is not longer required and the focus is solely laid upon the covariance matrix. For that reason this research paper takes the GMV profile into account, instead of the MV profile like in Ledoit et al. (2019). Next to that, this choice is based upon the fact that Haugen and Baker (1991), Jagannathan and Ma (2003) and Nielsen and Aylursubramanian (2008) have found that the GMV portfolios produce desirable out-of-sample properties, not only in terms of the minimized risk, however in terms of reward-to-risk, like the information ratio, as well.

Nonetheless, the past has shown that it is difficult to estimate the covariance matrix because of the possible high-dimensionality, *i.e.* the curse of dimensionality. Additionally, the covariance matrix is deemed to be varying over time, which is not taken into consideration in the simple sample covariance matrix. Nowadays, more and more research is conducted to deal with these problems. For example, Bollerslev (1986) has shown how to make the model compliant with a time varying covariance matrix by implementing the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) model. Additionally, R. Engle (2002) covered for the time variation by proposing the Dynamic Conditional Correlation (DCC) model. Furthermore, R. F. Engle, Ledoit, and Wolf (2017) attempt an effort to deal with both aforementioned problems by robustifying the DCC model of R. Engle (2002) for large universes of stocks. Ledoit et al. (2019) summarise and use the Dynamic Conditional Correlation Non Linear (DCC-NL) estimator of R. F. Engle et al. (2017) to substantially enhance the power of tests for cross-sectional anomalies. Within this covariance estimator, the Non Linear (NL) part covers for the issue of the high dimensionality in stocks. The time series direction is another interesting topic of this paper. Within the DCC part of the DCC-NL estimator, Ledoit et al. (2019) cope with the variation over time of the covariance matrix, whereas the simple sample covariance matrix is static and thus not geared to cope with the variation over time.

After being sorted and optimized for the weights, the portfolio is held for a certain period of time. Then, at some point in time it is re-balanced according to freshly updated factor data. This procedure generates a time series of portfolio returns. The factor is deemed successful if the average portfolio return exceeds some benchmark at a suitable level of statistical significance. The t-statistic of the long-short portfolio return is used in this respect to measure the one-tailed hypothesis stating that the expected portfolio returns should be equal or larger than zero. Considering this central quantity to test the optimized returns, Harvey et al. (2016) found that historical cutoffs used to claim research findings are likely false. This is because new factors need to clear a much higher hurdle rate then the usual cutoff of two according to their hypothesis. Harvey et al. (2016) state that the cutoff should be adjusted to the appropriate factor to be sufficient.

Subsequently, in addition to the t-test, the naive $\frac{1}{N}$ portfolio discussed by DeMiguel, Garlappi, and Uppal (2009) will be used as a benchmark portfolio to compare the computed portfolio returns against. The reason for choosing this simple allocation of weights as the benchmark is that DeMiguel et al. (2009) have shown that this strategy is tough to compete with in out-of-sample portfolio returns. The second hypothesis test concerns the comparison of the Sharpe Ratio of the two investment strategies, depicting the optimized portfolio and the benchmark portfolio. The *status quo* in the literature appears to be to look at the test introduced by Jobson and Korkie (1981), which was corrected by Memmel (2003). However, these two tests are not valid in case the returns of interest depict heavier tails than the normal distribution or are of time series nature. Moreover, literature has shown that it is common for financial returns to be characterized with heavier tails than the normal distribution. Additionally, the serial correlation of the actual returns is of relevance, since a long-short portfolio is used. Hence, the need for a more robust test is required. Ledoit and Wolf (2008) introduce a statistical inference test capable of dealing with the two aforementioned problems.

3 Methodology

This section elaborates upon the methods used to obtain the optimized weights of the dollar-neutral factor portfolio. First of all, the assets are ranked according to their factor scores. Then, a dollar-neutral long-short portfolio is constructed. Hence, the assets are discriminated by assigning positive weights to the top quantile of the ranked assets, *i.e.* take a long position, and going short in the assets that are ranked in the bottom quantile and thus tend to underperform. Eventually, the weights will add up to zero in this dollar-neutral portfolio, deviating from the classic long portfolios where the weights sum up to one. Furthermore, as described in Section 2, the naive $\frac{1}{N}$ portfolio is used as a benchmark to compare the results with since this is deemed to be a tough competitor, shown by DeMiguel et al. (2009).

To start, the return, R, and the factor value, x, are observed over N assets and T time periods. Let N denote the investment universe, where the stocks within this universe are indexed by $i = \{1, ..., N\}$. Let T denote the time series on which trading takes place, where a specific trading day is indexed by $t = \{1, ..., T\}$. The cross-section of returns and factor scores are respectively denoted as follows.

$$\boldsymbol{R_t} \coloneqq (R_{1t}, ..., R_{Nt})' \qquad\qquad \boldsymbol{x_t} \coloneqq (x_{1t}, ..., x_{Nt})'$$

The portfolios are constructed once a month, in order to avoid turnover resulting in large transaction costs. Let us denote the portfolio construction days by z, where they range from $z = \{1, \ldots, Z\}$. For simplicity, one month is set equal to 21 consecutive trading days. Furthermore, it is a common convention for quantitative investors to construct portfolios based upon past returns. Hence, the first portfolio construction will be generated after K trading days. Therefore, $z = \{1, \ldots, Z\}$ is a subset of $t = \{1, \ldots, T\}$.

3.1 Factor construction

In order to perform sorts, factor scores are needed to eventually come up with dollar-neutral factor models. This section will describe what factors are used and how they are calculated to represent the factor scores. In this section, the index, $[(z * 21) + K - 21]$, is referred to by t^* that is denoted as the trading day. Furthermore, the computations of the factor scores of subsection 3.1.3 are based upon the computations used in Ledoit et al. (2019).

3.1.1 Beta

The CAPM was introduced by Treynor (1961, 1962), Sharpe (1964), Lintner (1965a, 1965b) and Mossin (1966) independently, building on the earlier work of Markowitz (1952) on diversification and modern portfolio theory. The CAPM is a one factor model, where the systematic risk, β, tries to explain the variation of the cross-section of expected returns within markets. The β calculation is useful for investors in order to understand whether a stock moves in the same direction as the rest of the market. Furthermore, it can provide some insights about the volatility, or riskiness of a specific stock in relation to the overall market. The systematic risk is calculated as follows.

$$\beta := \frac{cov(R_{it}, R_{mt})}{var(R_{mt})}$$

where R_{it} and R_{mt} depict the specific asset and market return at a trading day t, respectively. Holding period returns (item ret) and value-weighted returns (item vwretd) are used in order to be able to make the computations.

3.1.2 Size

Building upon the CAPM, Fama and French (1993) came up with the three factor asset pricing model by adding size and value anomalies to the model as proxies for risk, where the economic intuition depicts that small companies should be expected to provide larger returns than bigger firms, *i.e.* small minus big (SMB). Therefore, the size anomaly can be proxied for by the market capitalization, which in turn can be resembled by multiplying the number of shares outstanding (item csho) with the price (item prc).

3.1.3 Value

Another anomaly proposed by Fama and French (1993) is the value factor, *i.e* high minus low (HML), or value versus growth. There exists a plethora of factors based upon the general value factor. A subset of the value factors is described in Ledoit et al. (2019), from which one is depicted below.

The dividend yield-to-price ratio (D/P) is one of the measures for the value characteristic. It is measured by the total dividends paid out from the previous year, that is $t^* - 252$ through $t^* - 1$, divided by the one day lagged market capitalization. The total dividends paid out are computed by accumulating daily dividends. In order to calculate the daily dividend, the difference between cum- and ex-dividend returns are taken, which are resembled by the holding period return (item ret) and return without dividends (item retx). There are many more ways to resemble the value characteristic, however, this paper is restricted to the aforementioned measure, like described in Ledoit et al. (2019).

3.1.4 Momentum

The most famous example for the momentum trend signal is the standard 12-1 months momentum investigated by Jegadeesh and Titman (1993), which is defined as the cumulative return of the prior 12 months with the exclusion of the most recent month. Long-short portfolio strategies based on this approach have shown to be able to generate abnormal positive returns. This anomaly is also detectable across different asset classes and markets as shown in Moskowitz, Ooi, and Pedersen (2011) and Asness, Moskowitz, and Pedersen (2013), respectively. The momentum pricing anomaly is depicted in this research paper because of its simple economic intuition, where it was added to the Fama and French factors, introduced as the four-factor model by Carhart (1997).

In this research paper, the 11-month momentum factor score is based upon the geometric average stock return over the previous twelve months, excluding the most recent month. Hence, the factor score is calculated by computing the geometric average stock return on portfolio construction day z, using holding period return (item ret) from $t - 252$ through $t - 22$.

3.2 Sorting

First of all, the decision has to be made whether to perform a sort based on a single factor, or a sequential sort is taken into account. This section describes the methodology used in this research paper in order to perform a sort on the set of stocks.

3.2.1 Single factor sorting

If only a single factor is taken into account, the following strategy is implemented. Let $\{(1),\ldots,(N)\}$ be a permutation of $\{1,\ldots,N\}$ that generates a ranked vector of factor scores from smallest to largest in accordance with the specific anomaly at each portfolio construction day z, denoted as follows.

$$x_{(1)z} \leq x_{(2)z} \leq \ldots \leq x_{(N)z}$$

After ordering the factor scores, quantile-based bins are constructed. For each time period z, Q disjoint portfolios P are formed. More specifically, the portfolios, P, are denoted as follows.

$$P_{qz} = \left(x_{\left(\frac{N(q-1)}{Q}\right)z}, x_{\left(\frac{Nq}{Q}\right)z} \right], \qquad \forall \quad q = 1, \ldots, Q \quad and \quad z = 1, \ldots, Z$$

Furthermore, the assumption is taken that the fraction $\frac{N}{Q}$ is equal to an integer. Hence, the portfolios will designate the assets in whole meaning that a certain asset is available only to a certain portfolio.

3.2.2 Conditional sorting

If a sort on another factor is desired, the characteristic values, x, are ranked once more from smallest to largest in accordance to the second factor, within each portfolio P_{qz} separately. In short, section 3.2.1 takes the whole set of N assets and ranks it according to some factor x. Then, it divides the whole set of ranked assets in subsets depending on the specific quantile chosen. From here on, this section ranks these subsets once again. Where these subsets, $i.e$ pre-formed portfolios, are considered as the whole set of assets, when ranking them. This is more formally denoted as follows. Let $\left\{[1], [2], \ldots, \left[\frac{N}{Q}\right]\right\}$ be a permutation of $\left\{\left(\frac{N}{Q}(q-1)+1\right), \left(\frac{N}{Q}(q-1)+2\right), \ldots, \left(\frac{N}{Q}(q-1)+\frac{N}{Q}\right)\right\}$, $\forall$ $q = 1, \ldots, Q$. This generates a ranked vector of factor scores from smallest to largest in accordance with the second anomaly within each of the Q portfolios at each portfolio construction day z, denoted as follows.

$$x_{q[1]z} \leq x_{q[2]z} \leq \ldots \leq x_{q[\frac{N}{Q}])z}$$

Then once more, quantile-based bins are constructed within the Q existing portfolios at portfolio construction day z. L disjoint portfolios P are formed at each portfolio construction day. More specifically, the portfolios P are denoted as follows.

$$P_{qlz} = \left(x_{q\left[\frac{\frac{N}{Q}(l-1)}{L}\right]z}, x_{q\left[\frac{\frac{N}{Q}l}{L}\right]z} \right], \qquad \forall \quad l = 1, \ldots, L, \quad q = 1, \ldots, Q \quad and \quad z = 1, \ldots, Z$$

Hence, after the conditional sort, $Q * L$ portfolios at each portfolio construction day, z, are obtained, where each portfolio comprises $\frac{N}{Q*L}$ assets.

3.3 Portfolio construction

Let us denote the portfolio returns on a portfolio construction day by $\boldsymbol{R_z^P} := \boldsymbol{R_z'w_z}$. The weight vector is defined by $\boldsymbol{w_z} := (w_{1z}, ..., w_{Nz})'$. The weights should satisfy the following condition for a single factor portfolio on the portfolio construction day z.

$$\sum_{P_{1z}} |w_{iz}| = \sum_{P_{Qz}} |w_{iz}| = 1 \quad (1)$$

And the weights should satisfy the following condition for a portfolio based on two anomalies.

$$\sum_{P_{11z}} |w_{iz}| = \sum_{P_{QLz}} |w_{iz}| = 1 \quad (2)$$

Where all weights of the assets contained in the P_{Qz} and P_{QLz} portfolios should be positive for the single factor model and the double factor model, respectively. Next to that, the short requirement comes from the fact that all weights of the assets contained in the P_{1z} and P_{11z} portfolios should be negative for the single factor model and the double factor model, respectively. The weights of the assets that are not in the extreme quantiles are forced to zero to form the dollar-neutral portfolio as denoted below for the single factor model and the double factor model. This makes sure that the portfolio adds up to zero, whereas the gross exposure of the portfolio sums up to two at portfolio construction day z.

Regarding the top and bottom quantile portfolios P_{QLz} and P_{11z}, the weights can be optimized in a similar fashion as the GMV problem. Estimating the GMV portfolio can be seen as a thorough problem in terms of evaluating the quality of a covariance matrix estimator, since it abstracts from having to estimate the vector of expected returns at the same time. Furthermore, empirical results of MV portfolios are often deemed to be very poor, whereas the GMV portfolio tends to be much more successful. Let w_z^T denote the vector of weights at time z of the top quantile portfolio and w_z^B of the bottom quantile portfolio. Hence, the optimization to be solved here for the top quantile portfolio is the following.

$$min_w \, w^{T\prime} \boldsymbol{\hat{H}_z} w^T \quad (3)$$

Such that the following holds.

$$w^{T\prime}\iota = 1$$

Where $\hat{\boldsymbol{H}}_{\boldsymbol{z}}$ is the DCC-NL estimator of the unobservable conditional covariance matrix Σ_z. The solution to this problem is the following.

$$w^{T*} = \frac{\hat{\boldsymbol{H}}_{\boldsymbol{z}}^{\mathbf{-1}}\iota}{\iota'\hat{\boldsymbol{H}}_{\boldsymbol{z}}^{\mathbf{-1}}\iota} \tag{4}$$

Where the whole set of derivations is represented in Appendix A.1. In a similar manner the results for the bottom quantile portfolio are obtained, where the weights are replaced by w_z^B and the constraint has to sum up to minus one instead of plus one. Furthermore, from the derived solution it becomes clear once more that the weights are found to be independent from the notoriously hard to estimate expected returns μ and do only depend on the estimated inverted covariance matrix $\hat{\boldsymbol{H}}_{\boldsymbol{z}}^{\mathbf{-1}}$.

Hence, the optimal weights w^* in the dollar-neutral factor portfolio P^* are distributed as follows for a single sort portfolio.

$$w_z^* := \begin{cases} w_z^{T*} & \text{for } P_{Qz} \\ w_z^{B*} & \text{for } P_{1z} \\ 0, & otherwise \end{cases}$$

The optimal weights with respect to a double sort are as follows.

$$w_z^* := \begin{cases} w_z^{T*} & \text{for } P_{QLz} \\ w_z^{B*} & \text{for } P_{11z} \\ 0, & otherwise \end{cases}$$

Finally, the optimal portfolio returns are denoted by $R_t^{P*} := R_t' s_t$. s_t specifies the number of shares hold after portfolio construction, until the next portfolio construction day. The number of shares s_t hold constant at a specific trading day t, refer to the portfolio construction day $z = \frac{t-K+21}{21}$, where the fraction will be rounded down for all cases where the fraction is not divided evenly. The number of shares hold are based upon the portfolio weights w^*, where they equal after portfolio construction. However, the number of shares remain constant over time, whereas the weights fluctuate. This fluctuation of the weights depends on the evolving price. In order to remain a dollar-neutral portfolio it is most probable that a daily rebalance is needed, which would incur additional transaction costs. Hence, the number of shares are hold constant instead of the weights, where the portfolios are updated only once a month, *i.e* 21 trading days, since this seems to be standard in the finance literature to reduce the turnover.

Regarding the benchmark, it can be viewed from two perspectives. One perspective is that the benchmark is considered to be the 'naive' portfolio as it is, where the whole universe of assets is assigned a long position with equal weights, $\frac{1}{N}$. However, this contradicts with the long-short portfolio discussed in this research paper and would be an unfair comparison. Therefore, the naive version is modified such that it will be constructed by dividing the weights, summing up to one, equally among the top quantile portfolio at each portfolio construction day. For a single factor portfolio this will be $\frac{Q}{N}$ for the assets of the top P_{Qz} portfolio. For a double sort this will be $\frac{Q*L}{N}$ for the assets of the top P_{QLz} portfolio. The weights, $\frac{-Q}{N}$ and $\frac{-Q*L}{N}$, summing up to minus one, will be equally divided for the assets of the bottom P_{1z} and P_{11z} portfolios respectively.

3.4 DCC-NL estimator

In this section $\mathbf{\Lambda}^{N,T}$ and $\mathbf{\Phi}^{N,T}$ refer to the deterministic functions of the sample eigenvalues and hybrid eigenvalues, respectively. Γ is the unobservable population covariance matrix. Furthermore, the $2MSCLE$ method is an efficient solution to inverting T matrices of dimension N times N. A more detailed description of the variables and functions is provided in Appendix A.2.

The DCC-NL estimator is used in order to calculate the covariance matrix. This specific estimator is chosen because it tackles two challenges, which are frequently encountered nowadays. Firstly, the estimator handles the time variation problem within the DCC part. Secondly, it deals with the curse of dimensionality in the NL part by being able to cope with large dimensions of stocks up to the order of one thousand. Ledoit et al. (2019) provided a summary of the DCC-NL estimator proposed by R. F. Engle et al. (2017) which is reproduced below.

1. Fit univariate GARCH models to devolatilize returns.

2. Compute the sample covariance matrix of devolatilized returns.

3. Decompose it into eigenvalues and eigenvectors.

4. Invert an approximation of the function $\mathbf{\Lambda}^{N,T}$ to estimate population eigenvalues.

5. Apply an approximation of the function $\mathbf{\Phi}^{N,T}$ to shrink eigenvalues non-linearly.

6. Recompose with the sample eigenvectors to estimate the unconditional covariance matrix Γ in Equation 18.

7. Transform the resulting estimator of Γ from a covariance matrix to a proper correlation matrix.[1]

8. Maximize the $2MSCLE$ composite likelihood to estimate the correlation dynamics.

9. Recombine the estimated conditional correlation matrix with the estimated univariate GARCH processes to obtain an estimated conditional covariance matrix.

The outside steps (1–2 and 7–9) compose the DCC part, while the inside steps (3–6) compose the NL part of the DCC-NL estimation procedure. The final product is a time series of T N-dimensional conditional covariance matrix estimates, which will be called $\{\hat{H}^t\}_{t=1}^{T}$. More detailed steps that will be used in this research paper in order to make use of the DCC-NL estimator are provided in Appendix A.2 and more explicit formulas are provided in R. F. Engle et al. (2017).

3.5 Portfolio testing

The following hypothesis is considered in order to test the constructed portfolios whether they are able to predict the variation within markets of the cross-section of expected returns.

$$H_0 : \mathbb{E}(R_t^{P^*}) \leq 0 \qquad versus \qquad H_1 : \mathbb{E}(R_t^{P^*}) > 0$$

[1] This is driven by the fact that Γ itself is a proper correlation matrix.

The hypothesis test is based upon the generated portfolio returns for each trading day t, where the used Student test statistic, t^{stat}, is defined as follows.

$$t^{stat} := \frac{\bar{R}^{P^*}}{s.e.(R^{P^*})} \tag{5}$$

Where the average portfolio return in the numerator is represented by the following.

$$\bar{R}^{P^*} := \frac{1}{T}\sum_{t=1}^{T} R_t^{P^*} \tag{6}$$

s.e.(.) denotes a standard error of (.). In the literature it is frequently chosen to use the naive standard error based upon the assumption of independent and identically distributed (i.i.d.) returns. More specifically, it is defined as $\frac{s.d.}{\sqrt{T}}$, where s.d. denotes the sample standard deviation of the observed returns $R_t^{P^*}$, for all trading days t. Nonetheless, Ledoit et al. (2019) advocate to use a heteroskedasticity and autocorrelation consistent (HAC) standard error estimator such that the used standard errors are robust against heteroskedasticity and serial correlation in the portfolio returns. In particular, the HAC standard error in the paper of Ledoit et al. (2019) is based upon the quadratic spectral (QS) kernel with automatic choice of bandwidth as described in Andrews (1991). In this paper, a Parzen kernel is depicted, since the QS kernel appears to be somewhat slow in the R programming language. Nonetheless, the deviation from the QS to the Parzen kernel should have little to no impact on the eventual results, since simulation studies have shown that both kernels yield virtually identical results. In addition, Monte Carlo results have shown that prewhitening is very effective in reducing bias and over-rejection of t-statistics constructed using kernel HAC estimators. Hence, a prewhitened Parzen kernel with automatic choice of bandwidth is added as well, based on the work of Andrews and Monahan (1992).

Nonetheless, the weights are determined by conducting the GMV portfolio, whereas Ledoit et al. (2019) optimized for a MV portfolio. The MV optimization is more suitable for this specific test, since the t^{stat} is defined as a Sharpe Ratio, whereas the GMV optimization minimizes the variance and thus the denominator only in this test. However, researchers have established that estimated GMV portfolios have desirable out-of-sample properties, not only in terms of risk however, also in terms of reward-to-risk, that is in terms of the information ratio. With respect to the t^{stat}, Harvey et al. (2016) found that the common historical cutoff of two, meaning that if the t-statistic is larger than two the strategy is deemed successful, is not sufficient nowadays. The critical value should be in proportion to the specific factor addressed due to multiple testing issues.

Furthermore, a second test will be performed where the difference of the Sharpe Ratios between the optimized portfolio returns and the benchmark portfolio returns is tested. More specifically, a two-sided p-value for the null hypothesis $H0 : \Delta = 0$ is given by the following.

$$\hat{p} = 2\Phi\left(-\frac{|\hat{\Delta}|}{s(\hat{\Delta})}\right) \tag{7}$$

Where Φ denotes the cumulative distribution function (cdf) of the standard normal distribution. Δ refers to the difference between the two Sharpe Ratios. $s(\hat{\Delta})$ denotes the standard error of the difference. A summary of the computations is provided in Appendix A.3, where the overall test is based upon the work of Ledoit and Wolf (2008). Once more, the computations of this test are made based upon the choice of a (prewhitened) Parzen kernel with automatic choice of bandwidth. This kind of test is chosen since the research paper is to a great extent written from the perspective of a portfolio manager. In practice, results are compared in a relative sense, meaning that one has to compare its results to some kind of benchmark, instead of regarding absolute results. In order to calculate the Sharpe Ratios the risk-free return is reduced from the portfolio's return and then divided by the simple standard deviation from the excess return of the specific portfolio.

By extension of backtesting the results, one could look at the turnover metric in order to measure if the portfolio returns are not shrunk too much by potential transaction costs. The computation of the turnover metric is done as follows. The change of the weights for every two consecutive portfolio construction days is calculated. The first turnover value is based on the second portfolio construction day, compared to the first portfolio construction day. Therefore, there will always be one less observation for the turnover, compared to the amount of portfolio returns. First of all, the unique stocks of the second comparator are determined from which the weights are summed up. Secondly, the stocks that are in both portfolios should be checked for differences as well. In this respect, the distinction should be made between the assets that partly gained or lost weight allocation in comparison with the previous portfolio construction day. In order to avoid double counting, only the assets that obtained more weight allocation in the latter portfolio construction day are taken into consideration. If the turnover values of the unique stocks and the stocks that remained the same but changed in weight allocation are summed up, this results in the overall turnover value.

4 Data

The data used in this research paper has become available by downloading daily stock return data from the Center for Research in Security Prices (CRSP)[2], which maintains historical stock return data of the New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and National Association of Securities Dealers Automated Quotations (NASDAQ) stock exchanges, going back to 1926. The main data set used in this research paper dates from 1 January 1990 through 31 December 2019, comprising stocks with a share code of 10 or 11. The main data set consists of over 17 thousand different assets. Finally, the out-of-sample period ranges from 8 December 1994 through 15 December 2019, resulting in 300 months.

Since there does not seem to be one specific empirical data set that should be used in practice, the main data set is split into two data sets at each portfolio construction day for completeness of the empirical analysis conducted. An overview of all the filters and assumptions used to obtain the two different data sets is mentioned in Appendix D, however, the important filters and assumptions are discussed here as well. The two data sets are referred to as the Ledoit, Wolf and Zhao (L-W-Z) data set and the Fama and French (F-F) data set. The assumption is taken that one month is represented by 21 consecutive trading days. For both data sets, the main data set is firstly filtered at each portfolio construction day for stocks that have recent past data of 1250 trading days, *i.e* five years. Five years is taken since the DCC-NL estimator requires the largest time frame of recent past data of all calculations used

[2] `http://www.crsp.org/`

in order to estimate the second moment over time. Secondly, the main data set is filtered for future return data of one period in which the portfolio assets are held constant in order to provide a genuine portfolio return. The latter, forward-looking restriction is not feasible in real life, however, with a portfolio management perspective, this is commonly applied in the related finance literature on the out-of-sample evaluation of portfolios. After filtering and cleaning the data, the L-W-Z data set comprises out of selecting one thousand assets, based upon the highest market capitalization as is done by Ledoit et al. (2019). Overall, the dimensionality remains equivalent and thus the final results can be compared with the results generated in the Ledoit et al. (2019) paper. The F-F data set excludes financial firms[3], since this is common practice in research and advocated by Fama and French (1992). They claim that the high leverage, which is normal for these firms, probably does not have the same meaning as for non-financial firms, where high leverage more likely indicates distress. Furthermore, the F-F data set is restricted to five thousand randomly selected stocks. This contrasts with the selection method favored by Ledoit et al. (2019), because this research paper allocates a large weight on the size factor and choosing stocks based upon market capitalization hints towards a bias in that case. Moreover, both data sets satisfy the high dimensionality categorization, which is to be dealt with in the optimization problem.

With respect to the observed variables in order to calculate the factor scores and portfolio returns, the daily data comprises of holding period returns (item ret), holding period returns without dividends (item retx), prices (item prc), number of shares outstanding (item csho) and index value-weighted returns including all distributions (item vwretd).

5 Empirical analysis

The purpose of this paper is to construct dollar-neutral portfolios that are exposed to one or two given factors within the top and bottom quintiles. In order to do so, the two data sets were depicted by filtering for the different assumptions, resulting in the L-W-Z based data set and F-F based data set. The summary statistics of the value weighted daily return

[3] All firms with a Standard Industrial Classification (SIC) code between 6.000 and 6.999.

series of the two data sets are displayed in Table 1. The minimum, mean, median and maximum represent the global statistics, where the volatilities and Sharpe Ratios are the average volatilities and Sharpe Ratios. In order to provide more insights into the general returns, the table is constructed in a panel data format, where the statistics of three equally divided time periods are provided as well. After cleaning and filtering the main data set of over 17 thousand stocks, accumulated over all the portfolio construction days, the F-F data set eventually comprises a universe of 2839 designated stocks, whereas the L-W-Z data set has a universe of 2959 stocks, represented by $\tilde{N}$ in Table 1. Over time there seems to appear some kind of consolidation, since the more recent the time period, the less companies pass the filters set out in order to come up with a stable stock universe. At the portfolio construction days, the number of stocks is a subset of the total amount $\tilde{N}$ and exactly equivalent to 1000 for the L-W-Z data set, *i.e.* N. Because of filtering the amount of stocks at each single portfolio construction day for the F-F data set ranges from 633 to 1362.

Comparing the two data sets shows that the research paper is dealing with data sets that work within the same bandwidth, regarding the extreme values of a minimum value weighted daily return of -2.99% and a maximum of 11.49% for both data sets. Both the extreme observations occured in the 2003 to 2011 time period. However, if the actual returns are regarded instead of the value weighted returns more extreme values are presented. An example is the maximum return with an outlier of 345.77% on a specific trading day t within the F-F data set. This company finds itself to be a relatively small company. This seems to be the case, since it is filtered out of the L-W-Z data set based on the market capitalization assumption. Hence, this specific observation might feature the size anomaly. However, the same reasoning holds for the minimum return value of -91.34% in the F-F data set, where the latter analogy can be reversed. This is based on the fact that the concerned company has a relatively small market capitalization as well and thus showing the possible connection to the added risk following from small firms. In addition to the risk, the highest volatility of 1.46% and 1.51% is observed in the 2003 through 2011 period, however the highest return is observed in this time frame as well, which could explain the increase in risk. In terms of the Sharpe Ratios, the highest values are observed in the 1995-2003 time period, where the results from 2011 through 2020 appear to pull the overall Sharpe Ratio down to 0.09.

Data set	$\tilde{N}$	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)	SR
1995 - 2020							
F-F	2839	-2.99	0.11	0.12	11.49	1.18	0.09
L-W-Z	2959	-2.99	0.11	0.13	11.49	1.19	0.09
1995 - 2003							
F-F	2338	-2.84	0.14	0.11	4.59	1.12	0.13
L-W-Z	2088	-2.84	0.14	0.10	4.59	1.14	0.12
2003 - 2011							
F-F	1498	-2.99	0.13	0.13	11.49	1.46	0.09
L-W-Z	1700	-2.99	0.15	0.15	11.49	1.51	0.10
2011-2020							
F-F	994	-2.17	0.04	0.17	2.15	0.82	0.05
L-W-Z	1505	-2.17	0.04	0.17	2.15	0.82	0.05

Table 1: Data return summary

Notes: This table shows the summary statistics for the value weighted daily return series of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, from 1995 through 2020. σ and SR represent the average volatilities and Sharpe Ratios.

After the data sets are constructed, one of the four factor scores are computed for all portfolio construction days Z, based on the definitions provided in Section 3.1. Subsequently, the single sort is performed in accordance with the chosen anomaly, from which a top and bottom portfolio is selected. The selection is established by using quintiles. If portfolios were, for example, split into more than five quantiles, like terciles, the focus would be narrowed down to a relatively small stocks universe. Although, this approach might work well statistically, it is not practical from a portfolio management perspective. It focuses a lot of the trading volume onto such a small amount of assets that it may not be implementable in practice without drastic price impact for institutional asset sizes. Hence, 200 stocks are listed in each of the extreme portfolios for the L-W-Z data set. Within the F-F data set the amount of stocks per portfolio construction day is variable due to filtering. Hence, the extreme portfolios of the F-F data set comprise an amount of stocks ranging from 127 to 272. Then, an additional factor is computed and ranked for the assets within the extreme portfolios, separately. Once more, a selection based upon quintiles is taken from these extreme portfolios. The assets within the top quintile based upon the single sort will form the basis of the stocks within the top portfolios of the conditional sort. Moreover, the specific selection of assets from the bottom portfolios of the single sort will consequently form the bottom portfolios of the conditional sort. Eventually, an investment universe of 40 stocks within each extreme portfolio of the specific L-W-Z data set and a range of 25 to 54 stocks within each extreme portfolio of the F-F data set at each portfolio construction day is left.

Once the extreme portfolios are created, the DCC-NL estimator is adopted in order to estimate the covariance matrix. The DCC-NL estimator itself is based on the most recent five years of return data of the specific stocks within the extreme portfolios. After that, the weights are divided by optimizing with the focus on the GMV profile, where the specific assumptions and requirements are described in Section 3.3. Next to that, the results for the MV profile are constructed in order to obtain insights in whether the statement that this strategy produces poor results in empirical analysis also holds in this specific case. From here on, the portfolio results can be computed, which is done by calculating the compounded returns for all 300 portfolio construction days, given the optimized weights.

In order to have a more intuitive interpretation of the results yielded by the implemented strategies, benchmarks were constructed as described in Section 3.3. Each strategy has a specific benchmark appointed, where the benchmark returns are computed by dividing the weights equally among the designated stocks of the extreme quintiles, for both the long and short positions and the F-F and L-W-Z data sets. Additionally, for each data set a naive benchmarks was constructed where the weights are divided equally in a long position over the whole assets universe of the two different data sets, prior to the segregation of the data sets into quintiles.

5.1 Portfolio return statistics

Section 3.1 describes the used factors. Hence, this provides four different constructions for the single sorted portfolios, whereas the sequential sorting provides twelve combinations. Since the focus in this paper is laid upon the size anomaly, the monthly return summary statistics of the portfolio based on the size anomaly for the single sorted portfolio is highlighted below in Table 2 as well as its benchmarks. In addition, the statistics for the combinations based on the size sort and the benchmark portfolios as discussed in Section 3.3 are included within Table 3. The results of the other portfolios are represented in Appendix B.1. With respect to the MV optimized weights, very similar results are obtained for all the different metrics used. For that reason and in order to preserve space, the tables that represent the results of the MV profile will not be displayed. Z displays the amount of portfolio construction days which depict the compounded results that were obtained over the previous month. Hence, in total an amount of 300 months, or 25 years of monthly portfolio returns are observed.

Note that all the displayed results are in percentages and that none of the statistics are annualized. The bandwidth, in which the results emerge, becomes more extreme as one goes from a single sort to a conditional sort. In the GMV setting for a single sort the portfolio monthly returns lie within -28.10% and 36.04%, whereas they lie within -52.92% and 102.43% for the double sort over all the strategies. When the attention is focused on the single sort of the size anomaly, the difference between the summary statistics of the F-F and L-W-Z data sets is remarkable. Applying the strategy to the F-F data set seems to do a decent job, where the strategy has not the desired outcome in the L-W-Z data set. This might be due to the exclusion of the financial firms in the F-F data set, although there is no theoretical reason why the size factor is less effective for financial firms. The only interpretation on this behalf is that leverage has a completely different value for financial firms in comparison with non-financial firms. Another reason could be that the deviation is caused by the filtering for the thousand largest stocks at each portfolio construction day for the L-W-Z data set. This filter reduces the effect of the size factor, which seems to be one of the most proper signals in this paper. This tendency is also visible regarding the modified benchmarks. Furthermore, comparing the strategies with their appropriate modified benchmarks, the results are quite similar, where both benchmarks slightly outperform in terms of returns. However, if only a long position and no specific quintiles are taken into consideration, like in the naive benchmarks, the exclusion of financial firms do not seem to have as much effect as in the modified benchmarks. Moreover, the naive benchmark for the L-W-Z data set outperforms the size sorted strategy substantially. This is not the case for the naive benchmark within the F-F data set that is underperforming the size sorted strategy. Furthermore, conditionally sorting based on the size factor does not improve upon the single sort strategy in terms of the summary statistics.

Regarding the risk in terms of the volatility does not hint at extreme differences between the different strategies for the same data set. For visualization the compounded return of the size sorted portfolio in a GMV profile and its benchmarks in the F-F data set is depicted in Figure 1. This figure provides an indication of how much money would have been made by an investor who started investing at the 12th of December 1994, with an initial investment of zero for the size and modified strategy, since those strategies go long one dollar and short one dollar, and one for the naive strategy. Given this time sample, the investor would have peaked around 2018 in terms of portfolio value given that he applied the size sorted strategy. Although this looks very promising, transaction costs are not taken into account, which could absorb a lot of the profit. Next to that, the line representing this value is in some ways misleading, as a quick glance at the solid line would seem to indicate that the returns were much more volatile toward the end of the sample period than at the beginning. This result is simply due to the scale however, as the same percentage gain or loss is indicated by a larger vertical distance on the chart for the more recent periods, as the cumulative returns are larger toward the end to the sample period than at the beginning of the sample period.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F size sorted	300	-14.21	2.41	1.90	36.04	6.40
L-W-Z size sorted	300	-11.29	0.13	0.10	16.32	3.89
Benchmark size sorted (F-F)	300	-8.83	2.24	1.36	35.30	5.95
Benchmark size sorted (L-W-Z)	300	-10.28	0.08	0.18	14.03	2.96
Naive benchmark (F-F)	300	-29.81	1.78	1.96	25.76	5.90
Naive benchmark (L-W-Z)	300	-28.56	1.14	1.46	18.68	5.23

Table 2: Size portfolio return summary statistics

Notes: This table shows the monthly size sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, in a global minimum variance setting. In addition, the size sorted benchmark statistics for both data sets as well as the naive benchmarks are included.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F sorted on beta	300	-29.19	1.20	0.57	33.46	7.90
L-W-Z sorted on beta	300	-25.80	0.08	-0.43	46.06	6.84
Benchmark sorted on beta (F-F)	300	-28.88	2.40	1.37	32.31	8.24
Benchmark sorted on beta (L-W-Z)	300	-26.17	-0.06	-0.17	37.42	6.41
F-F sorted on value	300	-16.77	1.61	0.79	25.90	6.72
L-W-Z sorted on value	300	-17.76	-0.17	-0.44	29.34	5.42
Benchmark sorted on value (F-F)	300	-10.15	1.88	0.90	27.92	6.44
Benchmark sorted on value (L-W-Z)	300	-15.72	-0.01	-0.03	17	4.40
F-F sorted on momentum	300	-19.20	2.36	1.34	57.65	8.85
L-W-Z sorted on momentum	300	-38.47	-0.03	-0.05	58.49	7.00
Benchmark sorted on momentum (F-F)	300	-17.20	2.08	1.28	43.72	7.24
Benchmark sorted on momentum (L-W-Z)	300	-38.70	0.20	0.00	42.77	6.56

Table 3: Sequentially sorted portfolio return summary statistics based on size sort

Notes: This table shows the monthly sequentially sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting. In addition, the sequential sorted benchmark statistics for both data sets are included as well.

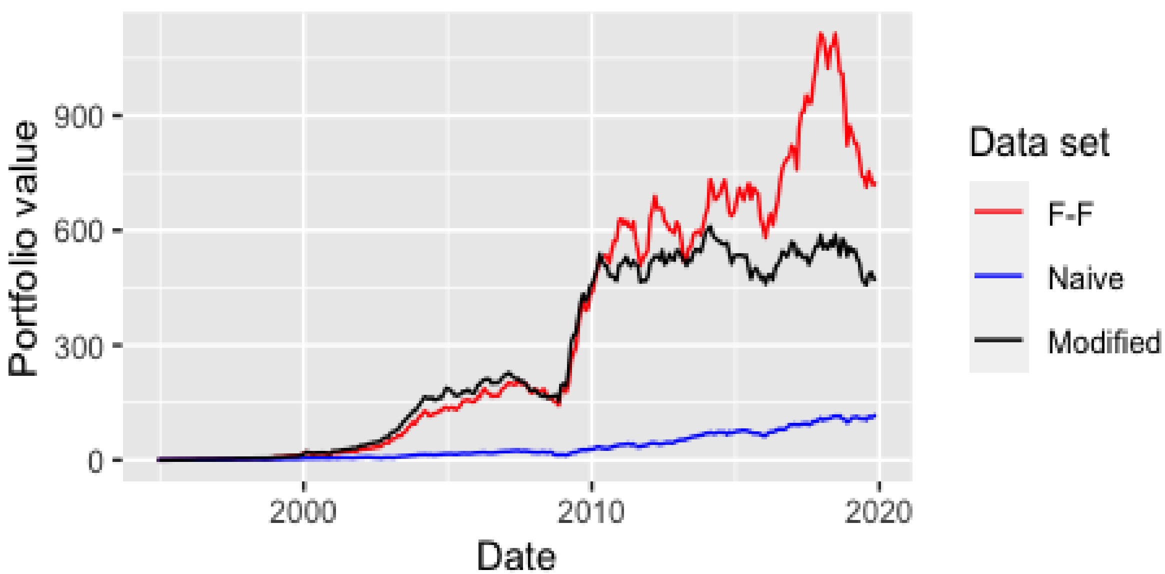

Figure 1: Cumulative return of the size sorted strategy

Notes: This figure shows the cumulative return of the size sorted long-short strategy in a global minimum variance setting for the French and Fama (F-F) data set. In addition, the modified as well as the naive benchmark cumulative returns are depicted as well.

5.2 Portfolio weights

Another set of percentages that could be of interest are the summary statistics of the weights, which resulted from the DCC-NL estimator, which was plugged in as the covariance matrix before optimizing. Tables 4 and 5 represent the summary statistics of the weights of the size sorted portfolios in a GMV context. The remaining set of statistics can be found in Appendix B.2. The strategies are constructed in a long-short manner, where the weights of the long as well as the short quintiles add up to one and minus one, respectively. Hence, the average weights should be equal to zero. Therefore, in order to avoid meaningless values, the absolute value of the weights in the short portfolios are taken before computing the summary statistics. Regarding these statistics, for all portfolios it holds that at least for one portfolio construction day a weight of almost 0% is accredited. Exactly 0% is not possible as per definition of the GMV portfolio, however, some of the weights are so small that a less than sign is used to indicate that the weight is not exactly equal to zero. The research paper deals with a very large dimension of stocks. Even though the weights are divided among quintiles there are still plenty of assets to choose from and in a parallel fashion determine the *bad* assets. Taking a closer look at the size sorted strategies, the minima, means and medians are all close or equal to zero, whereas the maxima are extremely large. Once more, the observation of a fat-tailed distribution of the weights is not eccentric, since the investment problem is optimized taking into consideration the volatility and dividing the weights over more stocks diversifies the systematic risk. Regarding the weights of the modified benchmarks based on the L-W-Z data set, each weight is set at 0.50% for the single sorted portfolios and at 2.50% for the double sorted portfolios. Hence, the optimized weights resemble the naive portfolios, unless the optimizer clinches some specific assets to stand out and addressing more or less weight to them.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F size sorted	300	< 0.01	0.51	0.20	81.80
L-W-Z size sorted	300	< 0.01	0.50	0.00	88.40

Table 4: Portfolio weights summary statistics based on the size sort

Notes: This table shows the monthly size sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	300	< 0.01	2.55	1.00	80.20
L-W-Z sorted on beta	300	< 0.01	2.50	0.40	83.20
F-F sorted on value	300	< 0.01	2.55	0.60	88.40
L-W-Z sorted on value	300	< 0.01	2.50	0.20	87.60
F-F sorted on momentum	300	< 0.01	2.55	1.00	88.80
L-W-Z sorted on momentum	300	< 0.01	2.50	0.20	85.80

Table 5: Portfolio weights summary statistics based on the sequential size sort

Notes: This table shows the monthly sequentially sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting.

5.3 Student's t-statistics

The name of the GMV strategy itself already denotes that it is designed to minimize the variance, rather than to maximize the expected return. Therefore, any portfolio that implements the GMV portfolio should be primarily evaluated by how successfully it achieves this goal. Nevertheless, a high out-of-sample average return and a high out-of-sample Sharpe Ratio are naturally desirable as well. Hence, whereas the portfolio return statistics only take into account the returns, two tests are performed in order to backtest the portfolio performances, taking into consideration the goal of the GMV strategy by introducing standard errors, which are subsequently linked to the variance. The one-sided null hypothesis is defined for the expected returns to be lower than zero. In order to support for the evidence of the null hypothesis a Student's t-statistic is used. With respect to the first hypothesis test, the Student's t-statistics of the portfolios based on the size anomaly and the sequential sorts thereon are represented in Tables 7 and 8, respectively. The other portfolios are included in Appendix B.3. In order to calculate the t-statistics three standard errors were taken into account. The simple sample standard error, the HAC standard error and the prewhitened HAC standard error. The latter two were both computed by using a Parzen kernel with automatic choice of bandwidth.

Not surprisingly, in some cases the t-statistic is negative. From the perspective of a portfolio manager the assumption can be made that the corresponding strategies will be discarded by a researcher, since they can never be established as successful based on a negative t-statistic. Therefore, in the evaluation of the t-statistics the attention is restricted to strategies that yield a positive t-statistic. The common cutoff in the literature for the value of a t-statistic is two. However, in a recent paper, Harvey et al. (2016) argue that a critical value of two is not sufficient anymore and should be updated to three because of multiple-testing issues. Hence, if the t-statistic transcends the cutoff the null hypothesis is rejected in favor of the alternative hypothesis. This is desirable, since it provides evidence in favor of the expected returns being higher than zero. Table 6 regards both critical values, two and three. Next to that, Ledoit et al. (2019) have shown that it is in the best interest of researchers to use as large an investment universe as possible since the proportion of significant factors increases in the investment universe. Hence, the choice of dealing with a very large dimension in this research paper should pay off in the amount of significant t-statistics as well, which are depicted in Table 6. The latter statement seems to be confirmed by the F-F data set, since two out of four of the single sorted strategies and nine out of the twelve sequentially sorted strategies in a GMV context have a positive t-statistic. In fact, seven out of sixteen strategies survive the historical cutoff of two, where five out of sixteen strategies survive the condition of having a critical value of three or higher, where the condition is set by Harvey et al. (2016).

The empirical advantage of the newly proposed procedure, considering the relatively good t-statistics and high returns, does not come from its ability to select more informative stocks, however, Ledoit et al. (2019) have shown in their relatively similar paper that this instead must come from its ability to minimize exposure to unpriced risk. Hence, in spite of the fact that the means and medians of the general data sets are close to zero, indicating the difficulty of obtaining large t-statistics given the way the t-statistic is constructed, the DCC-NL estimator, however, turned out to be able to generate large t-statistics when used in the optimization problem of non-zero weights in pre-sorted quintiles for the F-F data set. Next to that, the poor performance of the L-W-Z data set is confirmed by the t-statistics, since only six out of sixteen strategies have a t-statistic above zero. Moreover, not a single strategy surpasses the the critical value of two, let alone three. The fact that the L-W-Z data set is heavily underperforming should be explained by not excluding financials in data sets or by filtering on the one thousand largest stocks at each portfolio construction day, since these two are the only deviations of the applied procedure in comparison to the F-F data set. If the attention is paid to the size sorted strategy as well as the conditionally sorted strategies based on the sort size for the F-F data set, the results of the t-statistics stand out by all strategies surpassing the critical cutoff of two, where only the strategy that is sequentially sorted on beta falls slightly short to surpassing the critical value of three. Hence, the presentiment provided by the return summary statistics are approved by the t-statistics, showing that the high return statistics are not offset by an even higher volatility.

With respect to the single or conditional sorts a distinction can be made for the different factors and which ones are best to use in a portfolio strategy, according to this empirical analysis within the F-F data set. Regarding the t-statistics of the single sorts, the beta and the size sort yield the highest values. This trend continues in the sequential sorts. For every strategy based on one of the four applied factors in the first sort, however, sorted on a conditional sort based on either beta or size increases the t-statistic. The same reasoning holds for the value and momentum factor in the opposite way, where choosing one of the two in a secondary sort lowers the t-statistic. For example, the strategy based on the size sort, shown in the table below, illustrates this observation. Regarding the t-statistic* in the F-F data set, the size sort resulted in a value of 4.67. The t-statistic* of the conditional size sort, where the first sort is based on beta resulted in a value of 5.29, which is depicted in the appendix.

Data set	t-statistics ≥ 0	t-statistics ≥ 2	t-statistics ≥ 3
F-F single sorted	2 (50%)	2 (50%)	1 (25%)
L-W-Z single sorted	3 (75%)	0 (0%)	0 (0%)
F-F conditionally sorted	9 (75%)	5 (42%)	4 (33%)
L-W-Z conditionally sorted	3 (25%)	0 (0%)	0 (0%)

Table 6: Summary of Student's t-test results

Notes: This table shows the enumerations of the Student's t-statistics above zero, two and three for the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the global minimum variance setting. For each case, where all three differently calculated t-statistics of a specific strategy fulfill the requirement as mentioned in the columns of the table, one is accredited. In brackets is the percentage of fulfilled requirements divided by the total of possible strategies.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F size sorted	300	4.67	5.11	6.53
L-W-Z size sorted	300	0.61	0.61	0.60

Table 7: Student's t-test results based on the size sort

Notes: This table shows the Student's t-statistics for the portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F sorted on beta	300	2.82	2.98	2.63
L-W-Z sorted on beta	300	0.23	0.23	0.20
F-F sorted on value	300	3.42	4.04	5.06
L-W-Z sorted on value	300	-0.32	-0.32	-0.31
F-F sorted on momentum	300	3.57	3.90	4.61
L-W-Z sorted on momentum	300	-0.08	-0.09	-0.08

Table 8: Student's t-test results based on the sequential size sort

Notes: This table shows the Student's t-statistics for the sequentially sorted portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

5.4 Sharpe Ratio differences

The first hypothesis test only considers the applied strategy, where in order to backtest the portfolio returns even more, a two-sided hypothesis is constructed. In order to find out whether the applied strategy provides better results than the benchmark results, the two-tailed hypothesis states that the difference between the Sharpe Ratio of the applied strategy and the Sharpe Ratio of the benchmark equals zero. The second hypothesis is measured by a p-value, where the difference of the Sharpe Ratios between the optimized portfolio returns and the benchmark returns are tested. The Sharpe Ratios in this empirical analysis are computed by dividing the mean of the return series of the specific portfolio by the standard error of the same return series. In order to calculate the p-values, two standard errors were taken into account. The HAC standard error and the prewhitened HAC standard error. The latter two were both computed by using a Parzen kernel with automatic choice of bandwidth. The results of the single and sequentially sorted portfolios are displayed in Table 10 and 11, respectively. The other strategies can be found in Appendix B.4, where the exact computation, in order to calculate the p-values, is described in Appendix A.3. The rule of thumb is administered, where a p-value smaller or equal than 0.05 is considered to indicate strong evidence against the null hypothesis. *I.e* the null hypothesis is rejected if the p-value is smaller or equal to 0.05. A p-value larger than 0.05 indicates weak evidence against the null hypothesis and can therefore, not be rejected. Since the null hypothesis test is defined as a two-tailed test, one can still consider the direction of the value that is inferred on. Ideally, the null hypothesis gets rejected, whereas the Sharpe Ratio of the applied strategy exceeds the Sharpe Ratio of the benchmark, meaning that the strategy outperforms the benchmark with confidence. There will be referred to this as being the best case scenario.

Table 9 depicts the number of best case scenarios per strategy as well as the amount of cases where the null hypothesis is rejected. For the strategies applied in the F-F data set in a GMV context three out of sixteen cases show strong evidence against the null hypothesis and therefore, reject it. However, none of the scenarios where the null hypothesis is rejected showed a difference of the Sharpe Ratios in favor of the applied strategy. Moreover, all of the cases wherein the null hypothesis is rejected, the differences in the Sharpe Ratios are to the benefit of the benchmarks. Hence, once more the naive investment strategy described by DeMiguel et al. (2009) appears to be a tough competitor. Considering the size sort based strategies in Tables 10 and 11, including the naive long only benchmark, the applied strategies still do not seem to outperform the benchmarks. Although, the size sorted strategy in the F-F data set comes close by outperforming the modified naive benchmark, the evidence of this scenario happening most of the time is still not strong enough. If the size sorted portfolio is compared with the simplistic benchmark, there seems to be a quite strong outperformance since the Sharpe Ratios are 0.38 and 0.30, respectively. However, the hypothesis that the difference between the ratios is equal to zero is not rejected given that both p-values state the probability of the strategies ending up with the exact same performance is likely to happen in 21% of the cases. Applying the strategies to the L-W-Z data set does not improve upon the results of the F-F data set in terms of the Sharpe Ratios. However, in this case there is only one strategy where the null hypothesis is rejected, in favor of the benchmark. In that respect, the L-W-Z data set actually does better by resembling the respective benchmarks. The strategies applied to the F-F data set were able to produce similar results to their benchmarks in thirteen out of sixteen cases, where the L-W-Z data set provides fifteen out of sixteen such scenarios.

Data set	p-values ≤ 0.05	(SR 1 > SR 2) | p-value ≤ 0.05
F-F single sorted	0 (0%)	0 (0%)
L-W-Z single sorted	0 (0%)	0 (0%)
F-F conditionally sorted	3 (25%)	0 (0%)
L-W-Z conditionally sorted	1 (8%)	0 (0%)

Table 9: Summary of p-value results

Notes: This table shows the enumerations of the p-values below 0.05 for the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the global minimum variance setting. For each case where both p-values of a specific strategy fulfill the requirement as mentioned in the column of the table, one is accredited. Additionally, the amount of transcendments in Sharpe Ratio of the strategy versus the benchmark, given the p-value exceeded 0.05, is provided as well. In brackets is the percentage of fulfilled requirements divided by the total of possible strategies.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F size sorted vs. size sorted benchmark (F-F)	300	0.98	0.98	0.38	0.38
L-W-Z size sorted vs. size sorted benchmark (L-W-Z)	300	0.88	0.87	0.03	0.03
F-F size sorted vs. naive benchmark (F-F)	300	0.21	0.21	0.38	0.30
L-W-Z size sorted vs. naive benchmark (L-W-Z)	300	0.00	0.00	0.03	0.22

Table 10: p-values Sharpe Ratio test based on the size sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the size sorted portfolio returns in a global minimum variance setting and the benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set as well as the naive benchmarks. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F sorted on beta vs. sorted on beta benchmark (F-F)	300	0.00	0.00	0.15	0.29
L-W-Z sorted on beta vs. sorted on beta benchmark (L-W-Z)	300	0.56	0.58	0.01	-0.01
F-F sorted on value vs. sorted on value benchmark (F-F)	300	0.00	0.00	0.29	0.48
L-W-Z sorted on value vs. sorted on value benchmark (L-W-Z)	300	0.82	0.81	-0.02	0.00
F-F sorted on momentum vs. sorted on momentum benchmark (F-F)	300	0.55	0.55	0.27	0.29
L-W-Z sorted on momentum vs. sorted on momentum benchmark (L-W-Z)	300	0.40	0.43	0.00	0.03

Table 11: p-values Sharpe Ratio test based on the sequential size sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the sequentially sorted portfolio returns in a global minimum variance setting and the sequentially sorted benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

In order to make the comparison between the applied strategy and the methodology used by Ledoit et al. (2019), the Sharpe Ratio difference backtest is computed once more. The comparison between these two strategies does not fall within the scope of the research conducted in this paper, however, the results are depicted in Appendix C out of curiosity. No exact conclusion can be made from the results, since the applied strategy only outperforms in half of the cases in terms of the Sharpe Ratio. In addition, both strategies significantly outperform each other only once in terms of the Sharpe Ratio difference backtest.

5.5 Portfolio turnover

Although, the procedure of optimizing the portfolios did not integrate any kind of control against turnover, the turnover values are calculated in order to gain more knowledge of whether the eventual results are substantiated. The values for the turnover are represented below in Tables 12 and 13 for the size sorted strategies in a GMV setting. The turnover ratios for the remaining optimized portfolios can be found in Appendix B.5. In order to compute the turnover metric, the difference in the weights for every consecutive portfolio construction day, in this case each month, is taken into account. The percentage change of specific assets that are invested in two consecutive months is added up to the sum of the weights of the assets that do only appear in the succeeding month. Since the sum of the weights is restricted to one and the weights themselves are limited between zero and one within each extreme quintile, the value of the turnover must lie between zero and one as well. *I.e.* the turnover values lie between 0% and 100%. Note that there are only 299 reference points instead of 300, since at each portfolio construction day the change is measured with the previous portfolio, resulting in one less observation since for the very first portfolio there is no former portfolio to compare with. The bandwidth in which the turnover lies is 1.66% and 99.80% for the strategies based on a GMV setting in the F-F data set, with an average turnover of 59.94%. For the strategies based on the L-W-Z data set the range is from a minimum turnover of 4.44% to a maximum of 100.00%, with the mean being equal to 66.01%. Note that a 100.00% turnover depicts that at some portfolio construction day all current assets were sold and replaced by a whole set of different stocks. The observation that for the GMV setting the average turnover in general decreases, going from the single sorts to the sequentially sorted strategies confirms the intuition that the more assets available at each portfolio construction day, the more one would alter between the assets.

If the size sorted strategies are taken into account for the F-F data set, which are displayed in Tables 12 and 13, one can observe that these strategies are doing a relatively decent job in comparison to the other portfolios as well in terms of turnover. All portfolios based on a size sort have an average turnover lower or close to the overall average turnover of 59.94% for the F-F data set and 66.01% for the L-W-Z data set. The same accounts for the maximum turnover of the size sorted portfolios. In addition, Ledoit et al. (2019) have shown that the average turnover in strategies, which optimize the covariance matrix as well as take factor signals into account, can mostly be explained by changes in the factor scores themselves, rather than by the covariance matrix. Hence, the fact that the turnover of the strategy that is conditionally sorted on momentum acts as a misfit with an average of 81.60% might make sense, since it feels more intuitive for momentum to differ each portfolio construction day, than the size of a company. Nevertheless, half the portfolio construction days have more than 60% turnover. This observation somewhat invalidates the promising results discussed before.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F size sorted	299	16.99	66.03	62.80	94.60
L-W-Z size sorted	299	45.10	85.32	86.70	99.80

Table 12: Portfolio turnover summary statistics based on the size sort

Notes: This table shows the monthly size sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	299	21.45	55.59	55.60	91.90
L-W-Z sorted on beta	299	29.20	60.61	62.15	92.20
F-F sorted on value	299	19.03	58.14	59.30	85.60
L-W-Z sorted on value	299	26.90	60.91	61.20	98.40
F-F sorted on momentum	299	44.60	81.60	82.49	99.80
L-W-Z sorted on momentum	299	32.70	75.50	77.00	99.60

Table 13: Portfolio turnover summary statistics based on the sequential size sort

Notes: This table shows the monthly sequentially sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting.

6 Conclusion

In this research paper the standard methods of sorting portfolios have been investigated on two types of data sets, whereby the DCC-NL estimator is introduced in order to optimize the non-zero weights within the extreme quintiles. In this respect, four anomalies are taken into account, namely beta, size, value and momentum. Furthermore, not only single sorts are analysed, however, conditional sorted portfolios as well, where it turned out that the size factor excelled. Therefore, the size anomaly is highlighted in this conclusion.

In accordance with the standard sorting techniques and in addition the theory of minimum variance optimization, the research paper demonstrates that the portfolio construction in predictive tests of cross-sectional factors seem to be a proper solution to the question of how to optimize the non-zero weights for large investment universes when a suitable estimator of the covariance matrix of stock returns are incorporated. When a researcher upgrades from the standard sorting approaches and includes an optimizer for the non-zero weights within the quintiles of the extreme portfolios based on the DCC-NL covariance matrix estimator, Student's t-statistics exceed the recently updated critical value of three across a broad panel of return-predictive signals when the investment universe is large and constructed according to the F-F data principles. These high t-statistics are required for an anomaly to be deemed as successful since multiple-testing issues may justify raising the t-statistic significance threshold from the historical cutoff of two to a more demanding level of three. With respect to the sequentially sorted strategies the performances have a hard time keeping up with the results yielded from the single sorted portolios.

In addition to the first backtest, where the Student's t-statistics are used, a second backtest is performed. This test hypothesizes the difference of the Sharpe Ratios of the optimized portfolios against the naive and modified benchmarks to equalize zero. It follows that the modified benchmarks are tough competitors, where none of the applied strategies significantly outperform their specific benchmark.

Next to that, the size anomaly appears to be underestimated in the search for high returns, since the strategies based on a single sort show one of the best results across the signals investigated in this research paper. Additionally, the kind of data set that is used to base the research on appears to play an important role in the search for return-predictive strategies. The data set that produced decent returns did not take into account firms that have a financial SIC code. Furthermore, it did not exclude some percentage of companies based on their market capitalization, which reduces the size factor effectiveness a priori. Nevertheless, the turnover of the optimized portfolios can not be neglected. However, for statistical arbitrage hedge funds that specialize in systematically exploiting large ensembles of cross-sectional factors it is fairly representative in practice to have high turnover given their trade volumes. Moreover, their automated order-placing system usually brings the transaction costs down to a range as low as three to five basis points. Hence, from a portfolio management perspective, the gravity of the influence of the turnover ratio is subsidiary to the type of manager or investor that wants to apply the strategy.

In an extension to this research paper, some topics for further research are provided. One could apply the same methods and procedures as discussed in this paper to other factors or quantiles. In that respect, not only different quantiles can be investigated however, also one could take a look at other quintiles, instead of only taking into consideration the extremes. Furthermore, the focus of the research could be laid upon the kind of data used, as the returns generated based on two different data sets in this research paper turned out to be contrasting. Perhaps one could dive into the exclusion of a percentage of lower market capitalization firms, since this anomaly has proven to be underestimated. By placing a filter on this character one could possibly reduce the size signal effectiveness in advance. Next to that, some turnover control could be implemented in order to reduce transaction costs for portfolio managers, without affecting the returns. Furthermore, one could use more accurate univariate models than the straightforward GARCH(1,1) model to devolatilize the individual return series in the first step of the DCC-NL procedure, such as models that incorporate asymmetric responses or intraday prices.

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817–858.

Andrews, D. W., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica: Journal of the Econometric Society*, 953–966.

Asness, C., Moskowitz, T., & Pedersen, L. (2013). Value and momentum everywhere. *The Journal of Finance*, *No. 68*.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, *9*(1), 3–18.

Berk, J. B. (1995). A critique of size-related anomalies. *The Review of Financial Studies*, *8*(2), 275–286.

Black, F. (1992). Beta and return", presentation at the berkeley program in finance: Are betas irrelevant. *Evidence and Implications for Asset Management*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57–82.

Chen, N.-f., Copeland, T. E., & Mayers, D. (1987). A comparison of single and multifactor portfolio performance methodologies. *Journal of Financial and Quantitative Analysis*, *22*(4), 401–417.

Daniel, K., Mota, L., Rottke, S., & Santos, T. (2020). The cross-section of risk and returns. *The Review of Financial Studies*, *33*(5), 1927–1979.

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

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F., Ledoit, O., & Wolf, M. (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, *37*(2), 363–375.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of*.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Haugen, R. A., & Baker, N. L. (1991). The efficient market inefficiency of capitalization–weighted stock portfolios. *The Journal of Portfolio Management*, *17*(3), 35–40.

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

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91.

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

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *The Journal of Finance*, *36*(4), 889–908.

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

Lambert, M., & Hübner, G. (2013). Comoment risk and stock returns. *Journal of Empirical Finance*, *23*, 191–205.

Lambert, M., & Hübner, G. (2014). Size matters, book value does not! the fama-french empirical capm revisited. *The Fama-French Empirical CAPM Revisited (May 1, 2014)*.

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

Ledoit, O., Wolf, M., & Zhao, Z. (2019). Efficient sorting: a more powerful test for cross-sectional anomalies. *Journal of Financial Econometrics*, *17*(4), 645–686.

Lintner, J. (1965a). Security prices, risk, and maximal gains from diversification. *The journal of finance*, *20*(4), 587–615.

Lintner, J. (1965b). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic optimization models in finance* (pp. 131–155). Elsevier.

Lo, A., & MacKinlay, A. (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies*, *3*(1), 431–468.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. Retrieved from `http://www.jstor.org/stable/2975974`

Memmel, C. (2003). Performance hypothesis testing with the sharpe ratio. *Finance Letters*, *1*(1).

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*(1), 31–42.

Moskowitz, T., Ooi, Y. H., & Pedersen, L. H. (2011). Momentum crashes. *Chicago Booth Research Paper*, *No 12-21*.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, 768–783.

Nakagawa, K., Imamura, M., & Yoshida, K. (2018). Risk-based portfolios with large dynamic covariance matrices. *International Journal of Financial Studies*, *6*(2), 52.

Nielsen, F., & Aylursubramanian, R. (2008). Far from the madding crowd*–volatility efficient indices— april 2008.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425–442.

Treynor, J. L. (1961). Market value, time, and risk. *Time, and Risk (August 8, 1961)*.

Treynor, J. L. (1962). Jack treynor's' toward a theory of market value of risky assets'. *Available at SSRN 628187*.

Appendices

A Derivations and testing methods

A.1 Derivation of optimization problem

The following objective is optimized.

$$min_w \, w' \boldsymbol{\hat{H}_t} w \tag{8}$$

Such that

$$w'\iota = 1$$

The Lagrangian is as follows.

$$L = \frac{1}{2} w' \boldsymbol{\hat{H}_t} w - \lambda(w'\iota - 1) \tag{9}$$

When the first order conditions of the Lagrangian is taken, the following is obtained.

$$L_w = \boldsymbol{\hat{H}_t} w - \lambda\iota = 0 \tag{10}$$

$$L_\lambda = 1 - w'\iota = 0 \tag{11}$$

Re-arranging these first order conditions provides the following.

$$w = \lambda \boldsymbol{\hat{H}_t^{-1}} \iota \tag{12}$$

$$\iota' w = 1 \tag{13}$$

Then filling in equation 12 into equation 13 and rewriting for lambda gives the following.

$$\lambda \iota' \boldsymbol{\hat{H}_t^{-1}} \iota = 1 \iff \lambda = \frac{1}{\iota' \boldsymbol{\hat{H}_t^{-1}} \iota} \tag{14}$$

If equation 14 is filled in into 12, the system for the weight with respect to minimizing the variance is solved.

$$w_{GMV} = \frac{\boldsymbol{\hat{H}_t^{-1}} \iota}{\iota' \boldsymbol{\hat{H}_t^{-1}} \iota} \tag{15}$$

A.2 DCC-NL estimator

This brief recapitulation is only intended to make the present paper self-contained and is reproduced from Ledoit et al. (2019). The interested reader is referred to R. F. Engle et al. (2017) for the original exposition.

A.2.1 Time variation in the second moment

The modelling and estimation of time-varying variances, covariances, and correlations requires aggregating the contributions from three different ideas.

A key idea promoted by R. Engle (2002) is that modelling conditional heteroskedasticity is easy and successful in the univariate case. Hence, one should take care of that prior to looking at the covariance matrix as a whole. Thus, for every stock $i = 1, \ldots, N$, a GARCH(1,1) is fitted. Dividing the raw returns by the corresponding conditional standard deviations yields devolatilized returns that have unit variance. Let s_t denote the N-dimensional column vector of devolatilized residuals at trading day t, where $t = 1, \ldots, T$. Then the dynamics of the pseudo-correlation matrix Q_t can be specified as follows.

$$Q_t = \Theta + \alpha s_{t-1} s'_{t-1} + \beta Q_{t-1} \tag{16}$$

where α and β are non-negative scalars satisfying $\alpha + \beta < 1$ that govern the dynamics, and Θ is an N-dimensional symmetric positive definite matrix. Q_t is called a pseudo-correlation matrix because its diagonal terms are close, but not exactly equal, to one. Therefore, the following adjustment is needed to recover the proper correlation matrix R_t as described in the following equation.

$$R_t = Diag(Q_t)^{\frac{-1}{2}} \, Q_t \, Diag(Q_t)^{\frac{-1}{2}} \tag{17}$$

where Diag(.) denotes the function that sets all the off-diagonal elements of a matrix to zero.

The second ingredient is the notion of variance targeting. Although originally invented in a univariate context, the extension to the multivariate case of interest here is straightfoward. The basic idea is that a suitable rescaling of the matrix Θ in equation 16 can be interpreted as the unconditional covariance matrix. Therefore, it can be estimated using standard techniques that ignore time series effects, separately from the other parameters. This approach yields the reparametrized model described below.

$$Q_t = (1 - \alpha - \beta)\Gamma + \alpha s_{t-1}s'_{t-1} + \beta Q_{t-1} \tag{18}$$

where Γ is the long-run covariance matrix of the devolatilized returns s_t for each time t.[4]

After having dealt with the conditional variances and partialled out the problem of estimating the unconditional covariance matrix, the only remaining task is to estimate the dynamic correlation parameters α and β. These two scalars play the same role as their counterparts in the more familiar ARMA(1,1) and GARCH(1,1) models, but for conditional correlation matrices.

When the matrix dimension is large, say $N = 1000$, the standard likelihood maximization technique would require inverting T matrices of dimension 1000×1000 at every iteration, which is numerically challenging. A more efficient solution is called the 2MSCLE method. This method combines the individual likelihoods generated by 2×2 blocks of contiguous variables. Maximizing this composite likelihood yields asymptotically consistent estimators for α and β, as long as the DCC model is well-specified. The intuition is that every individual correlation coefficient shows traces of the dynamic parameters α and β in its own time series evolution, so a sufficiently large subset of individual correlations will reveal (a consistent approximation of) the true parameters. The advantage of this procedure is that it is numerically stable and fast in high dimensions.

The DCC Estimation Procedure can be summarized into three steps, as described below.

1. Fit a univariate GARCH(1,1) model to every stock return series individually, and divide the raw returns by their conditional standard deviations to devolatilize them.

[4]Since the devolatilized returns all have unit variance, Γ is actually a proper correlation matrix, that is, its diagonal elements are all equal to one.

2. Estimate the unconditional covariance matrix of devolatilized returns somehow.

3. Maximize the $2MSCLE$ composite likelihood to obtain consistent estimators of the two parameters of correlation dynamics in a numerically stable and efficient way.

At this juncture, it becomes apparent from step 2 that one needs an estimator of the unconditional covariance matrix of devolatilized returns that performs well when the dimension is large.[5]

A.2.2 Estimation of large-dimensional unconditional covariance matrices

The textbook estimator of Γ is the sample covariance matrix $\Sigma := \sum_{t=1}^{T} \frac{\boldsymbol{s_t s_t'}}{T}$. Both matrices admit spectral decompositions as follows.

$$\Sigma = \sum_{i=1}^{N} \lambda_i.\boldsymbol{u_i u_i'} \qquad \textit{and} \qquad \Gamma = \sum_{i=1}^{N} \tau_i.\boldsymbol{v_i v_i'} \tag{19}$$

where $(_1, \ldots, \lambda_N)$ and $(u_1, \ldots, u_N)$ denotes a system of eigenvalues and eigenvectors of the sample covariance matrix $\hat{\Sigma}$, and $(\tau_1, \ldots, \tau_N)$ and $(v_1, \ldots, v_N)$ denotes a system of eigenvalues and eigenvectors of the population covariance matrix Γ. Eigenvalues are indexed in ascending order without loss of generality.

In the traditional asymptotic framework, where the sample size T goes to infinity, while the number of assets N remains constant, the sample eigenvalue λ_i is a consistent estimator of its population counterpart τ_i, and the sample eigenvector u_i is a consistent estimator of its population counterpart v_i, for all assets i. However, this asymptotic framework is not robust against the curse of dimensionality. When N is no longer negligible with respect to T, the sample spectrum is far from its population counterpart.

This is why it is necessary to turn to another asymptotic framework that offers a different family of analytical solutions. Unlike the formulas from traditional asymptotics, they work also if N is greater than T. The key assumption is that the ratio $\frac{N}{T}$ converges to some limit

[5]Note that in practice the devolatilized returns have to be based on estimated univariate GARCH models rather than the true, unobservable univariate GARCH models.

c $\in [0, +\infty)$ called the concentration (ratio). This framework is called large-dimensional asymptotics, and it includes traditional (fixed-dimensional) asymptotics as a special case when the concentration c is equal to zero. Thus, it is a generalization of traditional asymptotics that is able to cope with the curse of dimensionality by making necessary corrections (whose intensity increases in c) to the standard formulas.

In the absence of a priori knowledge about the structure of the eigenvectors of the (unobservable) population covariance matrix Γ, estimators should preserve the sample covariance matrix eigenvectors $(u_1, \ldots, u_N)$, and correct the sample eigenvalues only. This framework is called rotation-equivariant because the economic outcome is immune to repackaging the N original stocks into a collection of N funds investing in these stocks, as long as the funds span the same investment universe as the stocks.

It is easy to show that, among rotation-equivariant estimators of the covariance matrix, the one that performs the best across all possible linear constraints for the purpose of portfolio selection in terms of minimizing out-of-sample variance is the following.

$$\tilde{\Sigma} := \sum_{i=1}^{N} (\boldsymbol{u_i'} \boldsymbol{\Gamma} \boldsymbol{u_i}).\boldsymbol{u_i} \boldsymbol{u_i'} \tag{20}$$

This makes economic sense because $\boldsymbol{u_i'\Gamma u_i}$ is the out-of-sample variance of the portfolio whose weights are given by the ith sample eigenvector u_i. Thus, the emergence of a third quantity is noted, after the sample eigenvalue $\lambda_i = u_i'\Sigma u_i$ and the population eigenvalue $\tau_i = v_i \Gamma v_i$, the hybrid $\phi_i = \boldsymbol{u_i'\Gamma u_i}$ is observed, which represents the best possible thing to do with the sample eigenvectors.

The key is that, under large-dimensional asymptotics, the vectors $\boldsymbol{\lambda} := (\lambda_i)_{i=1,\ldots,N}$, $\boldsymbol{\tau} := (\tau_i)_{i=1,\ldots,N}$ and $\boldsymbol{\phi} := (\phi_i)_{i=1,\ldots,N}$ are all far apart from one another. It is only as the concentration c goes to zero, that is, as the standard (fixed-dimension) asymptotics are approached, that their mutual differences vanish. When c > 0, which is the case when the investment universe is large, appropriate corrections must be applied to go from $\boldsymbol{\lambda}$ to $\boldsymbol{\tau}$ to $\boldsymbol{\phi}$.[6] Qualita-

[6]Correcting these relationships when the ratio of variables to observations is significant is analogous to correcting Newtonian relationships when the ratio of velocity to speed of light is significant.

tively, $\boldsymbol{\lambda}$, $\boldsymbol{\tau}$ and $\boldsymbol{\phi}$ have the same cross-sectional average, however $\boldsymbol{\lambda}$ is more dispersed than $\boldsymbol{\tau}$ that in turn is more dispersed than $\boldsymbol{\phi}$.

The ideal would be to have two deterministic functions $\boldsymbol{\Lambda^{N,T}}$ and $\boldsymbol{\Phi^{N,T}}$ from $[0, +\infty)^N$ to $[0, +\infty)^N$ mapping out the two important expectations.

$$\boldsymbol{\tau} \longmapsto \boldsymbol{\Lambda^{N,T}(\tau)} := (\boldsymbol{\Lambda_1^{N,T}(\tau)}, \ldots, \boldsymbol{\Lambda_N^{N,T}(\tau)}) = (\mathbb{E}[\lambda_1], \ldots, \mathbb{E}[\lambda_N]) = (\mathbb{E}[\boldsymbol{u_1'\Sigma u_1}], \ldots, \mathbb{E}[\boldsymbol{u_N'\Sigma u_N}])$$

$$\boldsymbol{\tau} \longmapsto \boldsymbol{\Phi^{N,T}(\tau)} := (\boldsymbol{\Phi_1^{N,T}(\tau)}, \ldots, \boldsymbol{\Phi_N^{N,T}(\tau)}) = (\mathbb{E}[\phi_1], \ldots, \mathbb{E}[\phi_N]) = (\mathbb{E}[\boldsymbol{u_1'\Gamma u_1}], \ldots, \mathbb{E}[\boldsymbol{u_N'\Gamma u_N}])$$

Then, the observed eigenvalues of the sample covariance matrix, $\boldsymbol{\lambda}$, would be used to reverse-engineer an estimator of the population eigenvalues by solving the following optimization problem.

$$\hat{\boldsymbol{\tau}} := argmin_{t \in [0,+\infty)^N} \frac{1}{N} \sum_{i=1}^{N} (\Lambda_i^{N,T}(\boldsymbol{t}) - \lambda_i)^2 \tag{21}$$

and the non-linear shrinkage estimator of the covariance matrix would follow as.

$$\hat{\Sigma} := \sum_{i=1}^{N} \boldsymbol{\Phi_i^{N,T}(\hat{\tau}).u_i u_i'} \tag{22}$$

Due to tractability issues, however, one only knows approximations to the functions $\boldsymbol{\Lambda^{N,T}}$ and $\boldsymbol{\Phi^{N,T}}$ that are valid asymptotically as the universe dimension N goes to infinity along with the sample size T, with their ratio $\frac{N}{T}$ converging to the concentration c. Replacing the true expectation functions with their approximations can be done at no loss, asymptotically. Therefore, this procedure yields a non-linear shrinkage estimator of the covariance matrix that is optimal in the large-dimensional asymptotic limit.

Qualitatively speaking, the effect of composing $\Phi^{N,T}$ with the inverse of $\boldsymbol{\Lambda^{N,T}}$ (or approximations thereof) moves the sample eigenvalues closer to one another, while preserving their cross-sectional average. The effect is increasing in $\frac{N}{T}$ and highly non-linear.

A.3 Testing with Sharpe Ratio

In order to perform a test on the difference off the Sharpe Ratio, the methodology described by Ledoit and Wolf (2008) is used and summarized below.

A.3.1 Notation

Take two investment strategies i and j with excess returns $r_{t,i}$ and $r_{t,j}$ for a time period $t = 1, \ldots, T$. Assume that $r_{t,i}$ and $r_{t,j}$ are strictly stationary time series processes with,

$$\mu = \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix} \quad and \quad \Sigma = \begin{pmatrix} \sigma_i^2 & \sigma_{i,j} \\ \sigma_{i,j} & \sigma_j^2 \end{pmatrix}$$

The difference between both Sharpe Ratios, depicted by SR, is then given by the following.

$$\Delta = SR_i - SR_j = \frac{\mu_i}{\sigma_i} - \frac{\mu_j}{\sigma_j}$$

The Sharpe Ratios are estimated by using the sample moments, which results in the following.

$$\hat{\Delta} = \hat{SR}_i - \hat{SR}_j = \frac{\hat{\mu}_i}{\hat{\sigma}_i} - \frac{\hat{\mu}_j}{\hat{\sigma}_j}$$

Denote $\mathbb{E}(r_{1,i}^2) = \gamma_i$ and $\mathbb{E}(r_{1,j}^2) = \gamma_j$ with their estimates $\hat{\gamma}_i$ and $\hat{\gamma}_j$. Further define $v = (\mu_i$, μ_j, γ_i, $\gamma_j)$ and $\hat{v} = (\hat{\mu}_i$, $\hat{\mu}_j$, $\hat{\gamma}_i$, $\hat{\gamma}_j)$', such that Δ = f(v) and $\hat{\Delta}$ = f($\hat{v}$) with f(a,b,c,d) = $\frac{a}{\sqrt{c - a^2}}$ - $\frac{b}{\sqrt{d - b^2}}$. Under relatively mild regularity conditions on the higher moments of excess returns the following holds.

$$\sqrt{T}(\hat{v} - v) \xrightarrow{d} N(0, \Psi)$$

where Ψ is an unknown symmetric, semi-definite matrix. Further using the Delta method, the following is found.

$$\sqrt{T}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0, \nabla' f(v) \Psi \nabla f(v))$$

In our notation, this becomes the following.

$$\nabla' f(a,b,c,d) = \left(\frac{c}{(c - a^2)^{1.5}}, -\frac{d}{(d - b^2)^{1.5}}, -\frac{1}{2}\frac{a}{(c - a^2)^{1.5}}, \frac{1}{2}\frac{b}{(d - b^2)^{1.5}} \right)$$

If a consistent estimator $\hat{\Psi}$ is available, the standard error for $\hat{\Delta}$ is given by the following.

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}{T}}$$

A.3.2 HAC inference

The limiting covariance matrix Ψ is given by the following.

$$\Psi = \lim_{T \to \infty} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}[y_s y_t'], \qquad with \quad y_t' = (r_{t,i} - \mu_i, r_{t,j} - \mu_j, r_{t,i}^2 - \gamma_i, r_{t,j}^2 - \gamma_j)$$

With a limit that alternatively can be expressed as follows.

$$\Psi = \lim_{T \to \infty} \Psi_T \qquad with \quad \Psi_T = \sum_{m=-T+1}^{T+1} \Gamma_T(m)$$

Where $\Gamma_T(m)$ depicts the following.

$$\Gamma_T(m) = \begin{cases} \frac{1}{T}\sum_{t=m+1}^{T} \mathbb{E}(y_t y'_{t-m}) & \text{for} \quad m \geq 0 \\ \frac{1}{T}\sum_{t=-m+1}^{T} \mathbb{E}(y_{t+m} y'_t) & \text{for} \quad m < 0 \end{cases}$$

In order to come up with a consistent estimator for $\hat{\Psi} = \hat{\Psi}_T$, a standard group of HAC robust kernel estimators exist. For this research paper, this involves choosing a real-valued kernel function $k(.)$ and a bandwidth S_T. The kernel $k(.)$ typically satisfies $k(0) = 1$, $k(.)$ is continuous at zero and $\lim_{x \to \pm\infty} k(x) = 0$. Then, the kernel estimate for Ψ is the following.

$$\hat{\Psi} = \hat{\Psi}_T = \frac{T}{T-4} \sum_{m=-T+1}^{T-1} k\left(\frac{m}{S_T}\right) \hat{\Gamma}_T(m) \tag{23}$$

Where the following holds.

$$\hat{\Gamma}_T(m) = \begin{cases} \frac{1}{T}\sum_{t=m+1}^{T} \hat{y}_t \hat{y}'_{t-m} & \text{for} \quad m \geq 0 \\ \frac{1}{T}\sum_{t=-m+1}^{T} \hat{y}_{t+m} \hat{y}'_t & \text{for} \quad m < 0 \end{cases}$$

and

$$\hat{y}'_t = (r_{t,i} - \hat{\mu}_i,\, r_{t,j} - \hat{\mu}_j, r^2_{t,i} - \hat{\gamma}_i,\, r^2_{t,j} - \hat{\gamma}_j)$$

A.3.3 Hypothesis test with HAC inference

A two-sided p-value for the null hypothesis $H0 : \Delta = 0$ is given by the following.

$$\hat{p} = 2\Phi\left(-\frac{|\hat{\Delta}|}{s(\hat{\Delta})}\right)$$

Where $\Phi(.)$ denotes the cdf of the standard normal distribution. Alternatively, a 1 - α confidence interval for Δ is given by the following.

$$\hat{\Delta} \pm z_{1-\frac{\alpha}{2}} s(\hat{\Delta})$$

where z_λ denotes the λ quantile of the standard normal distribution.

B Figures and tables

B.1 Portfolio returns in a global minimum variance profile

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F beta sorted	300	-21.93	0.72	0.51	22.36	5.70
L-W-Z beta sorted	300	-28.10	-0.06	-0.40	28.41	7.17
Benchmark beta sorted (F-F)	300	-27.14	0.88	0.76	27.31	6.71
Benchmark beta sorted (L-W-Z)	300	-28.95	-0.13	-0.46	29.19	6.81

Table 14: Beta portfolio return summary statistics

Notes: This table shows the monthly beta sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting. In addition, the beta sorted benchmark statistics for both data sets are included as well.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F size sorted	300	-14.21	2.41	1.90	36.04	6.40
L-W-Z size sorted	300	-11.29	0.13	0.10	16.32	3.89
Benchmark size sorted (F-F)	300	-8.83	2.24	1.36	35.30	5.95
Benchmark size sorted (L-W-Z)	300	-10.28	0.08	0.18	14.03	2.96

Table 15: Size portfolio return summary statistics

Notes: This table shows the monthly size sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting. In addition, the size sorted benchmark statistics for both data sets are included as well.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F value sorted	300	-19.39	-0.40	-0.41	16.73	4.79
L-W-Z value sorted	300	-21.49	0.09	-0.15	22.97	5.28
Benchmark value sorted (F-F)	300	-19.43	-0.45	-0.47	12.79	3.96
Benchmark value sorted (L-W-Z)	300	-15.96	0.06	-0.17	26.75	4.27

Table 16: Value portfolio return summary statistics

Notes: This table shows the monthly value sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting. In addition, the value sorted benchmark statistics for both data sets are included as well.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F momentum sorted	300	-28.00	-0.53	-0.43	25.16	6.86
L-W-Z momentum sorted	300	-25.21	0.01	0.68	24.17	5.99
Benchmark momentum sorted (F-F)	300	-31.00	-0.69	-0.09	17.33	5.83
Benchmark momentum sorted (L-W-Z)	300	-31.74	0.22	0.39	31.99	6.00

Table 17: Momentum portfolio return summary statistics

Notes: This table shows the monthly momentum sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting. In addition, the momentum sorted benchmark statistics for both data sets are included as well.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F sorted on size	300	-15.76	3.82	2.26	56.84	9.18
L-W-Z sorted on size	300	-20.60	-0.14	-0.39	24.04	5.93
Benchmark sorted on size (F-F)	300	-22.35	4.13	3.12	58.34	9.14
Benchmark sorted on size (L-W-Z)	300	-22.44	-0.06	-0.55	31.59	5.98
F-F sorted on value	300	-37.32	0.34	0.15	33.43	10.05
L-W-Z sorted on value	300	-39.12	-0.26	-0.53	54.04	10.45
Benchmark sorted on value (F-F)	300	-44.21	0.31	-0.25	47.71	10.73
Benchmark sorted on value (L-W-Z)	300	-44.15	-0.04	-0.56	59.16	10.42
F-F sorted on momentum	300	-52.92	0.01	0.35	34.52	11.29
L-W-Z sorted on momentum	300	-48.31	-0.21	0.02	39.99	9.49
Benchmark sorted on momentum (F-F)	300	-46.72	0.18	0.31	37.47	10.85
Benchmark sorted on momentum (L-W-Z)	300	-43.30	-0.02	0.03	45.48	9.90

Table 18: Sequentially sorted portfolio return summary statistics based on beta sort

Notes: This table shows the monthly sequentially sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the beta sort in a global minimum variance setting. In addition, the sequential sorted benchmark statistics for both data sets are included as well.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F sorted on beta	300	-29.19	1.20	0.57	33.46	7.90
L-W-Z sorted on beta	300	-25.80	0.08	-0.43	46.06	6.84
Benchmark sorted on beta (F-F)	300	-28.88	2.40	1.37	32.31	8.24
Benchmark sorted on beta (L-W-Z)	300	-26.17	-0.06	-0.17	37.42	6.41
F-F sorted on value	300	-16.77	1.61	0.79	25.90	6.72
L-W-Z sorted on value	300	-17.76	-0.17	-0.44	29.34	5.42
Benchmark sorted on value (F-F)	300	-10.15	1.88	0.90	27.92	6.44
Benchmark sorted on value (L-W-Z)	300	-15.72	-0.01	-0.03	17.00	4.40
F-F sorted on momentum	300	-19.20	2.36	1.34	57.65	8.85
L-W-Z sorted on momentum	300	-38.47	-0.03	-0.05	58.49	7.00
Benchmark sorted on momentum (F-F)	300	-17.20	2.08	1.28	43.72	7.24
Benchmark sorted on momentum (L-W-Z)	300	-38.70	0.20	0.00	42.77	6.56

Table 19: Sequentially sorted portfolio return summary statistics based on size sort

Notes: This table shows the monthly sequentially sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting. In addition, the sequential sorted benchmark statistics for both data sets are included as well.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F sorted on beta	300	-34.79	-0.14	-0.08	29.36	8.55
L-W-Z sorted on beta	300	-39.71	0.03	-0.82	43.28	9.48
Benchmark sorted on beta (F-F)	300	-37.20	0.23	-0.37	32.10	8.99
Benchmark sorted on beta (L-W-Z)	300	44.24	0.13	-0.11	56.26	10.10
F-F sorted on size	300	-15.90	0.20	0.08	39.95	5.64
L-W-Z sorted on size	300	-18.27	-0.13	-0.27	16.88	4.41
Benchmark sorted on size (F-F)	300	-18.11	0.70	-0.10	18.78	5.31
Benchmark sorted on size (L-W-Z)	300	-19.48	0.11	-0.55	25.96	5.55
F-F sorted on momentum	300	-40.05	-1.41	-0.23	21.82	9.13
L-W-Z sorted on momentum	300	-26.40	-0.23	-0.26	44.76	7.52
Benchmark sorted on momentum (F-F)	300	-34.42	-1.36	-0.83	29.46	8.92
Benchmark sorted on momentum (L-W-Z)	300	-35.32	0.03	-0.04	41.80	7.37

Table 20: Sequentially sorted portfolio return summary statistics based on value sort

Notes: This table shows the monthly sequentially sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the value sort in a global minimum variance setting. In addition, the sequential sorted benchmark statistics for both data sets are included as well.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)	σ (in %)
F-F sorted on beta	300	-47.41	0.25	-0.21	38.77	11.08
L-W-Z sorted on beta	300	-38.28	-0.32	-0.42	50.18	9.60
Benchmark sorted on beta (F-F)	300	-42.01	0.57	0.62	50.30	10.73
Benchmark sorted on beta (L-W-Z)	300	-42.75	-0.11	0.06	53.85	9.83
F-F sorted on size	300	-23.75	1.89	1.52	102.43	10.70
L-W-Z sorted on size	300	-27.87	-0.19	-0.08	34.46	6.54
Benchmark sorted on size (F-F)	300	-28.26	1.53	1.90	46.42	8.44
Benchmark sorted on size (L-W-Z)	300	-38.26	0.42	0.50	44.94	6.89
F-F sorted on value	300	-43.11	-1.94	-1.45	52.41	11.38
L-W-Z sorted on value	300	-34.51	0.09	0.08	63.50	9.23
Benchmark sorted on value (F-F)	300	-44.14	-2.15	-1.61	33.63	10.33
Benchmark sorted on value (L-W-Z)	300	-34.77	0.03	0.50	43.71	8.26

Table 21: Sequentially sorted portfolio return summary statistics based on momentum sort

Notes: This table shows the monthly sequentially sorted portfolio return summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the momentum sort in a global minimum variance setting. In addition, the sequential sorted benchmark statistics for both data sets are included as well.

B.2 Weights statistics in a global minimum variance profile

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F beta sorted	300	< 0.01	0.51	0.37	68.00
L-W-Z beta sorted	300	< 0.01	0.50	0.50	71.60

Table 22: Portfolio weights summary statistics based on the beta sort

Notes: This table shows the monthly beta sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F size sorted	300	< 0.01	0.51	0.20	81.80
L-W-Z size sorted	300	< 0.01	0.50	< 0.01	88.40

Table 23: Portfolio weights summary statistics based on the size sort

Notes: This table shows the monthly size sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F value sorted	300	< 0.01	0.51	0.20	81.60
L-W-Z value sorted	300	< 0.01	0.50	0.50	88.60

Table 24: Portfolio weights summary statistics based on the value sort

Notes: This table shows the monthly value sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F momentum sorted	300	< 0.01	0.51	0.20	85.00
L-W-Z momentum sorted	300	< 0.01	0.50	< 0.01	86.20

Table 25: Portfolio weights summary statistics based on the Momentum sort

Notes: This table shows the monthly momentum sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on size	300	< 0.01	2.55	1.89	80.60
L-W-Z sorted on size	300	< 0.01	2.50	0.60	69.60
F-F sorted on value	300	< 0.01	2.55	1.89	61.40
L-W-Z sorted on value	300	< 0.01	2.50	1.00	76.80
F-F sorted on momentum	300	< 0.01	2.55	1.85	76.60
L-W-Z sorted on momentum	300	< 0.01	2.50	0.80	79.60

Table 26: Portfolio weights summary statistics based on the sequential beta sort

Notes: This table shows the monthly sequentially sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the beta sort in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	300	< 0.01	2.55	1.00	80.20
L-W-Z sorted on beta	300	< 0.01	2.50	0.40	83.20
F-F sorted on value	300	< 0.01	2.55	0.60	88.40
L-W-Z sorted on value	300	< 0.01	2.50	0.20	87.60
F-F sorted on momentum	300	< 0.01	2.55	1.00	88.80
L-W-Z sorted on momentum	300	< 0.01	2.50	0.20	85.80

Table 27: Portfolio weights summary statistics based on the sequential size sort

Notes: This table shows the monthly sequentially sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	300	< 0.01	2.55	1.89	68.20
L-W-Z sorted on beta	300	< 0.01	2.50	0.80	65.40
F-F sorted on size	300	< 0.01	2.55	1.00	89.80
L-W-Z sorted on size	300	< 0.01	2.50	0.20	88.40
F-F sorted on momentum	300	< 0.01	2.55	1.85	77.40
L-W-Z sorted on momentum	300	< 0.01	2.50	0.40	88.40

Table 28: Portfolio weights summary statistics based on the sequential value sort

Notes: This table shows the monthly sequentially sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the value sort in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	300	< 0.01	2.55	1.80	69.00
L-W-Z sorted on beta	300	< 0.01	2.50	0.80	75.00
F-F sorted on size	300	< 0.01	2.55	1.20	85.20
L-W-Z sorted on size	300	< 0.01	2.50	0.20	87.20
F-F sorted on value	300	< 0.01	2.55	1.85	68.40
L-W-Z sorted on value	300	< 0.01	2.50	0.40	84.20

Table 29: Portfolio weights summary statistics based on the sequential momentum sort

Notes: This table shows the monthly sequentially sorted portfolio weights summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the momentum sort in a global minimum variance setting.

B.3 Student's t-statistic results in a global minimum variance profile

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F beta sorted	300	2.19	2.61	2.19
L-W-Z beta sorted	300	-0.15	-0.16	-0.15

Table 30: Student's t-test results based on the beta sort

Notes: This table shows the Student's t-statistics for the portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the beta sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F size sorted	300	4.67	5.11	6.53
L-W-Z size sorted	300	0.61	0.61	0.60

Table 31: Student's t-test results based on the size sort

Notes: This table shows the Student's t-statistics for the portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F value sorted	300	-1.43	-1.42	-1.44
L-W-Z value sorted	300	0.37	0.35	0.30

Table 32: Student's t-test results based on the value sort

Notes: This table shows the Student's t-statistics for the portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the value sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F momentum sorted	300	-1.34	-1.35	-1.33
L-W-Z momentum sorted	300	0.03	0.03	0.03

Table 33: Student's t-test results based on the momentum sort

Notes: This table shows the Student's t-statistics for the portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the momentum sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F sorted on size	300	5.29	6.13	7.20
L-W-Z sorted on size	300	-0.43	-0.47	-0.42
F-F sorted on value	300	0.62	0.65	0.59
L-W-Z sorted on value	300	-0.39	-0.41	-0.43
F-F sorted on momentum	300	0.02	0.02	0.02
L-W-Z sorted on momentum	300	-0.39	-0.42	-0.39

Table 34: Student's t-test results based on the sequential beta sort

Notes: This table shows the Student's t-statistics for the sequentially sorted portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the beta sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F sorted on beta	300	2.82	2.98	2.63
L-W-Z sorted on beta	300	0.23	0.23	0.20
F-F sorted on value	300	3.42	4.04	5.06
L-W-Z sorted on value	300	-0.32	-0.32	-0.31
F-F sorted on momentum	300	3.57	3.90	4.61
L-W-Z sorted on momentum	300	-0.08	-0.09	-0.08

Table 35: Student's t-test results based on the sequential size sort

Notes: This table shows the Student's t-statistics for the sequentially sorted portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F sorted on beta	300	-0.31	-0.30	-0.29
L-W-Z sorted on beta	300	0.05	0.05	0.05
F-F sorted on size	300	0.69	0.70	0.61
L-W-Z sorted on size	300	-0.49	-0.49	-0.49
F-F sorted on momentum	300	-2.45	-2.21	-2.68
L-W-Z sorted on momentum	300	-0.53	-0.53	-0.53

Table 36: Student's t-test results based on the sequential value sort

Notes: This table shows the Student's t-statistics for the sequentially sorted portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the value sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

Data set	Z	t-statistic*	t-statistic**	t-statistic***
F-F sorted on beta	300	0.39	0.39	0.39
L-W-Z sorted on beta	300	-0.61	-0.61	-0.57
F-F sorted on size	300	3.26	3.59	3.06
L-W-Z sorted on size	300	-0.59	-0.60	-0.50
F-F sorted on value	300	-2.55	-2.54	-2.95
L-W-Z sorted on value	300	0.16	0.17	0.16

Table 37: Student's t-test results based on the sequential momentum sort

Notes: This table shows the Student's t-statistics for the sequentially sorted portfolios of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the momentum sort in a global minimum variance setting. t-statistic* is based on a HAC standard error. t-statistic** is based on a PW-HAC standard error. t-statistic*** is based on a sample standard error.

B.4 Sharpe Ratio differences in a global minimum variance profile

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F beta sorted vs. beta sorted benchmark (F-F)	300	0.92	0.93	0.13	0.13
L-W-Z beta sorted vs. beta sorted benchmark (L-W-Z)	300	0.31	0.31	-0.01	-0.02

Table 38: p-values Sharpe Ratio test based on the beta sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the beta sorted portfolio returns in a global minimum variance setting and the benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F size sorted vs. size sorted benchmark (F-F)	300	0.98	0.98	0.38	0.38
L-W-Z size sorted vs. size sorted benchmark (L-W-Z)	300	0.88	0.87	0.03	0.03

Table 39: p-values Sharpe Ratio test based on the size sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the size sorted portfolio returns in a global minimum variance setting and the benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F value sorted vs. value sorted benchmark (F-F)	300	0.42	0.47	-0.08	-0.11
L-W-Z value sorted vs. value sorted benchmark (L-W-Z)	300	0.92	0.92	0.02	0.01

Table 40: p-values Sharpe Ratio test based on the value sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the value sorted portfolio returns in a global minimum variance setting and the benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F momentum sorted vs. momentum sorted benchmark (F-F)	300	0.28	0.27	-0.08	-0.12
L-W-Z momentum sorted vs. momentum sorted benchmark (L-W-Z)	300	0.44	0.46	0.00	0.04

Table 41: p-values Sharpe Ratio test based on the momentum sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the momentum sorted portfolio returns in a global minimum variance setting and the benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F sorted on size vs. sorted on size benchmark (F-F)	300	0.25	0.22	0.42	0.45
L-W-Z sorted on size vs. sorted on size benchmark (L-W-Z)	300	0.72	0.72	-0.02	-0.01
F-F sorted on value vs. sorted on value benchmark (F-F)	300	0.84	0.83	0.03	0.03
L-W-Z sorted on value vs. sorted on value benchmark (L-W-Z)	300	0.38	0.37	-0.02	0.00
F-F sorted on momentum vs. sorted on momentum benchmark (F-F)	300	0.57	0.58	0.00	0.02
L-W-Z sorted on momentum vs. sorted on momentum benchmark (L-W-Z)	300	0.45	0.45	-0.02	0.00

Table 42: p-values Sharpe Ratio test based on the sequential beta sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the sequentially sorted portfolio returns in a global minimum variance setting and the sequentially sorted benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the beta sort. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F sorted on beta vs. sorted on beta benchmark (F-F)	300	0.00	0.00	0.15	0.29
L-W-Z sorted on beta vs. sorted on beta benchmark (L-W-Z)	300	0.56	0.58	0.01	-0.01
F-F sorted on value vs. sorted on value benchmark (F-F)	300	0.00	0.00	0.29	0.48
L-W-Z sorted on value vs. sorted on value benchmark (L-W-Z)	300	0.82	0.81	-0.02	0.00
F-F sorted on momentum vs. sorted on momentum benchmark (F-F)	300	0.55	0.55	0.27	0.29
L-W-Z sorted on momentum vs. sorted on momentum benchmark (L-W-Z)	300	0.40	0.43	0.00	0.03

Table 43: p-values Sharpe Ratio test based on the sequential size sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the sequentially sorted portfolio returns in a global minimum variance setting and the sequentially sorted benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F sorted on beta vs. sorted on beta benchmark (F-F)	300	0.13	0.12	-0.02	0.03
L-W-Z sorted on beta vs. sorted on beta benchmark (L-W-Z)	300	0.67	0.69	0.00	0.01
F-F sorted on size vs. sorted on size benchmark (F-F)	300	0.02	0.01	0.04	0.13
L-W-Z sorted on size vs. sorted on size benchmark (L-W-Z)	300	0.38	0.38	-0.03	0.02
F-F sorted on momentum vs. sorted on momentum benchmark (F-F)	300	0.95	0.95	-0.16	-0.15
L-W-Z sorted on momentum vs. sorted on momentum benchmark (L-W-Z)	300	0.35	0.36	-0.03	0.00

Table 44: p-values Sharpe Ratio test based on the sequential value sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the sequentially sorted portfolio returns in a global minimum variance setting and the sequentially sorted benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the value sort. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F sorted on beta vs. sorted on beta benchmark (F-F)	300	0.32	0.34	0.02	0.05
L-W-Z sorted on beta vs. sorted on beta benchmark (L-W-Z)	300	0.43	0.43	-0.03	-0.01
F-F sorted on size vs. sorted on size benchmark (F-F)	300	0.89	0.88	0.18	0.18
L-W-Z sorted on size vs. sorted on size benchmark (L-W-Z)	300	0.02	0.02	-0.03	0.06
F-F sorted on value vs. sorted on value benchmark (F-F)	300	0.29	0.28	-0.17	-0.21
L-W-Z sorted on value vs. sorted on value benchmark (L-W-Z)	300	0.90	0.90	0.01	0.00

Table 45: p-values Sharpe Ratio test based on the sequential momentum sort

Notes: This table shows the p-values of the Sharpe Ratio differences between the sequentially sorted portfolio returns in a global minimum variance setting and the sequentially sorted benchmark returns of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the momentum sort. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

B.5 Turnover ratio in a global minimum variance profile

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F beta sorted	299	4.87	50.72	47.47	94.80
L-W-Z beta sorted	299	4.44	38.40	44.81	91.08

Table 46: Portfolio turnover summary statistics based on the beta sort

Notes: This table shows the monthly beta sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F size sorted	299	16.99	66.03	62.80	94.60
L-W-Z size sorted	299	45.10	85.32	86.70	99.80

Table 47: Portfolio turnover summary statistics based on the size sort

Notes: This table shows the monthly size sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F value sorted	299	1.66	64.04	80.60	96.70
L-W-Z value sorted	299	5.58	42.52	47.14	100.00

Table 48: Portfolio turnover summary statistics based on the value sort

Notes: This table shows the monthly value sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F momentum sorted	299	20.41	74.93	87.69	97.50
L-W-Z momentum sorted	299	47.10	90.64	92.90	100.00

Table 49: Portfolio turnover summary statistics based on the momentum sort

Notes: This table shows the monthly momentum sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on size	299	16.83	45.35	43.91	76.80
L-W-Z sorted on size	299	36.60	62.49	63.40	85.30
F-F sorted on value	299	5.66	44.03	42.04	76.60
L-W-Z sorted on value	299	12.50	56.07	58.60	85.60
F-F sorted on momentum	299	22.22	65.05	67.46	89.50
L-W-Z sorted on momentum	299	26.25	70.01	71.10	94.40

Table 50: Portfolio turnover summary statistics based on the sequential beta sort

Notes: This table shows the monthly sequentially sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the beta sort in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	299	21.45	55.59	55.60	91.90
L-W-Z sorted on beta	299	29.20	60.61	62.15	92.20
F-F sorted on value	299	19.03	58.14	59.30	85.60
L-W-Z sorted on value	299	26.90	60.91	61.20	98.40
F-F sorted on momentum	299	44.60	81.60	82.49	99.80
L-W-Z sorted on momentum	299	32.70	75.50	77.00	99.60

Table 51: Portfolio turnover summary statistics based on the sequential size sort

Notes: This table shows the monthly sequentially sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the size sort in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	299	4.63	38.70	36.25	74.10
L-W-Z sorted on beta	299	8.75	57.56	60.70	86.50
F-F sorted on size	299	14.00	56.01	57.23	84.37
L-W-Z sorted on size	299	25.60	61.84	61.10	98.80
F-F sorted on momentum	299	18.52	59.56	64.20	93.50
L-W-Z sorted on momentum	299	37.70	71.10	72.50	99.60

Table 52: Portfolio turnover summary statistics based on the sequential value sort

Notes: This table shows the monthly sequentially sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the value sort in a global minimum variance setting.

Data set	Z	Minimum (in %)	Mean (in %)	Median (in %)	Maximum (in %)
F-F sorted on beta	299	26.85	67.45	69.70	89.30
L-W-Z sorted on beta	299	27.50	73.37	74.50	95.90
F-F sorted on size	299	32.40	69.16	69.40	93.90
L-W-Z sorted on size	299	40.00	76.64	77.80	99.80
F-F sorted on value	299	22.22	62.68	66.20	90.80
L-W-Z sorted on value	299	37.50	73.16	74.60	99.20

Table 53: Portfolio turnover summary statistics based on the sequential momentum sort

Notes: This table shows the monthly sequentially sorted portfolio turnover summary statistics of the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set, based on the momentum sort in a global minimum variance setting.

C Ledoit, Wolf and Zhao strategy

The applied strategy described in this paper deviates from the methodology used by Ledoit et al. (2019) on several points. The biggest deviations will shortly be described here, where the Ledoit et al. (2019) strategy will be referred to as the LWZ strategy in this section. The first step of both portfolio optimizing methodologies is to consider factors by sorting. However, the applied strategy performs a straightforward sort in order to already reduce dimensionality before optimizing the specific quantiles. The LWZ strategy does not account for dimensionality by sorting and uses a target exposure to some specific anomaly within its optimizer. Next to that, the applied strategy takes into account a univariate as well as a bivariate sorting, where the LWZ strategy only considers the single sort. Another deviation noteworthy is the fact that the applied strategy uses the global minimum variance profile in order to optimize the sorted quantiles, where the LWZ strategy shows that their methodology can easily be linked to the mean variance profile.

In order to make the comparison between the two strategies, the Sharpe Ratio difference backtest is computed for which the results are presented in the table below. From the eight scenarios the applied strategy turns out to have a higher Sharpe Ratio in half of the cases and a break even occurs in one of the eight cases. In terms of statistical significance there are two events where the difference between the Sharpe Ratios are substantial. Within the F-F data set the applied strategy strongly outperforms when the size anomaly is implied, where even the threshold of a p-value of 0.01 could be used. If a less strict threshold is adopted of 0.05, the odds are turned with respect to the momentum factor, where the LWZ strategy significantly differs from the applied strategy in the F-F data set. Within all the other presented cases there is no significant deviation between the Sharpe Ratios. Hence, the conclusion to which strategy should be used at all times is inconclusive, where further research is needed to come up with a more strict answer.

Data set	Z	p-value*	p-value**	Sharpe Ratio 1	Sharpe Ratio 2
F-F beta sorted applied strategy vs. F-F beta sorted LWZ strategy	300	0.61	0.62	0.02	-0.01
L-W-Z beta sorted applied strategy vs. L-W-Z beta sorted LWZ strategy	300	0.53	0.55	0.00	-0.03
F-F size sorted applied strategy vs. F-F size sorted LWZ strategy	300	0.00	0.00	0.25	-0.01
L-W-Z size sorted applied strategy vs. L-W-Z size sorted LWZ strategy	300	0.48	0.47	-0.03	0.01
F-F value sorted applied strategy vs. F-F value sorted LWZ strategy	300	0.73	0.73	0.04	0.07
L-W-Z value sorted applied strategy vs. L-W-Z value sorted LWZ strategy	300	0.92	0.91	0.02	0.01
F-F momentum sorted applied strategy vs. F-F momentum sorted LWZ strategy	300	0.03	0.03	-0.05	0.08
L-W-Z momentum sorted applied strategy vs. L-W-Z momentum sorted LWZ strategy	300	0.96	0.96	0.02	0.02

Table 54: p-values Sharpe Ratio test

Notes: This table shows the p-values of the Sharpe Ratio differences between the applied strategy and the LWZ strategy within the two different data sets, referred to as the French and Fama (F-F) data set and the Ledoit, Wolf and Zhao (L-W-Z) data set. p-value* is based on a HAC standard error. p-value** is based on a PW-HAC standard error.

D Data

In order to perform the portfolio optimization techniques, the obtained data is cleansed and split into two data sets. The first data set is based upon the assumptions taken by Ledoit et al. (2019). The second data set has the same requirements as Fama and French (1993). The exact filters used in this research paper are listed below.

D.1 Cleaning raw data

First of all the values of all the different variables observed in the raw data are mutated to its proper value characteristic. For example, the returns (RET) should be interpreted as double values. Then, in case of a missing closing price (PRC), the PRC data was set to the negative of the average of the bid and ask price. In order to deal with this problem, the absolute values of the PRC variable are taken. Furthermore, some cases showed a lack of information, for which the RET data was set to -66, -77, -88, or -99, where each value has its specification of what exact information is missing. These observations are omitted from the data, since they have no further use. Observations with PRC equal to zero are removed as well, since the PRC information is missing, which is of relevance for the continuation of the program. Furthermore, observations with NAs are omitted as well.

D.2 Ledoit, Wolf and Zhao data set

After cleaning the data, the data set is split into two data sets. Here, the Ledoit, Wolf and Zhao (L-W-Z) data set is discussed. First of all, observations for which there is a lack of information within the past five years, or the succeeding month, are removed. Finally, at each portfolio construction day the one thousand biggest companies, depicted by market capitalisation, are chosen as the available stock universe.

D.3 Fama and French data set

For the Fama and French (F-F) data set, financial companies with a standard industrial classification (SIC) code between 6000 and 6999 are excluded. Then, 5000 assets are depicted randomly from the whole data set, in order to reduce some dimensionality. Finally, the F-F stock universe is produced by excluding assets with no full track record of the previous five years, or successive month, at each portfolio construction day.

E R code

The results discussed in Section 5 followed from the corresponding R programming code. Since the script is very elaborate it comes as an additional file. The code can be summarised into 23 chapters, where for each chapter a short description is provided here. The index of the chapter in this description corresponds with the chapter index within the programming code.

Chapter 1: Packages

The packages that are necessary in order to run the full extent of the programming code are listed and provided in this chapter.

Chapter 2: Functions

In order to reduce space and let R run in a more efficient way, some functions were defined.

Chapter 3: Reading in data

The code provided in chapter three provides the code to read in the corresponding data sets.

Chapter 4: Cleaning data

This chapter provides the code for the used filters and assumptions on the data sets described in Appendix D.

Chapter 5: Constructing the main data set

Since the raw data set was very high dimensional, chapter three and four were divided into parts in order to let R run in a more efficient way. Chapter five combines the separate data sets over time into one main data set.

Chapter 6: Unique companies

Chapter 6 is not a necessary chapter, however, it provides interesting insights into the the amount of stocks in the investment universe after filtering and cleaning the data.

Chapter 7: Creating Fama and French data set

In this chapter the code is provided in order to come up with the available investment universe for the Fama and French data set based on the assumptions that are described in Section 4.

Chapter 8: Creating Ledoit, Wolf and Zhao data set

In this chapter the code is provided in order to come up with the available investment universe for the Ledoit, Wolf and Zhao data set based on the assumptions that are described in Section 4.

Chapter 9, 10, 11 12: Repetition

These chapters read in, clean and reform the main data set once more since the previous chapters reduced the amount of variables and observations in order to reduce the required memory space ending up with the exact investment universe. However, in order to run the following chapters more variables are required, for which the data should be read in again to be filtered on the exact investment universe.

Chapter 13: Factor construction

In Chapter 13 the computations are done in order to come up with the factor signals.

Chapter 14: Single sort

In Chapter 14 the extreme quintiles of the single sort are constructed based on the factor signals.

Chapter 15: Factor construction for the sequential sort

Chapter 15 computes the factor signals within the extreme quantiles of the single sort.

Chapter 16: Conditional sort

The code is provided here to construct the extreme quintiles based on the factor signals of the conditional sort.

Chapter 17: Benchmark returns

This chapter calculates the benchmark returns.

Chapter 18: Optimizing portfolios

This chapter optimizes the single as well as the conditional portfolios by computing the DCC-NL covariance estimators within the extreme quintiles at each portfolio construction day. After the DCC-NL covariance matrix is calculated it can be directly inserted in the specific optimizers in order to come up with the weights of the stocks per portfolio.

Chapter 19: Portfolio returns

Once the weights are calculated the portfolio returns can be computed, for whicht the applied code is given in Chapter 19.

Chapter 20: Backtesting

The programming code for the backtest with respect to the Sharpe Ratio difference and the t-statistics is provided in Chapter 20.

Chapter 21: Turnover ratio

The turnover ratio for the different strategies are computed within this chapter.

Chapter 22: Visualizations

Chapter 22 is coded in order to provide some insights by producing graphics.

Chapter 23: Ledoit Wolf and Zhao strategy

Within this chapter the exact same portfolio methodology of Ledoit et al. (2019) is applied to a subset of the data sets used within this paper and backtested against the methodology applied within this research paper.