ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

MSc Econometrics and Management Science Quantitative Finance Thesis

Parametric Portfolio Policies: Is it still worthwhile to use firm characteristics for large-N portfolio allocation in the presence of transaction costs?

Academic supervisor:
Dr. Rasmus Lönn

Author:
Karan Samlal - (458921)

Second reader:
Dr. Onno Kleen

November 14, 2022

Abstract

This paper examines minimum variance parametric portfolios to determine to what extent investors still economically benefit from exploiting firm characteristics for large-N portfolio allocation when transaction costs are present.
The research finds considerable benefits for investors from exploiting the firm characteristics by using minimum variance parametric portfolios when imposing portfolio weight constraints such as turnover and short-sale constraints. The empirical results show strong out-of-sample performance for the constrained minimum variance parametric portfolios relative to the naive equally-weighted $1/N$ benchmark strategy and global minimum variance portfolios with estimation techniques to reduce the impact of the estimation error in the covariance estimate. Furthermore, the results indicate that the strongest out-of-sample performance from minimum variance parametric portfolios is when imposing a short-sale constraint and exploiting a smaller set of five characteristics highlighted in Fama and French (2015).
The empirical findings are consistent post-2003 when researchers document that characteristics-based predictability fell sharply and point out that when variance minimization is the objective, investors can still realize economic gains relative to naive diversification and traditional econometric approaches by exploiting the firm characteristics.

Keywords – *Minimum-variance portfolio; Estimation risk; Anomalies; Transaction costs; Portfolio weight constraints*

Contents

1 Introduction

The traditional mean-variance approach of Markowitz (1952) requires modeling the expected returns, variances, and covariances of all stocks.Minimum variance allocation is especially important for portfolio optimization since the optimal mean-variance allocation is generally found to be empirically poor out-of-sample. The notorious difficulty of forecasting asset returns consistently and reliably makes it challenging to get a reliable estimate of the mean vector. Jobson and Korkie (1980), and Chopra and Ziemba (1993) confirm this and show that the largest estimation error is in the mean estimate. At the same time, variance-covariance estimates are relatively more stable over time and are therefore more reliable than expected returns (Jorion, 1985; Merton, 1980).

There are now many explanations for this including low-risk anomalies, factor exposures, compositional characteristics and behavioural explanations (Ang, Hodrick, Xing, & Zhang, 2006; Baker, Bradley, & Wurgler, 2011; Scherer, 2011). Despite its appeal and advantages, minimum variance allocation still poses significant empirical challenges since it relies on the inverse covariance matrix. Thus, in portfolios of many assets, these allocations become burdensome to estimate since the number of unknown parameters grows quadratically in N. In addition, the matrix inversion may compound the measurement errors, which can be substantial due to the noisy returns. For this reason, it is relevant to consider alternative estimation methods that can circumvent these issues.

The literature finds that the estimation of the covariance matrix can be performed using various techniques developed to reduce the impact of estimation error; for instance, by imposing portfolio weight constraints (Chopra, 1993; Frost & Savarino, 1988; Jagannathan & Ma, 2003), imposing a factor structure on the estimator of the covariance matrix (Chan, Karceski, & Lakonishok, 1999), shrinking the covariance matrix towards a target matrix (Ledoit & Wolf, 2004b, 2012), and sparse (inverse) covariance estimation (Bickel & Levina, 2008; Cai & Liu, 2011; Friedman, Hastie, & Tibshirani, 2008). In contrast, DeMiguel, Garlappi, and Uppal (2009) demonstrate that an investor may be better off by ignoring stock return data and using the naive $1/N$ rule to allocate an equal proportion of wealth across each of the N assets. According to DeMiguel et al. (2009), "there are still many miles to go before the gains promised by optimal portfolio choice can be realized out of sample."

The literature finds that stock characteristics are related to the stock's expected return, variance, and covariance with other stocks, at least to some extent. Hundreds of variables have been proposed to explain the cross-section of stock returns (Harvey, Liu, & Zhu, 2016; Hou, Xue, & Zhang, 2015; McLean & Pontiff, 2016). Brandt, Santa-Clara, and Valkanov (2009) propose an alternative approach to optimizing portfolios with a large number of assets by introducing: the parametric portfolio policy. Instead of modeling the moments of returns, the parametric portfolio policy directly models the portfolio weights as a function of the firm characteristics. The coefficients of this function are found by optimizing the investor's average utility of the portfolio's return over the sample period. This leads to a substantial reduction in dimensionality. The traditional Markowitz approach requires modeling N first and $(N^2+N)/2$ second moments of returns for a problem with N stocks. In contrast, the parametric portfolio requires modeling only N portfolio weights. In addition, it captures implicitly the relation between the characteristics and expected returns, variances, covariances, and even higher order moments of returns, since they affect the distribution of the optimized portfolio's returns and, therefore, the investor's expected utility (Brandt et al., 2009).

However, transaction costs play a crucial role in portfolio choice. If the portfolio weights change too much over time and thus require frequent rebalancing, the investor faces unnecessarily high costs. Another critical thing to consider is portfolio weight constraints, such as turnover and short-sale constraints, which many funds face. These real-world constraints are not only relevant because of legal and other regulatory circumstances. Jagannathan and Ma (2003), for example, show that self-imposed short-sale constraints can significantly improve the out-of-sample performance of portfolios. Finally, Chordia, Subrahmanyam, and Tong (2014), and Green, Hand, and Zhang (2017) show that the economic and statistical significance of firm characteristics decreased significantly in the past decade.
This leads to the question:

"Can investors still economically benefit from using firm characteristics for large-N portfolio allocation in the presence of transaction costs?"

This paper answers this question by evaluating the model performance of several minimum variance parametric portfolios. To achieve the main research goal, four research questions are being answered:

1. Can parametric portfolios provide significant benefits over naive diversification and global minimum variance portfolios for investors that not only care about average returns but also about portfolio risk, transaction costs, and the stability of the portfolio weights?

2. Can an investor improve out-of-sample performance net of transaction costs by exploiting a large set of characteristics?

3. How do the out-of-sample results change when imposing portfolio weight constraints?

4. Does the decrease in characteristics-based predictability of the past decade have any implications on the out-of-sample results?

To answer the research questions, four parametric portfolios are studied: two portfolios that exploit a small number of characteristics typically considered in popular asset-pricing models and two portfolios that exploit a large set of characteristics. The first parametric portfolio exploits the characteristics highlighted in Fama and French (1993): the beta, size, and book-to-market ratio. The second parametric portfolio exploits the characteristics highlighted in Fama and French (2015): the beta, size, book-to-market ratio, operating profitability, and asset growth. The third parametric portfolio is the regularized parametric portfolio that exploits a large set of 39 characteristics with Lasso regularization. For the fourth parametric portfolio, I propose a new approach to estimate parametric portfolios when dealing with a large set of characteristics by introducing: K-threshold parametric portfolios. This method pre-selects the 39 characteristics based on in-sample scoring, then cross-validates the number of characteristics to include and finally estimates the standard parametric portfolio using these characteristics. Furthermore, the research considers global minimum variance portfolios with estimation techniques to reduce the impact of estimation error in the covariance estimate, such as by imposing a factor-covariance structure, (non) linear shrinkage, Adaptive Thresholding, and Factor Graphical Lasso.

The out-of-sample performance is compared, and the statistical significance is tested relative to the performance of the equally-weighted $1/N$ benchmark strategy using the following three performance criteria: (i) the out-of-sample portfolio volatility, (ii) Sharpe ratio, and (iii) the turnover (trading volume) for each portfolio strategy. In addition, the stability of the portfolio weights is examined by evaluating the extremeness of (individual) positions, the amount of leverage a portfolio strategy requires, and the fraction of total short positions. Further, this paper studies the effect of imposing portfolio weight constraints on the optimized portfolios by considering turnover and short-sale constraints. Finally, the out-of-sample results are examined for two equally sized sub-samples: January 1989 to December 2002 and January 2003 to December 2016, as Chordia et al. (2014), and Green et al. (2017) find that characteristics-based predictability fell sharply post-2003.

The primary contributions are two-fold. First, I propose a new method to optimize parametric portfolios when dealing with a large set of characteristics by introducing the K-threshold parametric portfolio. Although this paper finds no significant benefits to investors from exploiting a large set of 39 characteristics using the K-threshold parametric portfolio, this paper finds better out-of-sample performance than the regularized parametric portfolio and comparable performance to the parametric portfolios that exploit a small number of characteristics.

Second, and more importantly, I show that using firm characteristics for large-N portfolio allocation is still economically meaningful when transaction costs are present by demonstrating economic gains that can be achieved from constrained minimum variance parametric portfolios relative to global minimum variance portfolios and naive diversification. Furthermore, the research highlights that when optimizing a minimum variance allocation for the parametric portfolios, these gains tend to be relatively stable and persistent over time.

The main empirical finding of this paper is that there are considerable benefits for investors from exploiting the firm characteristics by using minimum variance parametric portfolios when imposing portfolio weight constraints. First, the empirical results show a substantial reduction in the portfolio volatility for the parametric and global minimum variance portfolios relative to the $1/N$ benchmark strategy, with the differences being statistically significant. Further, the results show that the Sharpe ratios increase and are generally higher for the parametric portfolios relative to the global minimum variance portfolios and naive diversification when imposing a turnover or short-sale constraint. This paper finds the strongest out-of-sample performance when imposing a short-sale constraint and exploiting the five characteristics highlighted in Fama and French (2015). Although the differences in the Sharpe ratio are found statistically insignificant, there are some considerable economic gains from constrained parametric portfolios. For example, when imposing a short-sale constraint, the results indicate a gain in the annualized out-of-sample Sharpe ratio of 30% relative to the $1/N$ benchmark strategy for the parametric portfolio that exploits five characteristics and 21.67% for the parametric portfolio that exploits three characteristics and the K-threshold parametric portfolio. In addition, the turnover ranges from just 7.30% to 8.20%, as opposed to the turnover for the $1/N$ benchmark strategy of 6.19%, suggesting a limited amount of additional transaction costs. Moreover, the parametric portfolios did not involve extreme weights or require high amounts of leverage, including the constrained versions.

In contrast to the findings in DeMiguel, Martin-Utrera, Nogales, and Uppal (2020), using a large set of characteristics did improve out-of-sample performance net of transaction costs. This paper finds that the regularized parametric portfolio using 39 characteristics with Lasso regularization does not perform well out-of-sample. However, the K-threshold parametric portfolio, which I propose, did achieve similar performance as the parametric portfolios that exploit only a small number of characteristics.

The empirical findings are consistent post-2003. The results indicate stronger out-of-sample performance from 2003 to 2016 for the short-sale constrained parametric portfolios and show that economic gains can still be realized by exploiting firm characteristics when focusing on minimizing the portfolio variance and omitting expected stock returns.

This thesis proceeds as follows. First, section 2 presents an overview of relevant literature on this research topic. Next, Section 3 gives a data description and discusses and contrasts some of the main characteristics used in the research. Furthermore, Section 4 describes the methods and set-up that are used in this research. Finally, Sections 5, 6, and 7 present the empirical results, while Sections 8 and 9 include the conclusion, discussion, and further research.

2 Literature

When the matrix dimension is large compared to the sample size, the sample covariance matrix is known to perform poorly and may suffer from ill-conditioning. In recent years researchers have proposed various techniques to deal with the problem of estimating the large number of elements in the covariance matrix. One technique is to impose portfolio weight constraints (Chopra, 1993; Frost & Savarino, 1988). Jagannathan and Ma (2003) show that imposing short-sale constraints on the global minimum variance portfolio is equivalent to shrinking the large elements of the covariance matrix. A second technique is to consider factor models to impose a covariance-factor structure. For example, Chan et al. (1999) propose several factor models, which reduce the number of parameters to be estimated and, therefore, mitigate the impact of estimation error. A third technique has been proposed by Ledoit and Wolf (2004b), who shrinks the sample covariance matrix toward a target matrix, also referred to as Linear shrinkage. Ledoit and Wolf (2012) extends the shrinkage estimation of Ledoit and Wolf (2004b) by introducing: Nonlinear shrinkage, which applies an individualized shrinkage intensity to every sample eigenvalue. A fourth technique is to incorporate additional knowledge in the estimation process, such as sparseness or a graph model. Bickel and Levina (2008) propose a thresholding method, which sets small elements to zero in the estimated covariance matrix. Cai and Liu (2011) introduce Adaptive Thresholding, which is an adaptive variant of the thresholding procedure of Bickel and Levina (2008). Finally, Friedman et al. (2008) propose a method for estimating sparse precision matrices by introducing: Graphical Lasso, which shrinks the elements of the inverse covariance matrix towards zero compared to the maximum likelihood estimates using l_1 regularization.

In contrast, DeMiguel et al. (2009) examine several Markowitz optimal portfolios and the various extensions designed to reduce the impact of estimation error. The authors find that in terms of the Sharpe ratio, certainty-equivalent return, and turnover, none of the optimal portfolios can consistently outperform the naive $1/N$ strategy in which each available asset is given an equal weight in the portfolio.

Brandt et al. (2009) propose a different approach to optimizing portfolios with a large number of assets and introduce the parametric portfolio policy. The parametric portfolio policy directly models the portfolio weights as a function of firm characteristics. The coefficients of this function are found by optimizing the investor's objective function, such as the CRRA utility as respectively used by Brandt et al. (2009) over the estimation period. Brandt et al. (2009) find that the size, book-to-market ratio, and one-year momentum could explain deviations of the optimal portfolio for a CRRA investor from the market portfolio. With a relative risk aversion of five, the parametric portfolio achieved an annualized certainty-equivalent gain of 11% relative to holding the market portfolio.

DeMiguel et al. (2020) extend the parametric portfolio framework from Brandt et al. (2009) to examine how transaction costs change the number of characteristics that are jointly significant for an investor's optimal portfolio. The authors introduce the *regularized parametric portfolio*, which is obtained by imposing a Lasso constraint on the parametric portfolio. The authors find that in the presence of transaction costs, the cross-section of stock returns is not fully explained by a small number of characteristics. The regularized parametric portfolio that exploits a large set of 51 characteristics, achieved in the presence of transaction costs an out-of-sample Sharpe ratio that is 139% higher than the market portfolio, 100% higher than the parametric portfolio that exploits the three characteristics considered in Brandt et al. (2009), and 25% higher than the parametric portfolio that exploits four investment and profitability characteristics, highlighted in Hou et al. (2015), and Fama and French (2015).

Chordia et al. (2014) examine the profitability of anomaly portfolio sorts from 12 characteristics. They find that due to increased liquidity and trading activity, the predictive power of most anomalies has weakened and that the average returns and the Sharpe ratio from a portfolio strategy based on the characteristics declined significantly in the past decade. McLean and Pontiff (2016) studied 97 firm characteristics to examine how academic publications and in-sample bias affect the significance of the respective characteristics. They find that portfolio returns are 26% lower out-of-sample and 58% lower post-publication and conclude that investors learn about mispricing from academic publications. Finally, Green et al. (2017) studied 94 firm characteristics to identify which can provide independent information about average US monthly stock returns. They find that while 12 characteristics are significant during 1980-2014, return predictability fell sharply after 2003, and just two characteristics have been independent determinants since then. They further show that the long-short portfolio returns from exploiting the characteristics have been insignificantly different from zero since 2003.

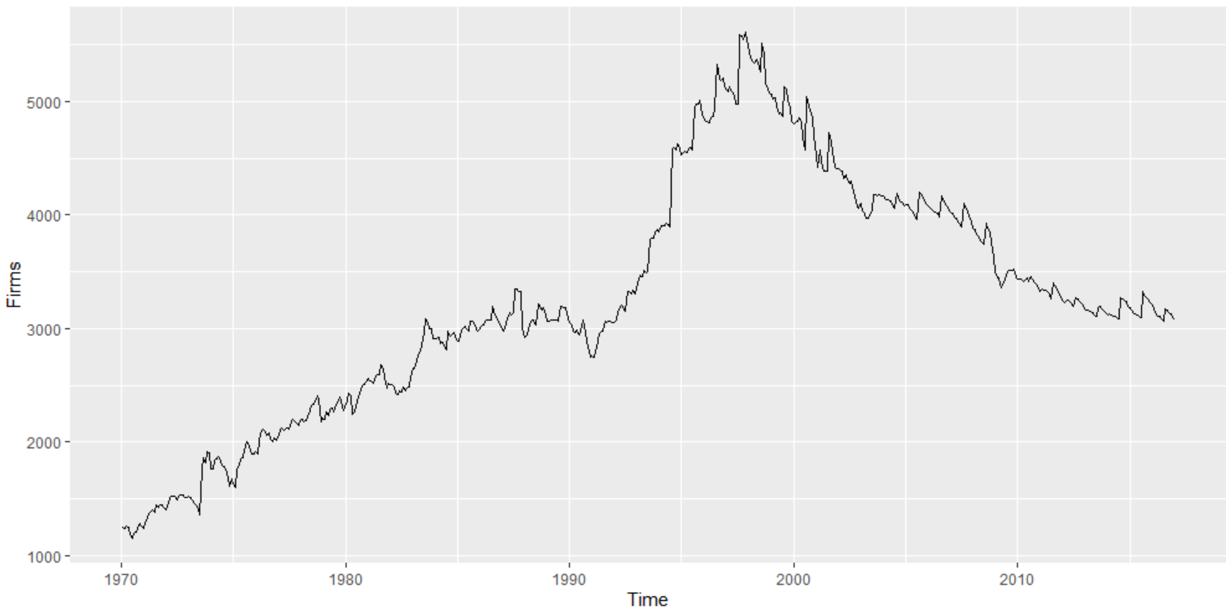
3 Data

This section discusses the data and preparation used in the empirical analysis. Monthly firm-level returns are obtained from CRSP for all listed firms on the NYSE, AMEX, and NASDAQ, from January 1970 through December 2016. The risk-free rate is obtained from the website of Kenneth R. French. In addition, the three factors in Fama and French (1993) are obtained, which will impose a factor structure for the global minimum variance portfolios. Finally, the firm characteristic data is obtained from the website of Shihao Gu, where the authors in Gu, Kelly, and Xiu (2020) published data of 94 lagged firm characteristics considered in Green et al. (2017).

Following Brandt et al. (2009), firms with negative book-to-market ratios are removed. In addition, firms below the 20th percentile of market capitalization are also removed since these are very small firms that are difficult to trade. Figure 1 shows the resulting number of firms over time, which trended upwards until the dot-com bubble burst, which took place in 2000. The average number of firms throughout the sample is 3,582, with the fewest firms in June of 1970 (1,155 firms) and the most firms in October of 1997 (5,610 firms).

As in DeMiguel et al. (2020), characteristics with a large proportion of missing observations are dropped. Specifically, characteristics are dropped with more than 10% of missing observations for more than 10% of firms with CRSP returns available for the entire sample from 1970 to 2016. In addition, characteristics are dropped without any observations for more than 1% of these firms. Table 1 in the Appendix Section A lists the resulting 39 characteristics together with the author's name (s), the date and journal of publication, the data source, and the frequency. Table 2 in the Appendix Section B shows a summary of statistics: mean, standard deviation, minimum, maximum, skewness, and kurtosis of the time-series averages of the (non-standardized) characteristics.

Figure 1: Plot of the number of firms over time



3.1 Discussion of the characteristics

The final dataset contains 39 characteristics, and this section briefly discusses and contrasts some of the main characteristics. Starting with Fama and MacBeth (1973), who study the relationship between expected stock returns and market risk (beta). The authors find a positive relationship between expected returns and beta. This supports the Capital Asset Pricing Model (CAPM) in that the market beta fully explains the cross-section of stock returns. However, numerous studies show that this generally does not hold anymore. For instance, Frazzini and Pedersen (2014) study a factor that goes long in low beta stocks and short in high beta stocks, also referred to as the betting against beta (BAB) factor. In contrast to the results in Fama and MacBeth (1973), the authors find that the BAB factor produces significant positive risk-adjusted returns and conclude that high risk is associated with low returns. Moreover, Litzenberger and Ramaswamy (1982) study the relationship between a stock's dividend yield and its expected return. They find that stocks with a high-dividend yield provide a higher return than stocks with a low-dividend yield. Bhandari (1988) examine if firm leverage can help to explain the cross-section of stock returns after both the beta and size are taken into account. They proxy the firm leverage by the debt-to-equity ratio. He finds that firms with higher debt-to-equity ratios tend to generate higher expected returns than firms with low debt-to-equity ratios and conclude that leverage also helps explain the cross-section of stock returns after both beta and size are included. Amihud and Mendelson (1989) study the impact of a stock's liquidity on its expected return, where they use the wideness of the bid-ask spread as a proxy for the liquidity. The authors find that illiquid stocks tend to generate higher expected returns than liquid stocks and conclude that investors generally price in the liquidity of a stock. Fama and French (1993) expand the CAPM by adding a size and value factor. In their research, they conclude that small firms with a high book-to-market ratio (small-cap value firms) tend to generate higher expected returns than large firms with a low book-to-market ratio (large-cap growth firms). Further, Jegadeesh and Titman (1993) earn a large paper profit by studying the effect between a stock's past performance on its future returns and uncover the well-known momentum effect. The momentum effect suggests that stocks that historically did well over a three to twelve-month period continue to do well for the next few months, and stocks that did poorly continue to do poorly. Ang et al. (2006) study the relationship between idiosyncratic volatility and expected returns. The authors find that stocks with low idiosyncratic volatility tend to have higher expected returns, and stocks with high idiosyncratic volatility have lower expected returns. Cooper, Gulen, and Schill (2008) study the relationship between a firm's asset growth and its expected return. They find that stocks with low-asset growth significantly outperform stocks with high-asset growth. Finally, Fama and French (2015) expand their original three-factor model by adding an investment and profitability factor. They conclude that stocks with high operating profitability and low asset growth have higher expected returns. They further provide evidence that the five-factor model can better explain the cross-section of stock returns than the three-factor model they earlier proposed.

Regardless, researchers document that with the improved liquidity in the stock market in recent years and increased arbitrage activity, the predictive power of many of the characteristics might have weakened (Chordia et al., 2014; Green et al., 2017). As an illustration, the size and value factor imply that small-cap value stocks should provide a higher expected return than large-cap growth stocks. However, large-cap growth has outperformed small-cap value in the past decade. Likewise, there are many more anomalies whose predictive power faded, as Chordia et al. (2014) point out. This naturally raises the question of to what extent these anomalies can still benefit investors for portfolio allocation, which is a motivation for this study.

4 Methodology

This section describes the collection of portfolio optimization methods and various estimation techniques. First, the global minimum variance portfolio is introduced. Second, the focus is on advanced methods to reduce the impact of estimation error in the covariance matrix estimate. Third, the parametric portfolio policy is introduced along with its extensions. Fourth, the hyper-parameter tuning via cross-validation is described. At last, the estimation of transaction costs is specified.

4.1 Global Minimum Variance Portfolio

The global minimum variance (GMV) portfolio introduced by Markowitz (1952) is the efficient portfolio with the lowest variance based on the mean-variance portfolio efficiency framework. Given N assets, the optimization problem is formulated as follows:

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

where w is the $N \times 1$ vector of portfolio weights, Σ is the $N \times N$ covariance matrix, and ι is a vector of ones. Solving for the optimal weights, the following solution is obtained:

$$w^* = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}. \tag{2}$$

Since the GMV portfolio does not require an estimate of the mean vector, the estimation of the GMV portfolio can be seen as a fundamental problem in evaluating the quality of the covariance matrix estimate. Moreover, researchers document that estimated GMV portfolios have desirable out-of-sample properties in terms of risk and reward-to-risk (Jagannathan & Ma, 2003).

4.2 Factor models

The main difficulty with estimating the covariance matrix in high dimensions is the accumulation of many estimation errors due to the large number of parameters that must be estimated. When the concentration ratio becomes large, the sample covariance matrix can lead to unsatisfactory out-of-sample portfolio performance (Kan & Zhou, 2007; Ledoit & Wolf, 2004a). In addition, when the concentration ratio exceeds one, it is well-known that the sample covariance matrix is non-invertible. One way to tackle this issue is by making use of factor models.
A factor model assumes that for every asset $i = 1, ..., N$

$$r_{i,t} = \alpha_i + \beta_i' f_t + \epsilon_{i,t} \tag{3}$$

where $r_{i,t}$ is the return of asset i at time t, $f_t = (f_{t1}, ..., f_{tZ})'$ is a vector of Z observed factors, $\beta_i = (\beta_{i1}, ..., \beta_{iZ})'$ is a vector of time-invariant factor loadings, and $\epsilon_{i,t}$ an error term that satisfies $E(\epsilon_{i,t}|f_t) = 0$. The covariance matrix of the returns is given by:

$$\Sigma = B\Sigma_f B' + \Sigma_\epsilon \tag{4}$$

where B is the $Z \times N$ matrix of factor loadings, Σ_f the $Z \times Z$ covariance matrix of the factor returns, and Σ_ϵ the $N \times N$ covariance matrix of the error terms from the factor representation in Equation (3). Chan et al. (1999) propose to assume an exact factor model (EFM), which corresponds to assuming that the error terms are uncorrelated across assets. This results in that Σ_ϵ becomes a diagonal matrix with residual variances on its diagonal. The idea behind this approach is that factor models may capture a significant part of the covariation among assets and effectively reduce the number of free parameters to estimate.

4.3 Linear shrinkage

Ledoit and Wolf (2003) point out that there is no consensus on the identity of the factors nor on the number of factors to include, and an exact factor model can suffer from severe biases. As an alternative approach, Ledoit and Wolf (2004b) propose to combine a structured estimator with high bias, such as the identity matrix, and an unstructured estimator with high variance, such as the sample covariance matrix. This is generally known as linear shrinkage. The intuition behind linear shrinkage is that it aims to find the optimal bias-variance trade-off between these two estimators. The linear shrinkage estimator takes a weighted average of the identity matrix and sample covariance matrix:

$$\widehat{\Sigma} = \nu I + (1 - \nu)S \tag{5}$$

where I is the identity matrix, S the sample covariance matrix, and $\nu \in [0, 1]$ the shrinkage intensity. The higher the value of ν, the more the sample covariance matrix is shrunk towards the identity matrix, and the more structure is imposed. As a trade-off, more information contained by the sample covariance matrix is lost in the process. This leads to the fundamental problem of how determining the optimal shrinkage intensity. In this study, we find the optimal shrinkage intensity via five-fold cross-validation to minimize the portfolio variance described in section 4.8. Different shrinkage intensities are considered, with values between 0.001 and 0.99 divided into ten equal steps[1].

[1]Although a shrinkage intensity of $\nu = 0.001$ is very low, the shrinkage estimator is still found to be semi-positive definite and results in a non-border solution.

4.4 Nonlinear shrinkage

Unlike linear shrinkage, in which all variance and covariance parameters shrink towards the corresponding target with the same amount, the nonlinear shrinkage method introduced by Ledoit and Wolf (2012) shrinks the eigenvalues of the covariance matrix by applying an individualized shrinkage intensity to every sample eigenvalue. Ledoit and Wolf (2017) conduct simulation studies and find that in terms of finite-sample performance, the nonlinear shrinkage estimator often performs better than the linear shrinkage estimator.

Analytical nonlinear shrinkage

This study considers the analytical nonlinear shrinkage method provided by Ledoit and Wolf (2020). In comparison to other nonlinear shrinkage methods, it has the advantage that it is fast and scalable. The analytical nonlinear shrinkage estimator is summarized in the following steps. First, the spectral decomposition of the sample covariance matrix is computed:

$$S = \sum_{i=1}^{N} \lambda_i u_i u_i' \tag{6}$$

where λ_i denotes an eigenvalue of the sample covariance matrix, sorted in nondecreasing order, and where u_i is the corresponding eigenvector. Second, by Ledoit and Wolf (2020), the following bandwidth per eigenvalue to compute the eigenvalue density is chosen:

$$h_i = \lambda_i T^{-1/3}$$

where T denotes the number of time-series observations in the estimation sample. Next, the asymptotically optimal shrunk eigenvalues $\tilde{d}_i$ are calculated, and the covariance matrix estimator is composed as follows:

$$\tilde{d}_i := \frac{\lambda_i}{[\pi \frac{N}{T} \lambda_i \tilde{f}(\lambda_i)]^2 + [1 - \frac{N}{T} - \pi \frac{N}{T} \lambda_i H_{\tilde{f}}(\lambda_i)]^2} \tag{7}$$

where $\tilde{f}(.)$ denotes the Epanechnikov kernel, and $H_{\tilde{f}}(.)$ the Hilbert transform, defined in the Appendix, Section C.

In the last step, the covariance matrix estimator is recomposed:

$$\widehat{\Sigma} = \sum_{i=1}^{N} \tilde{d}_i u_i u_i'. \tag{8}$$

To summarise, non-linear shrinkage uses different shrinkage intensities for all elements of the covariance matrix, which could lead to improvements like Ledoit and Wolf (2017) indicate.

4.5 Adaptive Thresholding

Bickel and Levina (2008) argue that shrinkage estimators shrink the overdispersed sample covariance eigenvalues, but they do not change the eigenvectors and do not result in sparse estimators. A sparse covariance matrix implies that only a small number of returns have non-zero covariances. As an alternative to the shrinkage estimator, Bickel and Levina (2008) propose a threshold method that directly sets the smaller elements in the covariance matrix to zero. Cai and Liu (2011) propose an adaptive variant of the threshold method by directly applying thresholding to the correlation matrix. Denote the sample correlation matrix by $\widehat{C} = (\hat{c}_{ij})_{1 \leq i,j \leq N}$, with $\hat{c}_{i,j} = \hat{\sigma}_{ij}/\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}$. By Cai and Liu (2011), the universal thresholding estimator of the correlation matrix is defined as: $\widehat{C}(\gamma) = (\hat{c}_{ij}^{thr})_{N \times N}$, with

$$\hat{c}_{ij}^{thr} = \hat{c}_{ij} 1[|\hat{c}_{ij}| \geq \gamma] \tag{9}$$

where 1[.] denotes an indicator function, and $\gamma \in [0, 1]$ denotes the threshold parameter that sets the amount of sparsity in the correlation matrix. The estimate of the covariance matrix is given by:

$$\widehat{\Sigma} = D^{0.5} \widehat{C}(\gamma) D^{0.5} \tag{10}$$

where $D = \text{diag}(S)$. The intuition behind the threshold parameter γ is that when $\gamma = 0$, the estimator becomes the sample covariance matrix as no thresholding is employed, and when $\gamma = 1$, the estimator becomes a diagonal matrix with marginal sample variances on its diagonal. Here, the optimal threshold parameter is found via five-fold cross-validation to minimize the portfolio variance described in section 4.8. Different thresholding values are considered, with values between 0.75 and 0.99, divided into ten equal steps.

4.6 Factor Graphical Lasso

The motivation for the sparse covariance estimators can likewise be applied to the precision matrix. Graphical Lasso, introduced by Friedman et al. (2008), is one of the most famous methods for estimating sparse precision matrices and uses l_1 regularization to promote sparsity in the precision matrix. The objective function for Graphical Lasso is given by:

$$\widehat{\Theta} = \text{argmin}_{\Theta \succeq 0} \left(\text{tr}(S\Theta) - \log(\det(\Theta)) + \rho \|\Theta - \text{diag}(\Theta)\|_1 \right) \tag{11}$$

where $\widehat{\Theta}$ is an estimator of the precision matrix $\Theta = \Sigma^{-1}$, $||\Theta - \text{diag}(\Theta)||$ denotes the sum of the absolute values of the off-diagonal elements of Θ, S is the sample covariance matrix, and $\rho \geq 0$ the regularization parameter setting the sparsity in the precision matrix. Here $\Theta \succeq 0$ denotes the set of positive semi-definite matrices, ensuring that the estimate is guaranteed to be positive semi-definite. When the regularization parameter $\rho = 0$, the estimator for Σ and Θ are the sample covariance matrix S and its inverse S^{-1}. Increasing ρ leads to more l_1 shrinkage and a sparser estimate for Θ. The optimal regularization parameter is found via five-fold cross-validation to minimize the portfolio variance described in section 4.8. Different regularization parameters are considered, with values between 0.01 and 0.99, divided into ten equal steps.

This study considers applying the Graphical Lasso method to a factor-covariance structure. Similarly to Fan, Liao, and Mincheva (2011), first, the residuals from the time-series regression in Equation (3) are estimated:

$$\hat{\epsilon} = (\hat{\epsilon}_1, ..., \hat{\epsilon}_N)$$

and take the sample covariance matrix of the estimated residuals to get the covariance estimate of the error terms $\widehat{\Sigma}_\epsilon$. We then apply the Graphical Lasso method on $\widehat{\Sigma}_\epsilon$, to obtain the sparse precision matrix $\widehat{\Theta}_\epsilon$. Finally, the precision matrix of the returns is given by:

$$\widehat{\Theta} = \widehat{\Theta}_\epsilon - \widehat{\Theta}_\epsilon \widehat{B}[\widehat{\Theta}_f + \widehat{B}' \widehat{\Theta}_\epsilon \widehat{B}]^{-1} \widehat{B}' \widehat{\Theta}_\epsilon. \tag{12}$$

4.7 Parametric portfolio policy

The previous methods aim to reduce the estimation error in the estimate of the covariance matrix, which can then be used to construct global minimum variance portfolios. Instead of modeling the moments of returns, the parametric portfolio policy introduced by Brandt et al. (2009) directly models the portfolio weights as a function of firm characteristics. The parametric portfolio policy optimises the investor's utility by using a set of firm characteristics to tilt a benchmark portfolio (such as the value-weighted portfolio) toward stocks that help to increase the investor's utility. The parametric portfolio at time t can be written as:

$$w_t(\theta) = w_{b,t} + (x_{1,t}\theta_1 + x_{2,t}\theta_1 + ... + x_{K,t}\theta_K)/N_t \tag{13}$$

where N_t denotes the number of assets in the investment universe at time t, $w_{b,t}$ is the benchmark portfolio at time t, $x_{k,t}$ is the kth firm characteristic standardized cross-sectionally to have zero mean and unit variance across all stocks at time t, and $\theta = (\theta_1, ..., \theta_K)'$ is a vector of coefficients to be estimated corresponding to K firm characteristics. The parametric portfolio return at time $t + 1$ is given by:

$$r_{p,t+1}(\theta) = w'_{b,t}r_{t+1} + \frac{1}{N_t}\theta' X'_t r_{t+1} = r_{b,t+1} + \theta' r_{c,t+1} \tag{14}$$

where $r_{t+1} \in \mathbb{R}^{N_t}$ is the vector of asset returns at time $t + 1$, $X_t = (x_{1,t}, x_{2,t}, ..., x_{K,t})$ is the $N_t \times K$ matrix containing K firm characteristics for all assets in the investment universe at time t, $r_{b,t+1} = w'_{b,t}r_{t+1}$ is the return of the benchmark portfolio at time $t+1$, and $r_{c,t+1} = \frac{1}{N_t}X'_t r_{t+1}$ is the characteristic return vector. The intuition behind the characteristic return vector is that it contains K zero-investment portfolios scaled by the number of firms N_t, as each firm characteristic is standardized cross-sectionally to have zero mean and unit variance. Equation (14) shows that the parametric portfolio return combines the benchmark portfolio return with the characteristic portfolio return.

Minimum variance parametric portfolios

This study considers optimising a minimum variance portfolio. An advantage of optimizing a minimum variance portfolio is that it fully captures the low-risk anomaly, which is found to be relatively stable and persistent over time (Scherer, 2011). At the same time, information in firm characteristics regarding expected returns can be unstable and might be weakened over time (Chordia et al., 2014; Green et al., 2017). Particularly, the investor solves the following problem:

$$\min_{\theta} \quad V_t[r_{p,t+1}(\theta)] \tag{15}$$

where $V_t[r_{p,t+1}(\theta)]$ is the variance of the parametric portfolio return, respectively. Given T historical observations of returns and characteristics, DeMiguel et al. (2020, Proposition 1) shows that the problem in Equation (15) is equivalent to solving:

$$\min_{\theta} \ \theta' \widehat{\Sigma}_c \theta + 2\theta' \hat{\sigma}_{bc} \tag{16}$$

where $\widehat{\Sigma}_c$ is the sample covariance matrix of the characteristic return vector r_c, and $\hat{\sigma}_{bc}$ is the sample vector of covariances between the benchmark portfolio return r_b and the characteristic return vector r_c. Equation (16) shows that the minimum variance parametric portfolio finds the optimal trade-off between the variance of the characteristic portfolio return $\theta' \widehat{\Sigma}_c \theta$, and the covariance of the characteristic portfolio return with the benchmark portfolio return $2\theta' \hat{\sigma}_{bc}$.

Regularized parametric portfolios

The standard parametric portfolio can lead to overfitting when dealing with a large set of characteristics. The regularized parametric portfolio introduced by DeMiguel et al. (2020) tries to mitigate this problem by imposing a Lasso constraint on the parametric portfolio. The regularized parametric portfolio problem is given by:

$$\begin{aligned} \min_{\theta} \ & \theta' \widehat{\Sigma}_c \theta + 2\theta' \hat{\sigma}_{bc} \\ s.t. \ & ||\theta||_1 < \delta \end{aligned} \tag{17}$$

where $||\theta||_1 = \sum_{k=1}^{K} |\theta_k|$ is the 1-norm of θ and δ is the threshold parameter. The idea behind the threshold parameter is that when $\delta = 0$, the regularized parametric portfolio becomes the benchmark portfolio. As δ increases, the regularized parametric portfolio change from the benchmark portfolio toward the standard parametric portfolio. Here, the optimal threshold parameter is found via a five-fold cross-validation to minimize the portfolio variance described in section 4.8. Fifteen different threshold parameters are considered, with values between 1 and 28.

***K*-threshold Parametric Portfolios**

Another way to overcome the problem of overfitting when dealing with many characteristics is a new approach I introduce, K-threshold parametric portfolios. This method first estimates K parametric portfolios with a single characteristic, obtains the in-sample scores (the portfolio variances in this study), sorts them from best (lowest variance) to worst (highest variance), and then selects the top K characteristics which provide the best in-sample scores. Finally, using these top K characteristics, the standard parametric portfolio is estimated. The optimal value for the number of characteristics $\widehat{K}$ is found via five-fold cross-validation to minimize the portfolio variance described in section 4.8. Ten different values for K are considered ranging from 2 to 20. The intuition behind this method is to reduce the number of parameters by cross-validating the number of characteristics to include. Moreover, when estimating parametric portfolios with a single characteristic, the chances of overfitting are relatively small. Thus the in-sample score can provide insights into how much information a specific characteristic can provide to create deviations from the benchmark portfolio to minimize the portfolio variance. Algorithm 1 describes the estimation procedure of the K-threshold parametric portfolios.

Algorithm 1 K-threshold parametric portfolios

1: Estimate K univariate parametric portfolios and obtain the in-sample portfolio variance.
2: Sort the portfolio variances from low to high.
3: Select the top $\widehat{K}$ characteristics that provide the lowest portfolio variance, where the number of characteristics $\widehat{K}$ is found using five-fold cross-validation to minimize the portfolio variance described in section 4.8.
4: Estimate the standard parametric portfolio using these $\widehat{K}$ characteristics.

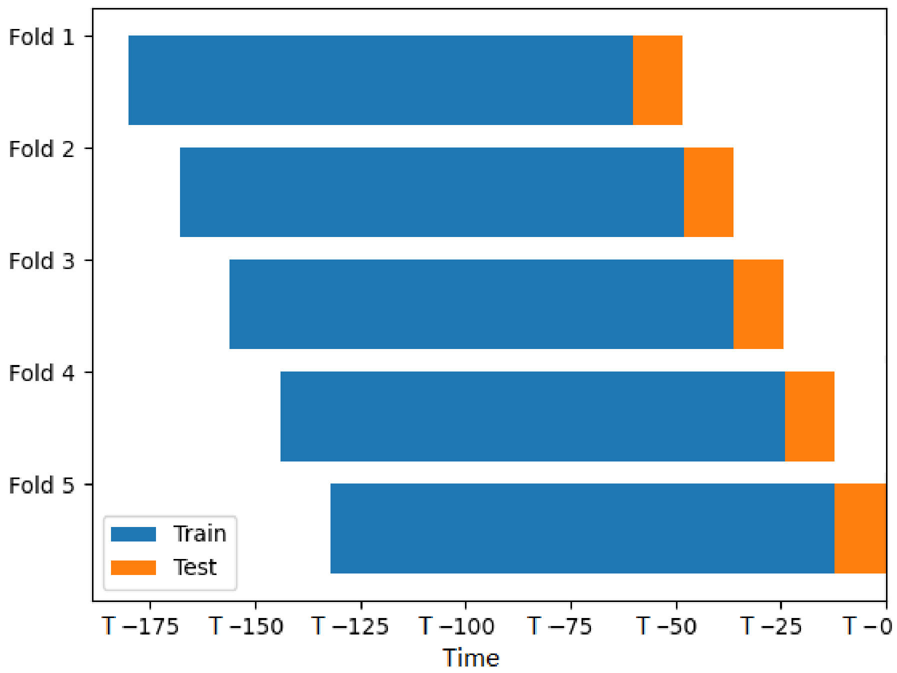
4.8 Time-series cross-validation

Some of the methods previously discussed methods rely on a choice of hyperparameters. Hyperparameters include, for example, the shrinkage intensity (ν) in linear shrinkage, the threshold parameter (γ) in Adaptive Thresholding, the regularization parameter (ρ) in Factor Graphical Lasso, the threshold parameter (δ) in the regularized parametric portfolio, and the number of characteristics (K) in the K-threshold parametric portfolio. This paper finds an optimal value for these hyperparameters via five-fold cross-validation to minimize the portfolio variance. As classical cross-validation techniques such as k-fold assume i.i.d. data, poor results might be obtained when this assumption does not hold. Time-series k-fold cross-validation, as described in Hyndman and Athanasopoulos (2018), is specially designed when dealing with time-series data. The time-series cross-validation is implemented with a blocking time-series split. Algorithm (2) explains how the cross-validation is implemented. Figure 2 illustrates how the data is split into five equal-sized folds with training and test sets.

Algorithm 2 Five-fold time-series cross-validation with a blocking time-series split

Ensure: The optimal hyperparameter that minimizes the portfolio variance.
1: Standing at date T, create five equal-sized validation sets using the most recent data where each training set consists of 120 observations and each test set of 12 observations, with each test-set non-overlapping.
2: For each validation set, compute the portfolios using the training set, obtain the portfolio returns on the test set and compute the portfolio variance over the 12 portfolio test returns.
3: Calculate the average portfolio variance over the five-folds.
4: Obtain the hyperparameter with the lowest average portfolio variance.
5: Use this tuned hyperparameter to compute the portfolio at date T, and obtain the out-of-sample portfolio returns $(r_{p,T+1}, ...r_{p,T+12})$.
6: Repeat from 1) now standing at $T + 12$
7: Repeat this process until the end of the out-of-sample period is reached.

Figure 2: Example of a five-fold time-series cross-validation with a blocking time-series split



4.9 Transaction costs

This section discusses the estimation of transaction costs, which are important to consider as they can easily erode the gains from a portfolio strategy. In this study, the transaction costs are estimated following the methodology in Brandt et al. (2009). The portfolio turnover is calculated as follows:

$$TO_t = \sum_{i=1}^{N_t} |w_{i,t} - w_{i,t-1}^{+}| \tag{18}$$

where $w_{i,t-1}^{+} = \frac{w_{i,t-1}(1+r_{i,t})}{1+r_{p,t}}$ is the portfolio weight of stock i at time t before rebalancing with $r_{p,t}$ denoting the portfolio return for a given portfolio strategy at time t. By Brandt et al. (2009), the portfolio return net of transaction costs at time $t+1$ is estimated as:

$$r_{p,t+1} = \sum_{i=1}^{N_t} w_{i,t} r_{i,t+1} - TC_{i,t}|w_{i,t} - w_{i,t-1}^{+}| \tag{19}$$

where $TC_{i,t} = 0.006 - 0.0025 \times \widehat{me}_{i,t}$ is the proportional transaction cost for stock i at time t. Here, $\widehat{me}_{i,t}$ is the relative size of stock i at time t, normalized to be between 0 and 1. The stock with the smallest relative size has a transaction cost of 0.6%, whereas the largest one has a cost of 0.35%. The reason for considering the relative size is that the literature finds that transaction costs are larger for small-caps than large-caps and have been gradually decreasing over time (Domowitz, Glen, & Madhavan, 2001; Hasbrouck, 2009; Keim & Madhavan, 1997). Reasons for these empirical facts have to do with liquidity, as large-cap stocks are often more liquid than small-cap stocks, increased institutional presence in the market resulting in a more competitive environment for trading services, and technological innovations in trading.

5 Empirical results

This section provides the out-of-sample analysis to find out if investors can still economically benefit from using firm characteristics for large-N portfolio allocation in the presence of transaction costs. Table 3 gives a summary of the notation for each portfolio studied. Following Brandt et al. (2009), the benchmark portfolio used to estimate the parametric portfolios is the value-weighted portfolio. As in Brandt et al. (2009), the research does not include the risk-free asset in the investment opportunity set. The reason for not including the risk-free asset is that the first-order effect of allowing investments in the risk-free asset is to vary the leverage of the portfolio, which only corresponds to a change in the scale of the stock portfolio weights (Brandt et al., 2009). The coding is done in R and available on GitHub[2].

[2]The code can be found under repository: Minimum-variance-parametric-portfolio-policies-Msc-thesis.

Table 3: List of various asset-allocation models considered

Abbreviation	Model
GMV	
EFM	Global minimum variance portfolio with an exact factor-covariance estimator using the three factors in Fama and French (1993)
LS	Global minimum variance portfolio with the Linear shrinkage covariance estimator
NLS	Global minimum variance portfolio with the Nonlinear shrinkage covariance estimator
AT	Global minimum variance portfolio with the Adaptive Thresholding covariance estimator
FGL	Global minimum variance portfolio with the Factor Graphical Lasso covariance estimator using the three factors in Fama and French (1993)
Naive / Passive	
$1/N$	Equally-weighted portfolio
VW	Value-weighted portfolio
Parametric	
PPP (FF3)	Minimum variance parametric portfolio using the characteristics: beta, me, bm
PPP (FF5)	Minimum variance parametric portfolio using the characteristics: beta, me, bm, agr, operprof
PPP (L1)	Minimum variance regularized parametric portfolio using all 39 characteristics
PPP (Thr)	Minimum variance K-threshold parametric portfolio using all 39 characteristics

5.1 Methodology for out-of-sample evaluation

A rolling-window procedure similar to that used in Jagannathan and Ma (2003) and Ledoit and Wolf (2004a) is implemented to evaluate the out-of-sample results. At the end of January, the largest $N = 500$ stocks are selected from all common stocks with a stock price greater than \$5, for which there is return data for the previous 120 months and the next 12 months. These 500 stocks are then selected as the investment universe for the next 12 months. Next, the various portfolios are computed using the return data for the preceding 120 months. Finally, each of these portfolios are held for 12 months, rebalanced every month, and their monthly returns are recorded. This procedure is continued until the end of the out-of-sample period is reached. For each portfolio studied, the following performance measures are computed:

- $\hat{\mu}_p$, the mean of the excess portfolio returns net of transaction costs.
- $\hat{\sigma}_p$, the volatility of the excess portfolio returns net of transaction costs.
- $\widehat{SR}_p = \hat{\mu}_p / \hat{\sigma}_p$, the Sharpe ratio net of transaction costs.
- $\overline{TO} = \frac{1}{\tau} \sum_{t=T}^{T+\tau} \sum_{i=1}^{N} |w_{i,t} - w_{i,t-1}^{+}|$, which denotes the average monthly turnover, with τ denoting the out-of-sample size.

5.2 Testing out-of-sample portfolio performance

The statistical significance of the out-of-sample portfolio variance and Sharpe ratio are tested against the $1/N$ benchmark strategy. As returns are hardly i.i.d., the studentized circular block bootstrapping methodology in Ledoit and Wolf (2008, 2011) is used to compute the bootstrap standard errors, and p-values with 10,000 bootstrap resamples and a block with the size of 5. The details of these tests are specified in the Appendix Section D.

5.3 Out-of-sample performance

Table 4 present the out-of-sample results for the unconstrained portfolios. The table is divided into two sections describing the (i) performance evaluation and (ii) distribution of the portfolio weights separately. The out-of-sample performance is compared and tested relative to the performance of the $1/N$ benchmark strategy, using the following three performance criteria: (i) the out-of-sample portfolio volatility, (ii) Sharpe ratio, and (iii) the turnover (trading volume) for each portfolio strategy. For ease of interpretation, the portfolio volatility and Sharpe ratio are annualized.

Table 4 shows that the parametric and global minimum variance (GMV) portfolios achieve significantly lower portfolio volatilities than the $1/N$ benchmark strategy, with the differences being statistically significant. There do not seem to be significant differences within the set of parametric portfolios in terms of portfolio volatility. The annualized portfolio volatility for the parametric portfolios ranges from 11.83% to 12.34% and seems to decrease slightly as the number of characteristics increases. The annualized portfolio volatility for the GMV portfolios ranges from 10.97% to 13.56%, and for the equally-weighted $1/N$ benchmark strategy, 15.48%. This suggests that the unconstrained parametric and GMV portfolios can successfully reduce the portfolio risk relative to the naive $1/N$ benchmark strategy.

Furthermore, the low portfolio volatility translates into higher Sharpe ratios for the K-threshold parametric portfolio (PPP(Thr)) and the parametric portfolio that exploits five characteristics (PPP(FF5)), which outperform all of the GMV portfolios and the $1/N$ benchmark strategy in terms of the Sharpe ratio. PPP(Thr) achieve an annualized Sharpe ratio of 0.64, and PPP(FF5) achieve 0.67. However, the gains are modest compared to Factor Graphical Lasso (FGL), Adaptive Thresholding (AT), and the $1/N$ benchmark strategy, which achieve annualized Sharpe ratios ranging from 0.60 to 0.63. Moreover, Table 4 shows that no asset allocations can provide a Sharpe ratio that is statistically significantly better than the Sharpe ratio of the $1/N$ benchmark strategy. Overall, the unconstrained minimum variance parametric portfolios do not seem to have significant economic gains.

Next, it is important to consider the amount of trading (turnover) each portfolio strategy requires. Table 4 reports two measures for the turnover: $\overline{TO}$ denotes the average monthly turnover, and $\overline{TO}_{t=1}$ denotes the average January turnover. The reason for also reporting the January turnover is to show that most of the turnover comes from January as the investment universe changes and new portfolios are constructed here. The unconstrained parametric portfolios tend to involve high trading costs in January. This is even more extreme for the GMV portfolios. This results in a turnover for the parametric portfolios ranging from 14.64% to 25.77% and for the GMV portfolio ranging from 10.93% to 50.00%, significantly higher than the turnover for the $1/N$ benchmark strategy of 6.19%. The turnover for the parametric portfolios seems to be increasing with the number of characteristics, resulting in a turnover that is almost two times higher for the regularized parametric portfolio (PPP(L1)) than the parametric portfolio that exploits only three characteristics (PPP(FF3)). Overall, this suggests that the out-of-sample performance could be impacted by trading costs in a significant matter, and imposing additional constraints could be helpful. Moreover, this suggests that although the portfolio risk can be significantly reduced by using unconstrained parametric portfolios relative to the $1/N$ benchmark strategy, the modest economic gains are offset mainly by the high turnover.

Further, the last five rows of Table 4 report the portfolio weights statistics, averaged across time to examine the stability of the portfolio weights. When analyzing the distribution of the weights, it is not surprising to see that the parametric portfolios take larger positions, as they are active portfolios. The average absolute weight for the parametric portfolios ranges from 0.34% to 0.41%, as opposed to the average absolute weight of 0.20% for the $1/N$ benchmark strategy. However, these positions are not extreme. For instance, the maximum weight for the parametric portfolios ranges on average from 1.17% to 2.43%, and the minimum weight ranges from -1.74% to -1.16%. Overall, the parametric portfolios do not reflect unreasonably extreme bets on individual stocks.

Finally, it is also important to consider the amount of leverage a portfolio strategy requires. The average sum of negative weights for the parametric portfolios ranges from -34% to -56%, which implies that the sum of long positions ranges, on average, from 134% to 156%, respectively. Finally, the average fraction of negative weights (shorted stocks) ranges from 0.27 to 0.35. Overall, the parametric portfolios do not seem to require unreasonable amounts of leverage and could be well implemented by a long-short equity fund.

Table 4: Unconstrained portfolios

This table shows the out-of-sample performance and distribution of the portfolio weights for the unconstrained optimal portfolios and benchmark portfolios specified in Table 3. The tables follow the structure of the tables in Brandt, Santa-Clara, and Valkanov (2009). The investment universe consists of a restricted set of the largest 500 companies each January. The out-of-sample period covers January 1985 to December 2016. The performance measures are presented net of transaction costs. The statistical significance of the out-of-sample portfolio variance and Sharpe ratio are tested against the $1/N$ benchmark strategy using the circular block bootstrapping method from Ledoit and Wolf (2008, 2011) with 10,000 bootstrap samples and a block with the size of 5. Here, $s(.)$ and p-value denote the standard error and p-value computed from the bootstrap data for a given performance measure. A significance level of $\alpha = 5\%$ is considered. *** denotes significance at 0.1%, ** denotes significance at 1%, and * denotes significance at 5% level. The last five rows report the portfolio weights statistics, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average maximum and minimum portfolio weight ($\max w_i$ and $\min w_i$), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N$).

	Parametric				**GMV**					**Naive / Passive**	
	PPP(Thr)	PPP(L1)	PPP(FF5)	PPP(FF3)	EFM	LS	NLS	FGL	AT	$1/N$	VW
$\hat{\mu}_p$ (%)	7.61	6.86	8.17	7.07	4.69	5.87	4.53	7.43	8.58	9.24	8.53
$\hat{\sigma}_p$ (%)	11.83***	11.71***	12.12***	12.34***	12.45***	11.17***	10.97***	11.84***	13.56***	15.48	14.80***
$s(\hat{\sigma}_p - \hat{\sigma}_b)$	(0.091)	(0.101)	(0.095)	(0.101)	(0.132)	(0.091)	(0.100)	(0.093)	(0.03)	(-)	(0.019)
$\hat{\sigma}_p$ p-value	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	-	0.00
$\widehat{SR}_p$	0.64	0.59	0.67	0.57	0.38	0.53	0.41	0.63	0.63	0.60	0.58
$s(\widehat{SR}_p - \widehat{SR}_b)$	(0.042)	(0.043)	(0.043)	(0.041)	(0.059)	(0.047)	(0.051)	(0.045)	(0.012)	(-)	(0.010)
$\widehat{SR}_p$ p-value	0.78	0.94	0.62	0.87	0.32	0.69	0.33	0.86	0.42	-	0.52
$\overline{TO}$ (%)	23.03	25.77	18.87	14.64	40.57	50.00	48.70	27.01	10.93	6.19	6.15
$\overline{TO}_{t=1}$ (%)	156.74	181.39	106.09	76.74	350.83	411.77	397.71	223.75	78.23	13.89	17.66
$100 \times \lvert w_i \rvert$	0.41	0.42	0.39	0.34	0.50	0.63	0.60	0.36	0.21	0.20	0.20
$100 \times \max w_i$	2.01	2.43	1.46	1.17	3.28	2.37	2.33	1.23	1.16	0.20	0.69
$100 \times \min w_i$	-1.70	-1.60	-1.74	-1.16	-1.62	-2.27	-2.28	-1.07	-0.54	0.20	0.04
$\sum w_i I(w_i < 0)$	-0.56	-0.52	-0.47	-0.34	-0.75	-1.08	-1.01	-0.41	-0.03	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.32	0.35	0.31	0.27	0.41	0.33	0.36	0.29	0.02	0.00	0.00

5.4 Economic interpretation of coefficients

To better understand which characteristics tend to be informative, I report the estimated coefficients averaged over time for the regularized parametric portfolio in Table 5 and the K-threshold parametric portfolio in Table 6. In addition, I added how often a specific characteristic is shrunk towards zero for the regularized parametric portfolio or set to zero for the K-threshold parametric portfolio. First, I discuss the regularized parametric portfolio's coefficients and then point out how they differ from the K-threshold parametric portfolio. Since the parametric portfolio policy imposes that all characteristics are standardized cross-sectionally, the coefficients are directly comparable. Despite not having standard errors, there can be accurately inferred which variables were dropped and evaluate the sign and magnitude of coefficients in economic terms.

Table 5 shows that the Lasso has performed a variable selection because many of the values are close to 0. Not surprising, of course, as Lasso performs l_1 shrinkage so that there are "corners" in the constraint, which in two dimensions corresponds to a diamond. If the sum of squares "hits" one of these corners, then the coefficient corresponding to the axis is shrunk to zero. The regularized parametric portfolio finds the beta and leverage (lev) most informative, as they are never shrunk towards zero and obtain the largest coefficients in absolute sign. The negative sign of -1.31 for beta and -0.48 for lev implies that the regularized parametric portfolio goes long on average in firms with low beta and low debt relative to equity and short in firms with high beta and high debt relative to equity. The negative sign of beta intuitively makes sense since the parametric portfolio optimizes a minimum variance portfolio, and beta is a measure of risk. Moreover, the sign is consistent with the findings in Frazzini and Pedersen (2014), who find that high beta is associated with low returns and low beta with high returns. The negative sign of lev differs from the empirical finding in Bhandari (1988), which document that stock returns are positively related to the debt-to-equity ratio. However, an explanation for this is that the expected stock returns are omitted, and the objective is variance minimization. Other characteristics that are worth pointing out are the beta-squared (betasq), industry momentum (indmom), one-month momentum (mom1m), dividend yield (dy), and size (me).

Furthermore, Table 6 shows that the coefficients for the K-threshold parametric portfolio differ from the regularized parametric portfolio coefficients and that the K-threshold parametric portfolio mostly depends on two characteristics: beta and beta-squared (betasq). The signs are similar to the regularized parametric portfolio but larger, which makes sense as the coefficients are estimated without imposing the Lasso constraint on the parameter vector. On average, The K-threshold parametric portfolio put a large negative sign of -3.18 on beta and a large positive sign of 1.74 on betasq. This implies that, on average, the K-threshold parametric portfolio overweights firms with low beta but high betasq and underweight firms with high beta but low betasq.

Table 5: Regularized coefficients

This table shows the time-series averages of the estimated coefficients for the regularized parametric portfolio (PPP(L1)). In addition, the table includes how many times a specific coefficient is shrunk towards 0 when estimating PPP(L1) (# Shrunk to 0). The investment universe consists of a restricted set of the largest 500 companies each January. The out-of-sample period covers January 1985 to December 2016. The description of each characteristic can be found in Table 1.

Coefficient	average value	# Shrunk to 0 (%)	Coefficient	average value	# Shrunk to 0 (%)
beta	-1.31	0.00			
betasq	0.26	9.38	ep	-0.01	43.75
chmom	0.05	43.75	gma	0.28	43.75
idiovol	-0.01	31.25	herf	0.03	50.00
indmom	-0.10	6.25	hire	0.06	46.88
mom1m	0.27	3.12	lev	-0.48	0.00
mom6m	-0.13	25.00	lgr	0.01	50.00
mom12m	0.11	28.12	mve_ia	-0.03	43.75
pricedelay	0.16	56.25	operprof	-0.03	43.75
age	-0.07	50.00	pchgm_pchsale	-0.18	50.00
agr	0.39	34.38	ps	0.28	31.25
bm	0.16	31.25	rd	0.28	34.38
bm_ia	0.26	34.38	salecash	0.27	40.62
cashpr	0.32	34.38	salerec	0.20	46.88
chcsho	-0.11	40.62	sgr	-0.13	46.88
chempia	-0.06	56.25	sp	-0.26	37.50
chpmia	0.24	28.12	baspread	0.17	31.25
convind	0.17	59.38	maxret	-0.02	53.12
dy	0.35	9.38	retvol	-0.19	43.75
egr	0.26	40.62	me	-0.29	18.75

Table 6: K-threshold coefficients

This table shows the time-series averages of the estimated coefficients for the K-threshold parametric portfolio (PPP(Thr)). In addition, the table includes how many times a specific coefficient is not used when estimating PPP(Thr) (# Set to 0). The investment universe consists of a restricted set of the largest 500 companies each January. The out-of-sample period covers January 1985 to December 2016. The description of each characteristic can be found in Table 1.

Coefficient	average value	# Set to 0 (%)	Coefficient	average value	# Set to 0 (%)
beta	-3.18	0.00			
betasq	1.74	0.00	ep	0.01	87.50
chmom	0.01	96.88	gma	0.00	100.00
idiovol	0.32	28.12	herf	0.03	78.12
indmom	0.00	96.88	hire	-0.01	84.38
mom1m	0.02	96.88	lev	-0.07	90.62
mom6m	0.00	100.00	lgr	-0.24	87.50
mom12m	0.00	100.00	mve_ia	0.09	90.62
pricedelay	0.19	68.75	operprof	0.04	90.62
age	0.03	59.38	pchgm_pchsale	0.00	100.00
agr	-0.04	75.00	ps	0.09	90.62
bm	0.01	96.88	rd	0.00	100.00
bm_ia	0.00	100.00	salecash	0.09	84.38
cashpr	0.15	87.50	salerec	0.02	96.88
chcsho	-0.14	75.00	sgr	-0.01	96.88
chempia	0.00	100.00	sp	-0.05	93.75
chpmia	0.00	100.00	baspread	-0.15	21.88
convind	0.06	75.00	maxret	0.27	56.25
dy	0.10	37.50	retvol	-0.08	28.12
egr	0.00	96.88	me	-0.04	81.25

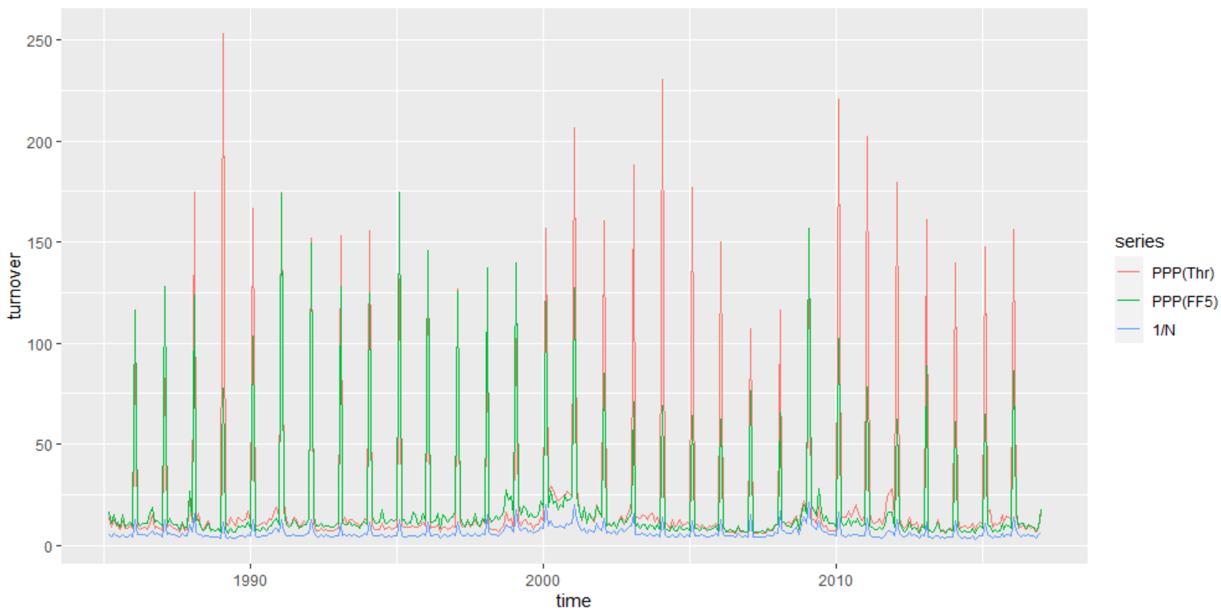
6 Portfolio Weight Constraints

The previous results indicate that unconstrained parametric portfolios could not benefit investors significantly. The gains in the Sharpe ratios were small and insignificant, yet the turnover was significantly higher compared to the equally-weighted $1/N$ benchmark strategy. This section studies how the out-of-sample results change when imposing portfolio weight constraints.

6.1 Turnover Constraint

Table 4 shows that most of the turnover comes from January. This intuitively makes sense as the investment universe changes here, and new portfolios are constructed, which are held throughout the year. The cost of rebalancing towards a new portfolio with a new set of stocks is much higher than simply rebalancing towards the previous portfolio. Figure 3 illustrates this and shows the portfolio turnover during the out-of-sample period for the K-threshold parametric portfolio (PPP(Thr)), the parametric portfolio that exploits five characteristics (PPP(FF5)), and the $1/N$ benchmark strategy. The unconstrained parametric portfolios seem to have very high turnover when new portfolios are constructed relative to the $1/N$ benchmark strategy.

Figure 3: A plot of the portfolio turnover during the out-of-sample period for the K-threshold parametric portfolio (PPP(Thr)), the parametric portfolio that exploits five characteristics (PPP(FF5)), and the equally-weighted $1/N$ benchmark strategy.



For this reason, I investigate the effect of imposing a turnover constraint when estimating the parametric and global minimum variance portfolios. Specifically, the portfolios are estimated with a turnover constraint so that the turnover is allowed to be 50%. The turnover-constrained

GMV portfolio problem can be formulated as:

$$\begin{aligned} \min_{w_t} \quad & w_t' \Sigma w_t \\ \text{s.t.} \quad & |w_t - w_{t-1}^{+}| < 50\% \\ & w_t' \iota = 1 \end{aligned} \tag{20}$$

Where w_{t-1}^{+} denotes the current weights in the portfolio before rebalancing. The turnover-constrained minimum variance parametric portfolio problem can be formulated as follows:

$$\begin{aligned} \min_{\theta} \quad & \theta' \hat{\Sigma}_c \theta + 2\theta' \hat{\sigma}_{bc} \\ \text{s.t.} \quad & |w_t(\theta) - w_{t-1}^{+}| < 50\% \end{aligned} \tag{21}$$

Where $w_t(\theta) = w_{b,t} + \frac{1}{N_t} X_t \theta$ denotes the parametric portfolio weights at time t which depend on the parameter vector θ that needs to be optimised. Table 7 presents the results.

Table 7 shows that the portfolio volatility for the parametric portfolios does not tend to increase significantly with the turnover constraint. The annualized portfolio volatility for the parametric portfolios ranges from 12.03% to 12.38%. This suggests that the turnover-constrained parametric portfolios can still significantly reduce the portfolio risk. The portfolio volatility for the global minimum variance (GMV) portfolios does increase slightly and ranges from 13.20% to 15.29%. Yet, except for the exact factor-covariance estimator (EFM) and nonlinear shrinkage (NLS), the parametric and GMV portfolios achieve portfolio volatilities statistically significantly lower than the $1/N$ benchmark strategy.

Further, the Sharpe ratios for the parametric and GMV portfolios increase, implying that the turnover constraint improves out-of-sample performance net of transaction costs. Similarly to the results in Table 4 for the unconstrained portfolios, the parametric portfolios achieve higher Sharpe ratios than the GMV portfolios and the $1/N$ benchmark strategy. The annualized Sharpe ratios for the parametric portfolios range from 0.69 to 0.75, with the strongest performance from the parametric that exploits five characteristics (PPP(FF5)). The annualized Sharpe ratios for the GMV portfolios range from 0.48 to 0.67, with Factor Graphical Lasso achieving the strongest out-of-sample performance of the GMV portfolios. Although the significance test shows that no asset allocations have a statistically significantly better Sharpe ratio than the $1/N$ benchmark strategy, the p-values for the turnover-constrained parametric portfolios are much lower than for the unconstrained parametric portfolios.

Moreover, Table 7 shows the turnover is substantially reduced for the parametric and GMV portfolios, which is not surprising given the turnover constraint. The turnover for the parametric portfolios ranges from 11.34% to 12.05%, and the turnover for the GMV portfolios ranges from 8.72% to 23.63%. The substantial reduction in portfolio volatility, the increase in the Sharpe ratios, and a significant turnover reduction suggest some considerable economic gains from using parametric portfolios when imposing a turnover constraint.

Table 7 also shows that the turnover constraint makes the parametric portfolio's weight distribution less extreme. The average absolute weight for the parametric portfolios ranges from 0.29% to 0.31%, the maximum weight ranges from 1.08% to 1.60%, and the minimum weight ranges from -0.92% to -0.59%. This suggests that there are not only economic benefits from imposing a turnover constraint on the parametric portfolios, but there is also slightly more diversification within the portfolio policy.

Finally, Table 7 shows that the turnover constraint reduced the leverage for the parametric portfolios. The average sum of negative weights for the turnover-constrained parametric portfolios ranges from -24% to -27%, which implies that the sum of long positions on average ranges from 124% to 127%, respectively. The fraction of short positions on average ranges from 0.24 to 0.29, which implies a lower amount of short positions on average compared to the unconstrained parametric portfolios.

Table 7: Turnover constraint

This table shows the out-of-sample performance and distribution of the portfolio weights for the optimal portfolios with a turnover constraint of 50%, and benchmark portfolios specified in Table 3. The tables follow the structure of the tables in Brandt, Santa-Clara, and Valkanov (2009). The investment universe consists of a restricted set of the largest 500 companies each January. The out-of-sample period covers January 1985 to December 2016. The performance measures are presented net of transaction costs. The statistical significance of the out-of-sample portfolio variance and Sharpe-ratio are tested against the $1/N$ benchmark strategy using the circular block bootstrapping method from Ledoit and Wolf (2008, 2011) with 10,000 bootstrap samples and a block with the size of 5. Here, $s(.)$ and p-value denote the standard error and p-value computed from the bootstrap data for a given performance measure. A significance level of $\alpha = 5\%$ is considered. *** denotes significance at 0.1%, ** denotes significance at 1%, and * denotes significance at 5% level. The last five rows report the portfolio weights statistics, averaged across time, specified in Table 4.

	Parametric				**GMV**					**Naive / Passive**	
	PPP(Thr)	PPP(L1)	PPP(FF5)	PPP(FF3)	EFM	LS	NLS	FGL	AT	$1/N$	VW
$\hat{\mu}_p$ (%)	8.71	8.80	9.31	8.54	7.94	8.54	7.38	8.87	8.80	9.24	8.53
$\hat{\sigma}_p$ (%)	12.03**	12.20**	12.35**	12.38**	14.79	13.50**	15.29	13.20**	13.92***	15.48	14.80***
$s(\hat{\sigma}_p - \hat{\sigma}_b)$	(0.072)	(0.073)	(0.076)	(0.082)	(0.085)	(0.046)	(0.079)	(0.036)	(0.024)	(-)	(0.019)
$\hat{\sigma}_p$ p-value	0.00	0.00	0.00	0.00	0.31	0.00	0.77	0.00	0.00	-	0.00
$\widehat{SR}_p$	0.72	0.72	0.75	0.69	0.54	0.63	0.48	0.67	0.63	0.60	0.58
$s(\widehat{SR}_p - \widehat{SR}_b)$	(0.031)	(0.031)	(0.033)	(0.033)	(0.035)	(0.015)	(0.034)	(0.018)	(0.010)	(-)	(0.010)
$\widehat{SR}_p$ p-value	0.27	0.27	0.20	0.46	0.63	0.52	0.35	0.22	0.30	-	0.52
$\overline{TO}$ (%)	11.34	11.82	12.05	11.59	21.29	11.52	23.63	10.37	8.72	6.19	6.15
$100 \times \lvert w_i \rvert$	0.30	0.29	0.31	0.30	0.65	0.31	0.72	0.26	0.21	0.20	0.20
$100 \times \max w_i$	1.24	1.60	1.16	1.08	8.05	1.10	7.34	1.24	0.97	0.20	0.69
$100 \times \min w_i$	-0.91	-0.59	-0.92	-0.86	-6.75	-1.21	-7.00	-1.30	-0.90	0.20	0.04
$\sum w_i I(w_i < 0)$	-0.24	-0.24	-0.27	-0.24	-1.13	-0.27	-1.31	-0.15	-0.01	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.25	0.29	0.26	0.24	0.51	0.24	0.50	0.23	0.01	0.00	0.00

6.2 Short-Sale Constraint

In this section, short-sale constraints are considered. The short-sale constrained global minimum variance (GMV) portfolios are computed by restricting the portfolio weights to be greater than zero during the optimization. Following Brandt et al. (2009), the short-sale constrained parametric portfolios are computed by first computing the unconstrained parametric portfolios, then setting all negative firm weights equal to zero, and finally normalizing the resulting vector so that the weights sum to one. This is implemented within the estimation procedure to ensure the results are robust. Table 8 presents the results.

Table 8 shows that the portfolio volatility for the parametric and GMV portfolios with short-sale constraints does not increase significantly compared to the unconstrained versions. The annualized portfolio volatility for the parametric portfolios ranges from 11.90% to 12.10%, and for the GMV portfolios ranges from 11.40% to 12.37%. This suggests that both the short-sale constrained parametric and the GMV portfolio can still significantly reduce the risk relative to naive diversification.

Moreover, the Sharpe ratios for the parametric and GMV portfolios increase relative to the unconstrained versions, implying that the short-sale constraint improves out-of-sample performance net of transaction costs. Moreover, Table 8 shows a slight improvement in the Sharpe ratios for the short-sale constrained parametric portfolios relative to the turnover-constrained parametric portfolios in Table 7, although these gains are modest. Similarly to the previous results, the parametric portfolios achieve higher Sharpe ratios than the GMV portfolios and the $1/N$ benchmark strategy, and the parametric portfolio that exploits five characteristics (PPP(FF5)) shows the strongest out-of-sample performance. The annualized Sharpe ratio for the short-sale constrained parametric portfolios ranges from 0.67 to 0.78, and for the GMV portfolios, from 0.55 to 0.64. This implies that PPP(FF5) achieve an annualized Sharpe ratio that is 30% higher than the $1/N$ benchmark strategy, and PPP(Thr) and PPP(FF3) achieve an annualized Sharpe ratio that is 21.67% higher. Although Table 8 shows that no asset allocations have a statistically significantly better Sharpe ratio than the $1/N$ benchmark strategy, the p-values for the short-sale constrained parametric portfolios are much lower compared to the unconstrained parametric portfolios and slightly lower compared to the turnover-constrained parametric portfolios, approaching significance, suggesting that there are some considerable economic gains in terms of the Sharpe ratio from using the short-sale constrained parametric portfolios.

Moreover, Table 8 shows that the turnover for the parametric portfolios is significantly reduced. The turnover for the short-sale constrained parametric portfolios is significantly lower than the unconstrained parametric portfolios and slightly lower than the turnover-constrained parametric portfolios. The turnover for the parametric portfolios ranges from 7.30% to 9.26%, which is only modestly higher than the turnover for the $1/N$ benchmark strategy of 6.19% and lower than the turnover for the GMV portfolio ranging from 13.54% to 18.39%. Overall, this suggests considerable economic gains from short-sale constrained parametric portfolios relative to the GMV portfolios and naive diversification.

Finally, Table 8 shows that the short-sale constrained parametric portfolios still do not involve extreme weights. In fact, the short-sale constraint makes the parametric portfolio weight distribution less extreme than the unconstrained versions. The average absolute weight for all parametric portfolios is 0.20%, which is the same as for the $1/N$ benchmark strategy. The maximum weight for the parametric portfolios ranges from 1.08% to 1.63%. On average, the short-sale constrained parametric portfolios invest only 57% to 59% of the stocks, which suggests that the portfolios contain, on average, around 285 to 295 long-positions out of the 500 stocks in the investment universe, which seems sufficient to maintain a well-diversified portfolio allocation.

Table 8: Short-Sale constraint

This table shows the out-of-sample performance and distribution of the portfolio weights for the optimal portfolios with short-sale constraints, and benchmark portfolios specified in Table 3. The tables follow the structure of the tables in Brandt, Santa-Clara, and Valkanov (2009). The investment universe consists of a restricted set of the largest 500 companies each January. The out-of-sample period covers January 1985 to December 2016. The performance measures are presented net of transaction costs. The statistical significance of the out-of-sample portfolio variance and Sharpe-ratio are tested against that of the $1/N$ benchmark strategy using the circular block bootstrapping method from Ledoit and Wolf (2008, 2011) with 10,000 bootstrap samples and a block with the size of 5. Here, $s(.)$ and p-value denote the standard error and p-value computed from the bootstrap data for a given performance measure. A significance level of $\alpha = 5\%$ is considered. *** denotes significance at 0.1%, ** denotes significance at 1%, and * denotes significance at 5% level. The last five rows report the portfolio weights statistics, averaged across time, specified in Table 5.

	Parametric				**GMV**					**Naive / Passive**	
	PPP(Thr)	PPP(L1)	PPP(FF5)	PPP(FF3)	EFM	LS	NLS	FGL	AT	$1/N$	VW
$\hat{\mu}_p$ (%)	8.72	8.07	9.37	8.80	6.74	7.77	6.73	7.85	6.62	9.24	8.53
$\hat{\sigma}_p$ (%)	11.90***	12.10***	12.05***	12.05***	11.94***	12.37***	11.40***	12.23***	12.04***	15.48	14.80***
$s(\hat{\sigma}_p - \hat{\sigma}_b)$	(0.060)	(0.053)	(0.055)	(0.059)	(0.112)	(0.072)	(0.070)	(0.084)	(0.057)	(-)	(0.019)
$\hat{\sigma}_p$ p-value	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.00	-	0.00
$\widehat{SR}_p$	0.73	0.67	0.78	0.73	0.56	0.63	0.59	0.64	0.55	0.60	0.58
$s(\widehat{SR}_p - \widehat{SR}_b)$	(0.025)	(0.023)	(0.025)	(0.026)	(0.044)	(0.025)	(0.034)	(0.029)	(0.024)	(-)	(0.010)
$\widehat{SR}_p$ p-value	0.14	0.43	0.13	0.16	0.84	0.75	0.96	0.69	0.62	-	0.52
$\overline{TO}$ (%)	8.20	9.26	7.59	7.30	18.07	13.54	18.39	13.75	16.63	6.19	6.15
$100 \times \lvert w_i \rvert$	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
$100 \times \max w_i$	1.17	1.63	1.11	1.02	9.41	3.53	5.93	2.98	5.09	0.20	0.69
$100 \times \min w_i$	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.20	0.04
$\sum w_i I(w_i < 0)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.41	0.43	0.43	0.42	0.90	0.43	0.87	0.42	0.61	0.00	0.00

7 Robustness checks

This section investigates the robustness of the main finding that parametric portfolios that optimize a minimum variance portfolio can benefit investors when imposing portfolio weight constraints. The following checks are performed: (i) sub-sample analysis (ii) allocating all stocks, and (iii) using a different objective function.

7.1 Sub-sample analysis

Chordia et al. (2014) and Green et al. (2017) show that the magnitude of asset return predictability has decreased in the last decade. To understand how the results vary over time, the out-of-sample results for the portfolios with short-sale constraints are examined for two different subperiods with a similar number of observations: January 1989 to December 2002 and January 2003 to December 2016. The split is motivated by the fact that Chordia et al. (2014), and Green et al. (2017) find that return predictability fell sharply after 2003. Table 9 reports the results.

Table 9 shows that the parametric portfolios achieve the strongest out-of-sample performance in both subperiods, whereas the global minimum variance (GMV) portfolios significantly underperform from January 1989 till December 2002. The short-sale-constrained minimum variance parametric portfolios performed better from January 2003 to December 2016. Moreover, the global minimum variance portfolios also show much better out-of-sample performance during this period. This suggests that when variance minimization is the objective, economic gains can still be realized by exploiting firm characteristics for portfolio allocation, even when taking into account the transaction costs. Another explanation for this strong performance has to do with stronger market conditions. The excess mean portfolio return for the $1/N$ benchmark strategy from January 2003 to December 2016 is 10.05% vs 7.63% from January 1989 to December 2002. This has also to do with a lower risk-free rate which is an annualized 1.18% from January 2003 to December 2016 vs 4.81% from January 1989 to December 2002.

Table 9: Pre and post 2003 analysis with Short-sale constraint

This table shows the out-of-sample performance and distribution of the portfolio weights for the optimal portfolios with short-sale constraints, and benchmark portfolios specified in Table 3. The tables follow the structure of the tables in Brandt, Santa-Clara, and Valkanov (2009). The investment universe consists of a restricted set of the largest 500 companies each January. The results are presented for two different subperiods with a similar number of observations: January 1989 to December 2002 and January 2003 to December 2016. The performance measures are presented net of transaction costs. The statistical significance of the out-of-sample portfolio variance and Sharpe-ratio are tested against the $1/N$ benchmark strategy using the circular block bootstrapping method from Ledoit and Wolf (2008, 2011) with 10,000 bootstrap samples and a block with the size of 5. Here, $s(.)$ and p-value denote the standard error and p-value computed from the bootstrap data for a given performance measure. A significance level of $\alpha = 5\%$ is considered. *** denotes significance at 0.1%, ** denotes significance at 1%, and * denotes significance at 5% level. The last five rows report the portfolio weights statistics, averaged across time, specified in Table 4.

	Parametric				**GMV**					**Naive / Passive**	
Jan 1989 - December 2002	PPP(Thr)	PPP(L1)	PPP(FF5)	PPP(FF3)	EFM	LS	NLS	FGL	AT	$1/N$	VW
$\hat{\mu}_p$ (%)	7.02	6.22	7.56	7.14	3.69	6.04	3.33	6.06	4.26	7.63	7.18
$\hat{\sigma}_p$ (%)	11.50***	11.65***	11.63***	11.74***	12.52*	13.04**	11.38***	12.95***	11.77***	14.57	14.28***
$s(\hat{\sigma}_p - \hat{\sigma}_b)$	(0.099)	(0.083)	(0.095)	(0.104)	(0.137)	(0.071)	(0.085)	(0.056)	(0.065)	(-)	(0.031)
$\hat{\sigma}_p$ p-value	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.00	-	0.00
$\widehat{SR}_p$	0.61	0.53	0.65	0.61	0.29	0.46	0.29	0.47	0.36	0.52	0.50
$s(\widehat{SR}_p - \widehat{SR}_b)$	(0.044)	(0.039)	(0.045)	(0.048)	(0.063)	(0.037)	(0.052)	(0.033)	(0.039)	(-)	(0.020)
$\widehat{SR}_p$ p-value	0.64	0.94	0.60	0.66	0.36	0.69	0.28	0.66	0.33	-	0.76
$\overline{TO}$ (%)	8.71	9.46	8.34	7.89	19.11	13.37	19.47	13.15	17.88	6.63	6.75
$100 \times \lvert w_i \rvert$	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
$100 \times \max w_i$	1.17	1.35	1.15	1.06	10.13	2.05	5.96	1.50	5.36	0.20	0.61
$100 \times \min w_i$	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.20	0.04
$\sum w_i I(w_i < 0)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.39	0.44	0.45	0.43	0.91	0.29	0.87	0.25	0.59	0.00	0.00
	Parametric				**GMV**					**Naive / Passive**	
Jan 2003 - December 2016	PPP(Thr)	PPP(L1)	PPP(FF5)	PPP(FF3)	EFM	LS	NLS	FGL	AT	$1/N$	VW
$\hat{\mu}_p$ (%)	9.55	8.70	9.34	9.22	8.30	8.22	7.62	8.56	7.33	10.05	8.78
$\hat{\sigma}_p$ (%)	11.21***	11.45***	11.36***	11.26***	10.57***	11.16***	10.58***	10.90***	11.46***	15.32	14.11***
$s(\hat{\sigma}_p - \hat{\sigma}_b)$	(0.075)	(0.074)	(0.068)	(0.070)	(0.161)	(0.109)	(0.113)	(0.119)	(0.108)	(-)	(0.020)
$\hat{\sigma}_p$ p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	0.00
$\widehat{SR}_p$	0.85	0.76	0.82	0.82	0.78	0.74	0.72	0.78	0.64	0.66	0.62
$s(\widehat{SR}_p - \widehat{SR}_b)$	(0.034)	(0.034)	(0.031)	(0.033)	(0.068)	(0.036)	(0.053)	(0.044)	(0.038)	(-)	(0.011)
$\widehat{SR}_p$ p-value	0.12	0.40	0.13	0.15	0.61	0.57	0.41	0.43	0.91	-	0.37
$\overline{TO}$ (%)	7.88	9.43	7.00	6.75	17.99	13.19	18.22	14.11	15.68	5.90	5.72
$100 \times \lvert w_i \rvert$	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
$100 \times \max w_i$	1.04	1.99	1.08	0.97	8.60	3.70	5.52	3.12	4.14	0.20	0.80
$100 \times \min w_i$	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.20	0.03
$\sum w_i I(w_i < 0)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.41	0.42	0.40	0.39	0.89	0.42	0.86	0.45	0.54	0.00	0.00

7.2 Allocating all stocks

This section examines the out-of-sample results when allocating all common stocks in the investment universe, as done by Brandt et al. (2009), and DeMiguel et al. (2020). As in the previous empirical set-up, penny stocks or stocks with a stock price lower than \$5 are removed[3]. The investment universe now consists of all common stocks in the sample with a stock price greater than \$5 based on the last in-sample date. Here, the investment universe changes every month, so in contrast to the previous results, where the portfolios are held for a year, the portfolios are now updated monthly. To save computation time and since the previous results indicate that the out-of-sample performance did not necessarily improve from exploiting a large set of characteristics, I report the results for the parametric portfolio that exploits five characteristics (PPP(FF5)) and the parametric portfolio that exploits three characteristics (PPP(FF3)) unconstrained and with a short-sale constraint. Table 10 present the results.

Table 10 shows that the results differ from the restricted set of the largest 500 stocks, as the Sharpe ratios of the unconstrained parametric portfolios are now slightly higher than those with short-sale constraints. However, the short-sale constrained parametric portfolios achieve Sharpe ratios that are statistically significantly better than the $1/N$ benchmark strategy, whereas the unconstrained parametric portfolios do not. Nevertheless, the turnover for the unconstrained parametric portfolios is more than three times higher than the turnover for the short-sale constrained parametric portfolios, and both achieve portfolio volatilities significantly lower than the $1/N$ benchmark strategy, with the differences being statistically significant. Overall, the main empirical finding that parametric portfolios can benefit investors when imposing portfolio weight constraints seems robust when allocating all common stocks.

[3]Penny stocks are often highly volatile and lack adequate liquidity. Investing in penny stocks can thus be especially risky and entail higher trading costs.

Table 10: Allocating all stocks

This table shows the out-of-sample performance and distribution of the portfolio weights for the unconstrained and short-sale constrained parametric portfolios (Parametric-C), and benchmark portfolios specified in Table 3. The tables follow the structure of the tables in Brandt, Santa-Clara, and Valkanov (2009). The investment universe consists of all common stocks in the investment universe. The out-of-sample period covers January 1985 to December 2016. The performance measures are presented net of transaction costs. The statistical significance of the out-of-sample portfolio variance and Sharpe-ratio are tested against that of the $1/N$ benchmark strategy using the circular block bootstrap method from Ledoit and Wolf (2008, 2011) with 10,000 bootstrap samples and a block with the size of 5. Here, $s(.)$ and p-value denote the standard error and p-value computed from the bootstrap data for a given performance measure. A significance level of $\alpha = 5\%$ is considered. *** denotes significance at 0.1%, ** denotes significance at 1%, and * denotes significance at 5% level. The last five rows report the portfolio weights statistics, averaged across time, specified in Table 4.

	Parametric		**Parametric-C**		**Naive**	
	PPP(FF5)	PPP(FF3)	PPP(FF5)	PPP(FF3)	$1/N$	VW
$\hat{\mu}_p$ (%)	9.79	8.55	9.16	9.04	8.69	8.64
$\hat{\sigma}_p$ (%)	11.24***	10.49***	11.59***	11.62***	17.46	15.52***
$s(\hat{\sigma}_p - \hat{\sigma}_{bench})$	(0.109)	(0.118)	(0.055)	(0.057)	(-)	(0.034)
$\hat{\sigma}_p$ p-value	0.00	0.00	0.00	0.00	-	0.00
$\widehat{SR}_p$	0.87	0.81	0.79**	0.78**	0.50	0.56
$s(\widehat{SR}_p - \widehat{SR}_{bench})$	(0.062)	(0.063)	(0.029)	(0.028)	(-)	(0.020)
$\widehat{SR}_p$ p-value	0.11	0.15	0.01	0.01	-	0.35
$\overline{TO}$ (%)	37.46	33.18	13.19	12.53	9.41	2.49
$100 \times \lvert w_i \rvert$	0.08	0.07	0.03	0.03	0.03	0.03
$100 \times \max w_i$	0.47	0.46	0.20	0.21	0.03	0.46
$100 \times \min w_i$	-0.38	-0.30	0.00	0.00	0.03	0.00
$\sum w_i I(w_i < 0)$	-0.63	-0.57	0.00	0.00	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.34	0.34	0.43	0.42	0.00	0.00

7.3 Using a different objective function

This section investigates how the results change when optimizing a different objective function. Brandt et al. (2009) propose to optimize a CRRA utility, whereas DeMiguel et al. (2020) propose to optimise a mean-variance utility. I run the empirical analysis for the CRRA and mean-variance utility using a relative risk aversion parameter of five, but this gave extreme outcomes resulting in unsatisfactory out-of-sample performance. For this reason, I consider optimizing a maximum Sharpe ratio portfolio. The results are examined for the restricted set of the largest 500 companies each January and all common stocks in the investment universe. Since the previous results indicate that the out-of-sample performance did not necessarily improve from exploiting a large set of characteristics, I report the results for the parametric portfolio that exploits five characteristics (PPP(FF5)) and the parametric portfolio that exploits three characteristics (PPP(FF3)) unconstrained and with a short-sale constraint to save computational time. Table 11 present the results.

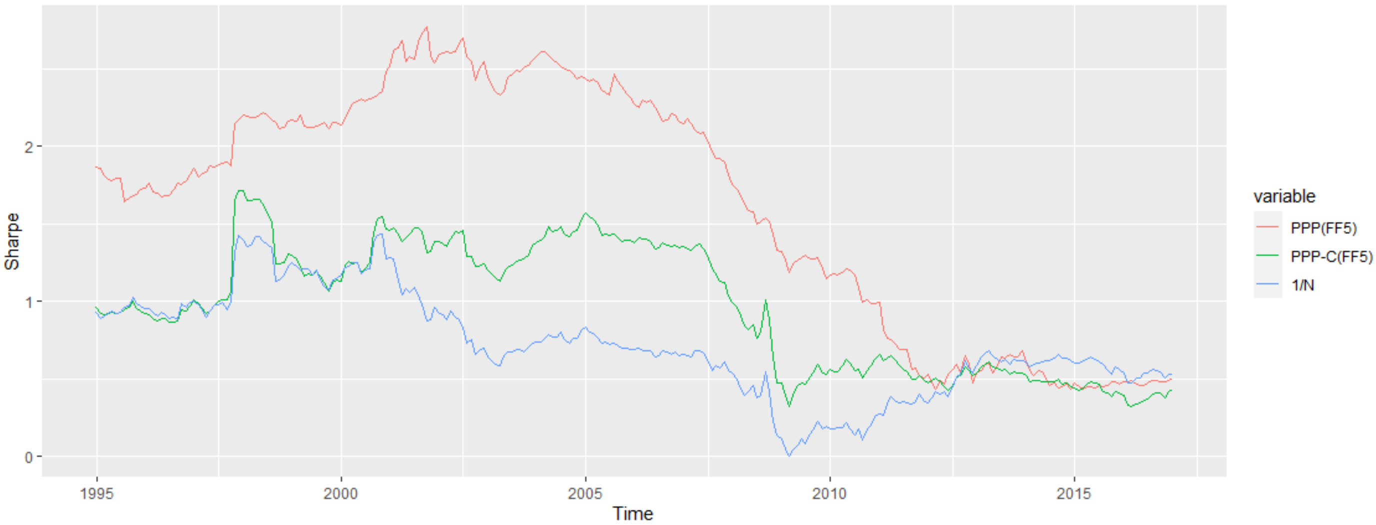
Table 11 shows that for the top 500 stocks, imposing a short-sale constraint on the maximum Sharpe ratio parametric portfolios improves out-of-sample performance as the Sharpe ratios increases and the turnover decreases significantly. This is also what the results for the minimum variance parametric portfolios indicated. Again, the results for all stocks differ for the parametric portfolio that exploits five characteristics (PPP(FF5)) which obtain a notably high out-of-sample Sharpe ratio, with an annualized Sharpe ratio of 1.04. However, the amount of turnover and leverage to realize this performance is substantial. The turnover for PPP(FF5) is 137.22%, and the average sum of negative weights is -279%, which implies an average sum of long positions of 379%, which seems quite extreme and might indicate unstable results over time. Moreover, Table 11 shows that PPP(FF5) and the short-sale constrained version both achieve a Sharpe ratio statistically significantly better than the $1/N$ benchmark strategy. Figure 4 shows the trailing 10-year Sharpe ratio for PPP(FF5), the short-sale constrained version (PPP-C(FF5)), and the $1/N$ benchmark strategy. Figure 4 confirms the results in Chordia et al. (2014) and Green et al. (2017), as the out-of-sample performance decreased substantially over time, resulting in a lower Sharpe ratio for PPP(FF5) and PPP-C(FF5) than the $1/N$ benchmark strategy after approximately 2003. This suggests that investors might be better off optimising a minimum variance portfolio as information regarding expected returns declined and anomalies seem to be priced out in recent years.

Table 11: Maximum Sharpe Ratio

This table shows the out-of-sample performance and distribution of the portfolio weights for the unconstrained and short-sale constrained maximum Sharpe ratio parametric portfolios (Parametric-C) and benchmark portfolios. The tables follow the structure of the tables in Brandt, Santa-Clara, and Valkanov (2009). The results are presented for two investment universes: i) Top 500 stocks consist of a restricted set of the largest 500 companies each January, and ii) All stocks consist of all common stocks in the investment universe. The out-of-sample period covers January 1985 to December 2016. The performance measures are presented net of transaction costs. The statistical significance of the out-of-sample portfolio variance and Sharpe-ratio are tested against the $1/N$ benchmark strategy using the circular block bootstrap method from Ledoit and Wolf (2008, 2011) with 10,000 bootstrap samples and a block with the size of 5. Here, $s(.)$ and p-value denote the standard error and p-value computed from the bootstrap data for a given performance measure. A significance level of $\alpha = 5\%$ is considered. *** denotes significance at 0.1%, ** denotes significance at 1%, and * denotes significance at 5% level. The last five rows report the portfolio weights statistics, averaged across time, specified in Table 4.

	Parametric		**Parametric-C**		**Naive**	
Top 500 stocks	PPP(FF5)	PPP(FF3)	PPP(FF5)	PPP(FF3)	$1/N$	VW
$\hat{\mu}_p$ (%)	11.83	9.32	9.71	9.50	9.24	8.53
$\hat{\sigma}_p$ (%)	20.65***	15.59	14.75*	14.87*	15.48	14.80***
$s(\hat{\sigma}_p - \hat{\sigma}_b)$	(0.096)	(0.054)	(0.043)	(0.042)	(-)	(0.019)
$\hat{\sigma}_p$ p-value	0.00	0.85	0.05	0.05	-	0.00
$\widehat{SR}_p$	0.57	0.60	0.66	0.64	0.60	0.58
$s(\widehat{SR}_p - \widehat{SR}_b)$	(0.050)	(0.027)	(0.014)	(0.014)	(-)	(0.010)
$\widehat{SR}_p$ p-value	0.90	0.96	0.16	0.35	-	0.52
$\overline{TO}$ (%)	61.75	21.81	9.06	9.48	6.19	6.15
$100 \times \lvert w_i \rvert$	1.11	0.47	0.20	0.20	0.20	0.20
$100 \times \max w_i$	3.43	1.37	1.14	1.08	0.20	0.69
$100 \times \min w_i$	-5.33	-1.15	0.00	0.00	0.20	0.04
$\sum w_i I(w_i < 0)$	-2.26	-0.67	0.00	0.00	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.42	0.38	0.47	0.46	0.00	0.00
	Parametric		**Parametric-C**		**Naive**	
All stocks	PPP(FF5)	PPP(FF3)	PPP(FF5)	PPP(FF3)	$1/N$	VW
$\hat{\mu}_p$ (%)	24.04	7.90	9.00	8.33	8.69	8.64
$\hat{\sigma}_p$ (%)	22.07*	15.14	13.69***	13.65***	17.46	15.52***
$s(\hat{\sigma}_p - \hat{\sigma}_b)$	(0.197)	(0.169)	(0.069)	(0.071)	(-)	(0.034)
$\hat{\sigma}_p$ p-value	0.03	0.14	0.00	0.00	-	0.00
$\widehat{SR}_p$	1.09*	0.52	0.66*	0.61	0.50	0.56
$s(\widehat{SR}_p - \widehat{SR}_b)$	(0.067)	(0.054)	(0.023)	(0.027)	(-)	(0.020)
$\widehat{SR}_p$ p-value	0.02	0.90	0.04	0.28	-	0.35
$\overline{TO}$ (%)	137.22	59.29	16.72	16.18	9.41	2.49
$100 \times \lvert w_i \rvert$	0.24	0.13	0.03	0.03	0.03	0.03
$100 \times \max w_i$	1.05	0.55	0.25	0.26	0.03	0.46
$100 \times \min w_i$	-1.70	-0.42	0.00	0.00	0.03	0.00
$\sum w_i I(w_i < 0)$	-2.79	-1.29	0.00	0.00	0.00	0.00
$\sum I(w_i \leq 0)/N$	0.35	0.41	0.47	0.49	0.00	0.00

Figure 4: A plot of the 10-year trailing out-of-sample Sharpe ratio when allocating all common stocks in the investment universe for the maximum Sharpe ratio parametric portfolio that exploits five characteristics (PPP(FF5)), the short-sale constrained maximum Sharpe ratio parametric portfolio that exploits five characteristics (PPP(FF5-C)), and the equally-weighted $1/N$ benchmark strategy.



8 Conclusion

This paper examined the model performance of minimum variance parametric portfolios to determine if investors can still economically benefit from using firm characteristics for large-N portfolio allocation in the presence of transaction costs. Four parametric portfolios are studied: two that exploit only a small number of characteristics typically considered in asset-pricing models, and two that exploit a large set of 39 characteristics with lasso regularization and a new method I propose referred to as K-threshold parametric portfolios. Moreover, the analysis includes traditional econometric approaches by considering global minimum variance (GMV) portfolios with estimation techniques to reduce the errors in the covariance estimate, such as by imposing a factor structure, (non) linear shrinkage, Adaptive Thresholding, and Factor Graphical Lasso. The out-of-sample performance is compared and tested relative to the performance of the equally-weighted $1/N$ benchmark strategy, using the out-of-sample portfolio volatility, Sharpe ratio, and the turnover for each portfolio strategy. In addition, the stability of the portfolio weights is examined. Next, the effect of imposing portfolio weight constraints on the optimal portfolios is examined by considering turnover and short-sale constraints. At last, this paper examined whether the decrease in return predictability of the past decade has any implications on the out-of-sample results.

This paper finds that imposing portfolio weight constraints on parametric portfolios leads to stronger out-of-sample performance; as the Sharpe ratios increase, the portfolio turnover decreases significantly, and the portfolio volatility does not change significantly. The empirical results show that imposing a short-sale constraint and exploiting the smaller set of five characteristics highlighted in Fama and French (2015) leads to the strongest out-of-sample performance. The portfolio volatility for the parametric and GMV portfolios is significantly lower than the $1/N$ benchmark strategy, with the differences being statistically significant, suggesting that the portfolio risk can be significantly reduced. Moreover, the constrained minimum variance parametric portfolios have strong out-of-sample performance in terms of the Sharpe ratio relative to the GMV portfolios and naive diversification. For example, when imposing a short-sale constraint, the results indicate a gain in the annualized Sharpe ratio of 30% for the parametric portfolio that exploits five characteristics, and 21.67% for the parametric portfolio that exploits three characteristics and the K-threshold parametric portfolio. Although the differences in the Sharpe ratio are statistically insignificant, the p-values decreased substantially compared to the unconstrained parametric portfolios, approaching significance. In addition, turnover for the short-sale constrained parametric portfolios ranges from just 7.30% to 8.20%, which is only modestly higher than the turnover for the $1/N$ benchmark strategy of 6.19%. Moreover, the minimum variance parametric portfolios did not involve extreme weights or require high leverage.

In contrast to DeMiguel et al. (2020), using a large set of 39 characteristics did not result in better out-of-sample performance. The regularized parametric portfolio using 39 characteristics with Lasso regularization did not perform well out-of-sample. However, the K-Threshold parametric portfolio, which I propose, did achieve similar performance as the parametric portfolios that exploit only a small number of characteristics.

The empirical findings are consistent post-2003, where Chordia et al. (2014), and Green et al. (2017) find that return predictability fell sharply. The results indicate that when variance minimization is the objective, economic gains can still be realized by exploiting firm characteristics. However, this paper does confirm the finding in Chordia et al. (2014) that the magnitude of asset return predictability has decreased in the last decade, as the out-of-sample performance for the maximum Sharpe ratio parametric portfolios decreased significantly over time. Overall, the results show considerable benefits for investors from exploiting firm characteristics using minimum variance parametric portfolios when imposing portfolio weight constraints such as turnover or short-sale constraints.

However, some possible limitations of this study are worth pointing out. The main limitation is the calculation of the firm characteristics. The data set used in the empirical analysis is the dataset by Gu et al. (2020). The problem is that the firm characteristics could be calculated differently than, for example, in Brandt et al. (2009). In the graphs of the characteristics in Brandt et al. (2009), I noticed significant differences compared to the characteristic data set used in this research. As a result, this makes the results of Brandt et al. (2009) challenging to replicate.

9 Discussion and further research

In this research, I find out to what extent investors still economically benefit from using firm characteristics for large-N portfolio allocation when transaction costs are present. I evaluate minimum variance parametric portfolios and consider a dataset with 39 firm characteristics. I also evaluate the traditional econometric approaches by considering global minimum variance (GMV) portfolios with estimation techniques to reduce the impact of estimation error in the covariance estimate. I compare and test the statistical significance of the out-of-sample performance of the optimal portfolios against the naive equally-weighted $1/N$ strategy, which DeMiguel et al. (2009) find a tough competitor.

Regarding the results, I find that the parametric portfolios achieve better out-of-sample performance than the global minimum variance (GMV) portfolios and the equally-weighted $1/N$ benchmark strategy. This is also what DeMiguel et al. (2009) find as they conclude that using information about the cross-sectional characteristics of assets, rather than just statistical information about the moments of asset returns, does lead to an improvement in Sharpe ratios.

Further, I show that imposing portfolio weight constraints on minimum variance parametric portfolios improves out-of-sample performance net of transaction costs. Brandt et al. (2009) find other results for CRRA investors when allocating all common stocks. From January 1974 till December 2002, the unconstrained parametric portfolio that exploits three characteristics and optimizes a CRRA utility achieved net of transaction costs, an annualized Sharpe ratio of 0.89, and a turnover of 134.10%. With short-sale constraints, the Sharpe ratio decreased to 0.62 and the turnover to 32.40%. In the robustness checks, I provide the out-of-sample results when allocating all common stocks and confirm this finding as the Sharpe ratios are slightly higher for the unconstrained parametric portfolios. DeMiguel et al. (2009) also find that the performance of the parametric portfolio improves when the number of assets available for investment is much larger. Regardless, this comes at the cost of having a much higher turnover. Also, the out-of-sample considered by Brandt et al. (2009) covers January 1974 to December 2002. This includes periods when many anomalies documented in the finance literature were not discovered. At the same time, McLean and Pontiff (2016) find that portfolio returns using firm characteristics are 58% lower post-publication. Moreover, Chordia et al. (2014), and Green et al. (2017), find that return predictability fell sharply after 2003.

Next, I find that investors do not improve out-of-sample performance net of transaction costs by exploiting a larger set of characteristics. The regularized parametric portfolio using 39 characteristics with lasso regularization did not perform well out-of-sample, and the K-threshold parametric portfolio showed approximately similar out-of-sample performance as the parametric portfolios that exploit only a small number of characteristics. On the other hand, DeMiguel et al. (2020) find that for mean-variance investors, the cross-section of stock returns is not fully explained by a small number of characteristics, as the regularized parametric portfolio achieves significantly better out-of-sample performance than the parametric portfolio that exploits only a small number of characteristics. However, the authors confirm the findings in Chordia et al. (2014) that the magnitude of asset return predictability has decreased in the last decade. They find that the number of significant characteristics is smaller from 2003 to 2014. Moreover, in their out-of-sample results for the unconstrained mean-variance parametric portfolios, they find that the regularized parametric portfolio achieved an annualized mean return net of transaction costs of 73.80%, a Sharpe ratio of 1.21 and a monthly turnover of 385.90%, which seems to be quite extreme and might indicate unstable results over time. A turnover constraint did improve out-of-sample performance, which is also what I found for the minimum variance parametric portfolios.

At last, I find that the short-sale constrained minimum variance parametric portfolios achieved significant economic gains from 2003 to 2016. Chordia et al. (2014), and Green et al. (2017) find that characteristics-based predictability fell sharply post-2003. I find that by focusing on variance minimization and omitting expected returns, economic gains can still be realized by exploiting firm characteristics for portfolio allocation. However, in the robustness check, I also evaluate maximum Sharpe ratio parametric portfolios and confirm the finding in Chordia et al. (2014) and Green et al. (2017) that anomalies seem to be priced out in recent years, as the out-of-sample Sharpe ratio decreased substantially in the past decade for the maximum Sharpe ratio parametric portfolios.

For further research, one can try to extend parametric portfolio policy into other asset classes, such as cryptocurrencies. Crypto assets and the vast universe have grown rapidly in recent years, and institutional acceptance and adoption of crypto continue to build. One could investigate if investors can economically benefit from exploiting cryptocurrency characteristics for portfolio allocation using parametric portfolios.

Another further research topic is more of a technical aspect. One could examine different parametric portfolio weight functions. In this research, the portfolio policy is specified as a linear function of the characteristics, similarly done as in Brandt et al. (2009). However, researchers have documented the importance of accounting for non-linearities when studying the relation between stock returns and firm characteristics (Freyberger, Neuhierl, & Weber, 2020; Gu et al., 2020). They demonstrate that accounting for non-linearities results in more precise estimates of the equity risk premium than linear models. Therefore, it could be interesting to investigate whether the parametric portfolio policy can be improved by considering non-linear portfolio weight functions. One could try to directly parameterize the portfolio weights as a non-linear function of firm characteristics.

10 References

Amihud, Y., & Mendelson, H. (1989). The effects of beta, bid-ask spread, residual risk, and size on stock returns. *The journal of Finance*, *44*(2), 479–486.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The journal of finance*, *61*(1), 259–299.

Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, *67*(1), 40–54.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *The journal of Finance*, *43*(2), 507–528.

Bickel, P. J., & Levina, E. (2008). Covariance regularization by thresholding. *The Annals of statistics*, *36*(6), 2577–2604.

Brandt, M. W., Santa-Clara, P., & Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, *22*(9), 3411–3447.

Cai, T., & Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, *494*(106), 672–684.

Chan, L. K., Karceski, J., & Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The review of Financial studies*, *12*(5), 937–974.

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

Chopra, V. K. (1993). Improving optimization. *The Journal of Investing*, *2*(3), 51–59.

Chordia, T., Subrahmanyam, A., & Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, *58*(1), 41–58.

Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *the Journal of Finance*, *63*(4), 1609–1651.

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.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J., & Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, *33*(5), 2180–2222.

Domowitz, I., Glen, J., & Madhavan, A. (2001). Liquidity, volatility and equity trading costs across countries and over time. *International Finance*, *4*(2), 221–255.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1–22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, *81*(3), 607–636.

Fan, J., Liao, Y., & Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of statistics*, *39*(6), 3320.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of financial economics*, *111*(1), 1–25.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). "dissecting characteristics nonparametrically.". *The Review of Financial Studies*, *33*(5), 2326–2377.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Frost, P. A., & Savarino, J. E. (1988). For better performance: Constrain portfolio weights. *Journal of portfolio management*, *15*(1), 29.

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, *30*(12), 4389–4436.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

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

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, *64*(3), 1445–1477.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice, chapter 5. *OTexts*.

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.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, 259–278.

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

Keim, D. B., & Madhavan, A. (1997). Transactions costs and investment style: An inter-exchange analysis of institutional equity trades. *Journal of Financial Economics*, *46*(3), 265–292.

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

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

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

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. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, (55), 86–89.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 1024–1060.

Ledoit, O., & Wolf, M. (2017). Direct nonlinear shrinkage estimation of large-dimensional covariance matrices. *(Working paper)*, (264).

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

Litzenberger, R. H., & Ramaswamy, K. (1982). The effects of dividends on common stock prices tax effects or information effects? *The journal of Finance*, *37*(2), 429–443.

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

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Business and Economic Statistics*, *71*(1), 5–32.

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

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

Scherer, B. (2011). A note on the returns from minimum variance investing. *Journal of Empirical Finance*, *18*(4), 652–660.

Appendix

A Characteristics considered

Table 1: List of characteristics considered

No.	Acronym	Firm characteristic	Paper's author(s)	Year, Journal	Data Source	Frequency
1	age	# years since first Compustat coverage	CJiang, Lee & Zhang	2005, RAS	Compustat	Annual
2	agr	Asset growth	Cooper, Gulen & Schill	2008, JF	Compustat	Annual
3	baspread	Bid-ask spread	Amihud & Mendelson	1989, JF	CRSP	Monthly
4	beta	Beta	Fama & MacBeth	1973, JPE	CRSP	Monthly
5	betasq	Beta squared	Fama & MacBeth	1973, JPE	CRSP	Monthly
6	bm	Book-to-market	Rosenberg, Reid & Lanstein	1985, JPM	Compustat + CRSP	Annual
7	bm_ia	Industry-adjusted book to market	Asness, Porter & Stevens	2000, WP	Compustat + CRSP	Annual
8	cashpr	Cash productivity	Chandrashekar & Rao	2009, WP	Compustat	Annual
9	chcsho	Change in shares outstanding	Pontiff & Woodgate	2008, JF	Compustat	Annual
10	chempia	Industry-adjusted change in employees	Asness, Porter & Stevens	1994, WP	Compustat	Annual
11	chmom	Change in 6-month momentum	Gettleman & Marks	2006, WP	CRSP	Monthly
12	convind	Convertible debt indicator	Valta	2016, JFQA	Compustat	Annual
13	currat	Current ratio	Ou & Penman	1989, JAE	Compustat	Annual
14	dy	Dividend to price	Litzenberger & Ramaswamy	1982, JF	Compustat	Annual
15	egr	Growth in common shareholder equity	Richardson, Sloan, Soliman & Tuna	2005, JAE	Compustat	Annual
16	ep	Earnings to price	Basu	1977, JF	Compustat	Annual
17	gma	Gross profitability	Novy-Marx	2013, JFE	Compustat	Annual
18	herf	Industry sales concentration	Hou & Robinson	2006, JF	Compustat	Annual
19	hire	Employee growth rate	Bazdresch, Belo & Lin	2014, JPE	Compustat	Annual
20	idiovol	Idiosyncratic return volatility	Ali, Hwang & Trombley	2003, JFE	CRSP	Monthly
21	indmom	Industry momentum	Moskowitz & Grinblatt	1999, JF	CRSP	Monthly
22	lev	Leverage	Bhandari	1988, JF	Compustat	Annual
23	lgr	Growth in long-term debt	Richardson, Sloan, Soliman & Tuna	2005, JAE	Compustat	Annual
24	maxret	Maximum daily return	Bali, Cakici & Whitelaw	2011, JFE	CRSP	Monthly
25	me	Natural log of market capitalization	Banz	1981, JFE	CRSP	Monthly
26	mom12m	12-month momentum	Jegadeesh	1990, JF	CRSP	Monthly
27	mom1m	1-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly
28	mom6m	6-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly
29	mve_ia	Industry-adjusted size	Asness, Porter & Stevens	2000, WP	Compustat	Annual
30	operprof	Operating profitability	Fama & French	2015, JFE	Compustat	Annual
31	pchgm_pchsale	% change in gross margin - % change in sales	Abarbanell & Bushe	1998, TAR	Compustat	Annual
32	pricedelay	Price delay	Hou & Moskowitz	2005, RFS	CRSP	Monthly
33	ps	Financial statements score	Piotroski	2000, JAR	Compustat	Annual
34	rd	R&D increase	Eberhart, Maxwell & Siddique	2004, JF	Compustat	Annual
35	retvol	Return volatility	Ang, Hodrick, Xing & Zhang	2006, JF	CRSP	Monthly
36	salecash	Sales to cash	Ou & Penman	1989, JAE	Compustat	Annual
37	salesrec	Sales to receivables	Ou & Penman	1989, JAE	Compustat	Annual
38	sgr	Sales growth	Lakonishok, Shleifer & Vishny	1994, JF	Compustat	Annual
39	sp	Sales to price	Barbee, Mukherji, & Raines	1996, FAJ	Compustat	Annual

B Summary statistics characteristics

Table 2: Summary statistics for the (non-standardized) characteristics

	Mean	St Dev	Min	Max	Skewness	Kurtosis
age	12.53	3.79	4.90	20.00	0.29	-0.43
agr	-0.16	0.05	-0.34	-0.02	-0.20	0.54
baspread	0.04	0.01	0.02	0.11	2.18	8.18
beta	1.07	0.08	0.72	1.24	-1.48	3.60
betasq	1.52	0.22	0.81	1.97	-0.90	0.59
bm	0.78	0.23	0.47	1.76	1.60	3.43
bm_ia	-0.02	0.10	-0.39	0.23	-1.11	2.43
cashpr	-1.56	5.53	-14.26	12.58	-0.16	-0.54
chcsho	0.11	0.05	0.03	0.25	0.54	-0.28
chempia	-0.05	0.07	-0.62	0.01	-5.87	44.80
chmom	0.00	0.25	-0.78	1.24	0.65	1.70
chpmia	0.12	0.97	-3.04	3.00	0.64	2.44
convind	0.16	0.07	0.08	0.32	0.91	0.08
dy	0.02	0.01	0.01	0.06	0.99	0.13
egr	0.16	0.06	-0.04	0.37	0.11	1.37
ep	0.03	0.07	-0.25	0.18	-0.90	3.51
gma	0.35	0.05	0.25	0.44	0.05	-1.03
herf	0.09	0.01	0.06	0.11	0.08	-1.41
hire	0.09	0.04	-0.01	0.19	0.25	-0.06
idiovol	0.05	0.01	0.03	0.09	1.00	0.65
indmom	0.13	0.24	-0.47	1.13	0.60	1.34
lev	2.26	0.79	0.58	4.80	0.64	0.22
lgr	0.22	0.08	0.04	0.44	0.38	-0.07
maxret	0.06	0.02	0.03	0.17	1.97	5.81
me	12.32	0.64	11.35	13.69	0.59	-0.84
mom12m	0.16	0.23	-0.44	1.17	0.58	1.30
mom1m	0.01	0.06	-0.28	0.26	-0.33	2.35
mom6m	0.07	0.15	-0.41	0.68	0.32	1.10
mve_ia	-350.83	467.87	-1383.52	90.63	-0.75	-1.08
operprof	0.19	0.07	0.09	0.33	0.72	-0.79
pchgm_pchsale	-0.04	0.04	-0.17	0.04	-0.51	0.34
pricedelay	0.10	0.06	-0.47	0.27	-1.26	12.44
ps	3.93	0.64	2.81	5.04	-0.23	-1.31
rd	0.11	0.04	0.01	0.19	-0.66	0.85
retvol	0.03	0.01	0.02	0.08	2.09	6.84
salecash	46.37	16.85	23.47	81.01	0.45	-1.11
salerec	11.22	1.13	6.14	14.08	-1.14	6.42
sgr	0.17	0.07	-0.02	0.37	0.03	0.25
sp	2.00	0.92	0.92	5.47	1.29	1.50

C Elaboration on Analytical Non-Linear Shrinkage

In line with Ledoit and Wolf (2020), the spectral density with the Epanechnikov kernel used in Equation (7) is given by:

$$\tilde{f}(\lambda_i) := \frac{1}{N}\sum_{j=1}^{N}\frac{3}{4\sqrt{5}h_j}[1-\frac{1}{5}\left(\frac{\lambda_i-\lambda_j}{h_j}\right)^2]^+$$

The Hilbert transform is defined as:

$$H_{\tilde{f}}(\lambda_i) := \frac{1}{N}\sum_{j=1}^{N}\left\{\frac{-3(\lambda_i-\lambda_j)}{10\pi h_j^2}+\frac{3}{4\sqrt{5}\pi h_j}[1-\frac{1}{5}\left(\frac{\lambda_i-\lambda_{t,j}}{h_j}\right)^2]\times\log|\frac{\sqrt{5}h_j-\lambda_i+\lambda_j}{\sqrt{5}h_j+\lambda_i-\lambda_j}\right\}$$

D Robust hypothesis testing by Ledoit and Wolf (2008, 2011)

We observe T pairs of returns, $(r_{p,t}, r_{b,t})', ..., (r_{p,T}, r_{b,T})'$, with a bivariate return distribution over time:

$$\mu = \begin{bmatrix}\mu_p \\ \mu_b\end{bmatrix} \qquad \Sigma = \begin{bmatrix}\sigma_p^2 & \sigma_{p,b} \\ \sigma_{b,p} & \sigma_b^2\end{bmatrix}$$

We do not assume the distribution to be normal, nor do we assume that returns are independent over time. The parameter of interest is:

$$\Delta = \xi_p - \xi_b \tag{22}$$

where ξ is a given performance measure. Hence, ξ_p is the performance measure for the strategy p and ξ_b is the performance measure for the benchmark strategy b. Its estimator using sample moments is given by:

$$\hat{\Delta} = \hat{\xi}_p - \hat{\xi}_b \tag{23}$$

We are interested in testing:

$$H_0 : \Delta = 0 \quad \text{vs} \quad H_1 : \Delta \neq 0$$

In order to test this, we need some notation first. Denote $E(r_p^2) = \mu_{2,p}$ and $E(r_b^2) = \mu_{2,b}$ as the second moments, with their corresponding estimates $\hat{\mu}_{2p}$ and $\hat{\mu}_{2b}$. We define $\eta = (\mu_p, \mu_b, \mu_{2p}, \mu_{2b})'$ and $\hat{\eta} = (\hat{\mu}_p, \hat{\mu}_b, \hat{\mu}_{2p}, \hat{\mu}_{2b})'$ such that $\Delta = f(\eta)$ and $\hat{\Delta} = f(\hat{\eta})$ with:

$$\Delta := f(\eta) = \xi_p - \xi_b \tag{24}$$

and the estimator of Δ given by:

$$\hat{\Delta} := f(\hat{\eta}) = \hat{\xi}_p - \hat{\xi}_b \tag{25}$$

We assume that:

$$T^{1/2}(\hat{\eta} - \eta) \xrightarrow{d} N(0, \Psi) \tag{26}$$

where Ψ is an unknown symmetric positive semidefinite matrix. This relation holds under mild regularity conditions. If we apply a function on the vector η of parameters, the Taylor expansion (Delta method) implies:

$$T^{1/2}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0, \nabla' f(\eta)\Psi\nabla f(\eta)) \tag{27}$$

where $\nabla f(.)$ is the gradient of $f(.)$. Therefore, if a consistent estimator $\hat{\Psi}$ of Ψ is available, then an asymptotic standard error of for $\hat{\Delta}$, $s(\hat{\Delta})$, is given by:

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{\eta})\hat{\Psi}\nabla f(\hat{\eta})}{T}} \tag{28}$$

Bootstrap Inference

The two-sided distribution function of the studentized statistic is approximated via the bootstrap as follows:

$$\mathcal{L}\left(\frac{|\hat{\Delta} - \Delta|}{s(\hat{\Delta})}\right) \approx \mathcal{L}\left(\frac{|\hat{\Delta}^{*} - \hat{\Delta}|}{s(\hat{\Delta}^{*})}\right) \tag{29}$$

where Δ is the true difference in the performance measure, $\hat{\Delta}$ is the estimated difference from the original data, $\hat{\Delta}^{*}$ is the estimated difference from the bootstrap data. $\mathcal{L}(X)$ is the distribution function of random variable X. As recommended by Ledoit and Wolf (2008, 2011), the Circular Boostrap of Politis and Romano (1992) is used, resampling now blocks of pairs from the observed pairs $(r_{p,t}, r_{b,t})'$, $t = 1, ..., T$ with replacement. These blocks have a fixed size. The standard error is computed as:

$$se(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{\eta})\hat{\Psi}\nabla f(\hat{\eta})}{T}} \tag{30}$$

The estimator $\hat{\Psi}$ is obtained via kernel estimation. Standard error $se(\hat{\Delta}^{*})$ is the "natural" standard error computed from the bootstrap data.

$$se(\hat{\Delta}^{*}) = \sqrt{\frac{\nabla' f(\hat{\eta}^{*})\hat{\Psi}^{*}\nabla f(\hat{\eta}^{*})}{T}} \tag{31}$$

Denote the original studentized test statistic by

$$y = \frac{|\hat{\Delta}|}{s(\hat{\Delta})} \tag{32}$$

and denote the centered studentized test statistic computed from the mth bootstrap sample by

$$\tilde{y}^{*,m} = \frac{|\hat{\Delta}^{*,m} - \hat{\Delta}|}{s(\hat{\Delta}^{*,m})} \tag{33}$$

where M is the number of bootstrap resamples. Then the p-value is computed as

$$\hat{p}^* = \frac{\#\{\tilde{y}^{*,m} \geq d\} + 1}{M + 1} \tag{34}$$

Sharpe Ratio testing

Define the difference in Sharpe Ratios as a function of the vector of primitive statistics:

$$\Delta = f(\eta) = SR_p - SR_b = \frac{\mu_p}{\sqrt{\mu_{2,p} - \mu_p^2}} - \frac{\mu_b}{\sqrt{\mu_{2,b} - \mu_b^2}} \tag{35}$$

The gradient of $f(\eta)$ is:

$$\nabla' f(\eta) = \left(\frac{\partial f(\eta)}{\partial \mu_p}, \frac{\partial f(\eta)}{\partial \mu_b}, \frac{\partial f(\eta)}{\partial \mu_{2,p}}, \frac{\partial f(\eta)}{\partial \mu_{2,b}} \right) \tag{36}$$

$$= \left(\frac{\mu_{2,p}}{(\mu_{2,p} - \mu_p^2)^{3/2}}, \frac{\mu_{2,b}}{(\mu_{2,b} - \mu_b^2)^{3/2}}, \frac{-\mu_p}{2(\mu_{2,p} - \mu_p^2)^{3/2}}, \frac{\mu_b}{2(\mu_{2,b} - \mu_b^2)^{3/2}} \right) \tag{37}$$

And the standard error of $\hat{\Delta}$, $se(\hat{\Delta})$, is given by:

$$se(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{\eta}) \hat{\Psi} \nabla f(\hat{\eta})}{T}} \tag{38}$$

where $\hat{\Psi}$ is a consistent estimator for Ψ.

Variance Testing

Define the difference in Sharpe Ratios as a function of the vector of primitive statistics:

$$\Delta = \log(\sigma_p^2) - \log(\sigma_b^2) = \log(\sigma_p^2 / \sigma_b^2) \tag{39}$$

Defining Δ as a function of the first two uncentered moments, $\eta = (\mu_p, \mu_b, \mu_{2,p}, \mu_{2,b})'$:

$$\Delta = f(\eta) = \log(\mu_{2,p} - \mu_p^2) - \log(\mu_{2,b} - \mu_b^2) = \log(\tfrac{\mu_{2,p} - \mu_p^2}{\mu_{2,b} - \mu_b^2})$$

The gradient of the function $f(\eta)$ is:

$$\nabla' f(\eta) = \left(\frac{\partial f(\eta)}{\partial \mu_p}, \frac{\partial f(\eta)}{\partial \mu_b}, \frac{\partial f(\eta)}{\partial \mu_{2,p}}, \frac{\partial f(\eta)}{\partial \mu_{2,b}} \right) \tag{40}$$

$$= \left(\frac{-2\mu_p}{\mu_{2,p} - \mu_p^2}, \frac{2\mu_b}{\mu_{2,b} - \mu_b^2}, \frac{1}{\mu_{2,p} - \mu_p^2}, \frac{1}{\mu_{2,b} - \mu_b^2} \right) \tag{41}$$

And the standard error of $\hat{\Delta}$, $se(\hat{\Delta})$, is given by:

$$se(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{\eta})\hat{\Psi}\nabla f(\hat{\eta})}{T}} \tag{42}$$

where $\hat{\Psi}$ is a consistent estimator for Ψ.