Optimal portfolios based on stock characteristics: Construction and Performance for the US Market

Nádia Raquel de Araújo Oliveira

UMinho | 2022

Universidade do Minho
Escola de Economia e Gestão

Nádia Raquel de Araújo Oliveira

Optimal portfolios based on stock characteristics: Construction and Performance for the US Market

outubro de 2022

Universidade do Minho
Escola de Economia e Gestão

Nádia Raquel de Araújo Oliveira

Optimal portfolios based on stock characteristics: Construction and Performance for the US Market

Dissertação de Mestrado
Mestrado em Finanças

Trabalho efetuado sob a orientação do
Professor Doutor Nelson Areal

outubro de 2022

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Acknowledgements

I would like to thank the following people, without whom I would not have been able to complete this dissertation.

First of all, I would like to express my sincere gratitude to my supervisor, Professor Doutor Nelson Areal, for his suggestion of the topic, time, patience and encouragement. His guidance and feedback were vital for the development of this research.

I would also like to thank my family, especially my mother, for the support they gave me and for always making me feel confident in my abilities.

Finally, I would like to thank my work colleagues, who were always supporting and interested to know the conclusions of this study.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Carteiras ótimas baseadas em características de ações: Construção e performance no mercado dos E.U.A

Resumo

Existe uma vasta literatura sobre criação de carteiras de investimentos, iniciada por Markowitz (1952) e prolongada ao longo dos anos com contribuições sobre novas abordagens e técnicas. Outro ramo relevante da literatura em Finanças concentra-se na análise fundamental e na investigação de métodos para prever rendibilidades futuras. Esta dissertação situa-se na interseção destes dois ramos, uma vez que tem como principal objetivo incorporar a análise fundamental na otimização de carteiras.

Para atingir este objetivo, este estudo foca-se no mercado de ações dos EUA entre 1998 e 2021. Os dados são provenientes da Refinitiv Datastream e filtrados e processados seguindo as orientações de Landis e Skouras (2021). Para a criação das carteiras, as rendibilidades esperadas e as matrizes de covariância são produzidas com a estimação de um modelo que relaciona diretamente as rendibilidades esperadas com características das ações, como rácios contabilísticos. Para compreender se possíveis ganhos de performance advêm das rendibilidades esperadas, das matrizes de covariância, ou de ambas, carteiras com diferentes pressupostos são construídas e comparadas com as "Carteiras fundamentais totalmente otimizadas" (*Fully optimized fundamental portfolios*) através de várias medidas de performance. Estas são construídas seguindo Markowitz (1952), incorporando tanto as rendibilidades esperadas como as estimativas da matriz de covariância. Os resultados são comparados com os encontrados no estudo de Lyle e Yohn (2021).

As principais conclusões desta dissertação são de que a incorporação de estimativas de rendibilidades esperadas com base nas características das ações e matrizes de covariância na otimização pode traduzir-se em ganhos na performance da carteira. No entanto, estes ganhos são, em quase todos os casos, apenas notórios quando há uma restrição de pesos superiores a zero (sem vendas a descoberto). Estes resultados são também dependentes do período de tempo em análise e da inclusão de ações de baixa capitalização no universo de investimentos.

Palavras-chave: Análise de performance, Análise fundamental, Construção de carteiras, Mercado de ações, Rendibilidades

Optimal portfolios based on stock characteristics: Construction and Performance for the US Market

Abstract

There is a vast literature on investment portfolio creation, starting with Markowitz (1952) and prolonging itself throughout the years with numerous contributions on new approaches and improved techniques. Another relevant branch of Finance literature focuses on fundamental analysis and seeks to find methods to predict future returns. This dissertation lies in the intersection between these two branches, as its main objective is to incorporate fundamental analysis in mean-variance portfolio optimization.

In order to achieve this objective, this study focuses on the US stock market between 1998 and 2021. The data is retrieved from Refinitiv Datastream and cleaned and processed following Landis and Skouras (2021) guidelines. For the creation of the portfolios, both expected returns and covariance matrices are produced with the estimation of a model that directly links expected returns with stock characteristics, including accounting ratios. To understand whether possible gains in performance arise from the expected returns, covariance matrixes, or both, several portfolios with different assumptions are constructed and compared to the "Fully optimized fundamental portfolios" with several performance measures. These Fully optimized portfolios are constructed in Markowitz's (1952) style, incorporating both the expected returns and covariance matrix estimations. Results are compared with the ones found in Lyle and Yohn (2021) study.

The main conclusions of this dissertation are that incorporating estimates of expected returns based on stock characteristics and covariance matrices in mean-variance optimization can produce an improved portfolio performance. However, this improvement is almost in all cases only noticeable when there is a constraint on weights to be higher than zero (no short selling). These results are also dependent on the time period under analysis and the inclusion of small stocks in the investment set.

Keywords: Performance analysis, Fundamental analysis, Portfolio construction, Stock market, Returns

List of Contents

List of tables

List of Figures

1. Introduction

There is a vast literature on investment portfolio creation. Constructing a portfolio generally consists of estimating expected future stock returns and a covariance matrix and optimally allocating wealth across those stocks to maximize the expected return subject to a given risk constraint (Lyle & Yohn, 2021). The first contributor to the topic was Markowitz (1952) with its mean-variance portfolio approach. Since then, multiple alternatives to portfolio creation have been discussed and analyzed in the Finance literature.

The ability to predict market returns is important in portfolio allocation decisions (Lee et al., 1999). According to Neely & Rapach (2010), there are two main methods of predicting aggregate stock returns: (1) Fundamental analysis and (2) Technical analysis. Fundamental analysis uses valuation ratios, interest rates, term and credit spreads, and other economic variables to forecast excess stock returns while technical analysis relies on past stock price behavior and trade volume to determine future price movements (Neely et al., 2010). Throughout this dissertation, a more specific definition of fundamental analysis will be used: the use of fundamentals-based ratios to estimate a stock's intrinsic value (Lyle & Yohn, 2021).

The main objective of this dissertation is to incorporate fundamental analysis in mean-variance portfolio optimization. To do so, data on the United States stock market between July 1998 and September 2021 is used. After the creation of the "Fundamental Portfolios", their ex-post performance is compared with other standard portfolio construction alternatives, namely Equal-Weighted, Minimum Variance, and Expected Returns portfolios. Several performance measures are used, including more traditional ones such as the Sharpe and Treynor Ratios, and also alphas and information ratios based on multifactor models.

Accordingly, the main research question is finding if there is any improvement in portfolio performance for considering expected returns based on fundamentals for mean-variance portfolio optimization. The results are also compared with previous literature with the same methodology and approach, namely the recent study of Lyle and Yohn (2021).

According to Baker and Haugen (1996), there is a common bias in Finance literature related to data snooping. This occurs when researchers examine the properties of a database or the results of other studies of that database, build predictive models based on the previous results, and then test the power of the models on the same database. A relevant feature of this dissertation is that this direct comparability of the results with Lyle and Yohn (2021) study can also help to understand possible differences in the databases used for both studies. The data for this dissertation is retrieved from Refenitiv Eikon Datastream whereas the original study and a large part of the research for the US market are based on data from CRSP and Compustat. Another relevant feature is the analysis of results for different time periods from the ones in the original study, which only uses data up to 2017. Finally, I also consider some possible technical problems with the original study and innovate by providing solutions for them.

The remaining of this Dissertation is organized as follows: Section 2 is the Literature Review, which brings into light the state of the art concerning fundamental analysis, portfolio construction, and performance measurement; Section 3 presents the methods used to develop the study; Section 4 explains the data that will be collected; Section 5 presents the main results found; and, finally, Section 5 concludes the dissertation.

2. Literature Review

2.1 Return predictability and fundamental analysis

The literature on return predictability is quite extensive. Some academic finance economists believe that aggregate stock returns are predictable (Campbell & Thompson, 2008) and others argue that market efficiency implies that stock returns are random, and so, cannot be predicted.

Allen et al. (2019) assert that the partial predictability of returns through the use of publicly available information challenges the idea of pure market efficiency. They assume that some forecasting ability seems reasonable. However, the forecasts are noisy and the estimation error must be considered.

Campbell and Thompson (2008) argue that several variables are correlated with subsequent stock returns, including stock market valuation ratios, levels of short- and long-term interest rates, patterns in corporate finance and cross-sectional pricing of individual stocks, and the level of consumption to wealth.

Also, Lyle and Yohn (2021) point to the predictive ability of fundamentals-based ratios, such as book-to-market and profitability. Brandt et al. (2009) argue that stock characteristics, such as market capitalization, book-to-market ratio, or lagged returns, are related to the expected return, variance, and covariance with other stocks.

Fundamental analysis uses valuation ratios, interest rates, term and credit spreads, and other economic variables to forecast excess stock returns (Neely et al., 2010). According to Lyle and Yohn (2021), Fundamental analysis can be used to derive an "intrinsic value" estimate of a stock, assuming that the stock's market value will converge to its intrinsic value over time. Also Lee et al. (1999) find that when the price is a noisy measure of value, fundamental measures can be evaluated in terms of their ability to contribute to return prediction: an investor can use the difference between the intrinsic value and the market value to estimate future stock returns.

One common ratio studied in the literature is the book-to-market value ratio. According to Fama and French (2006), a higher book-to-market value ratio implies a higher expected stock return. Firms with higher book-to-market ratios (value firms) grow less rapidly than low book-to-market ratio firms (growth firms). The authors find that, when size and the book-to-market ratio are used alone to explain returns, there is a strong positive relation between average return and book-to-market ratios and, so, assert that high book-to-market firms have higher average returns than low book-to-market firms.

Baker and Haugen (1996) provide a review of possible explanations across the literature for this phenomenon. Some authors believe that value stocks are riskier, and so it is expected for these stocks to have a premium. Other authors believe that the premium returns to value stocks are unexpected by investors. Investors overreact to firms' history of success and failure. By projecting prolonged and rapid growth, investors in growth stocks can drive prices too high. As the market competition forces that drive profits to normal levels come into play sooner than expected, future dividends and capital gains on these stocks tend to be smaller than expected and returns tend to be relatively low. The opposite tends to occur for value stocks. An important conclusion of the authors is that, regardless of whether the relationship is a consequence of risk or overreaction, a higher book-to-market value ratio is associated with a higher expected stock return.

Also according to Baker and Haugen (1996), more profitable firms will tend to grow faster, at least until market competition forces profits to normal levels. Regarding the association between market capitalization and average returns, Fama and French (2006) find that small firms have higher average returns than big firms.

A wide literature originating from Sloan's (1996) study demonstrates a negative relation between accruals and future profitability. Sloan (1996) hypothesizes that investors do not understand that earnings are composed of both operating cash flows and accruals, and that accruals tend to reverse in future periods. This faster mean reversion of the accruals part of earnings leads to a negative relationship between current accruals and future stock returns.

Many papers extended this original hypothesis and the research has generally found that the accrual component of earnings has a negative association with future returns (Richardson et al., 2010). However, the fact that accruals do not mean revert faster than the cash component of earnings suggests that Sloan's (1996) inference cannot in itself explain large spreads in average returns associated with accruals (Fama & French, 2006). More recent papers attribute this relation to a combination of earnings management and accounting distortions. These papers tend to show that the negative relation between accruals and future returns is stronger for accruals reflecting earnings management or accounting distortions (Richardson et al., 2010).

Sloan (1996) uses changes in balance sheet items and defines accruals as changes in non-cash working capital minus depreciation scaled by average total assets. This is the narrow definition of accruals. A more general definition of accruals is introduced by Richardson et al. (2005) – the change in net operating assets. They separate operating from financing activities and reorganize the standard balance sheet identity of assets equal to liabilities plus book value of equity: assets can be divided into operating assets (operating component) and financial assets (financing component), as well as liabilities can be operating liabilities or financial liabilities. Thus, net operating assets (operating assets less operating liabilities) are equal to net financial obligations (short-term debt plus long debt less financial assets) plus the book value of equity.

The change in net operating assets includes not only the current accruals of the original Sloan's (1996) measure such as changes in inventory, accounts receivables, and accounts payable, but also the noncurrent accruals such as intangibles, property, plant and equipment, and deferred employment obligations. The main differences are that Sloan’s measure ignores non-current accruals, excludes taxes payable, and treats depreciation expense as a current accrual (Richardson et al., 2010).

It is relevant to note that this dissertation has not the intent to give an insight into whether or not the predictability of returns and the relationship of each of these ratios with future returns is a challenge to market efficiency. According to Baker and Haugen (1996), there are three main interpretations for these findings: (1) some believe that the evidence is derived from biases in the studies (for instance survival bias and data snooping), others believe that, even with possible exaggeration through flaws in estimation, the relationships are still observed and (2) are expected by investors and caused by differences in risk premiums, that although not explained by the Capital Asset Pricing Model, can be consistent with other models, and so the market efficiency is not rejected or (3) are not expected by investors and derive from the market over or underreactions to numerous events.

2.2 Portfolio construction approaches

Markowitz’s influential articles on portfolio selection support many advances in financial literature. Its approach is one of the most widely used quantitative methods of portfolio construction by practitioners (Allen et al., 2019). However, this traditional approach of first modeling the distribution of returns and then solving for the corresponding optimal portfolio weights is difficult to implement for a large number of assets and produces notoriously noisy and unstable results (Brandt et al., 2009).

In the past, investors sought to maximize wealth while minimizing the volatility of returns (Bender et al., 2019). However, according to Bender et al. (2019), new portfolio construction ideas have proliferated rapidly in the last few years. Now, more than having the objective of identifying robust systematic and repeatable sources of return, investors should refocus on isolating sources of return that are rewarded and build portfolios that capture these sources in the most risk-efficient and cost-efficient way (Bender et al., 2019).

2.2.1 Markowitz (1952) mean-variance optimal portfolios

According to Markowitz (1952), the process of selecting a portfolio may be divided into two stages: (1) the analysis of the future performance of available securities and (2) the choice of the portfolio. The first phase should involve estimates of individual stock's expected returns, variances, and covariances (Brandt et al., 2009).

Markowitz (1952) discusses that the rule for portfolio choice should not be the maximization of discounted expected returns, but the investor should both diversify and maximize expected returns instead. He refers to the latter as the "expected returns-variance of returns" rule (Markowitz, 1952). The portfolio weights should be chosen to maximize the expected portfolio return subject to a given risk constraint (Lyle & Yohn, 2021). This implies that the portfolio with the maximum expected return is not necessarily the one with minimum variance and there is a trade-off between expected return and risk (Lyle & Yohn, 2021; Markowitz, 1952).

According to Lyle and Yohn (2021), although this method is intuitive and one of the most prevalent models for capturing investor preferences (Bender et al., 2019), it has some issues: it is difficult to implement because of unreliable expected returns estimates and there can be significant errors in the estimation of the covariance matrix. Allen et al. (2019) point out a consensus in the literature that estimation error makes mean-variance portfolio strategies inferior to passive equal-weighted approaches.

It is important to note, however, that errors in estimation can be addressed and sensible parameters can be calibrated, solving the issues concerning the mean-variance approach (Bender et al., 2019). Brandt et al. (2009) point to different solutions that, although generally improve the properties of the optimized portfolio, require substantial resources. These include the shrinkage of estimates, imposing a factor structure on the covariance matrix, the estimation of expected returns from an asset pricing model, and constraining portfolio weights.

2.2.2 Alternative portfolio construction approaches

Taking into account the potential issues arising from mean-variance optimization, there are several portfolio construction approaches referenced throughout the literature. When it comes to weighted portfolios, they are constructed using some set of rules focusing on diversification or risk (Bender et al., 2019). There are several definitions and categorizations of portfolio construction approaches.

For instance, Amenc et al. (2014) create two broad definitions: (1) scientific diversification, which includes, for example, minimum variance portfolios and maximum Sharpe ratio portfolios, and (2) naive diversification, including equal-weighted portfolios and equal risk contribution portfolios.

Also, Bender et al. (2019) divide portfolio approaches into two types: (1) optimization-based weighting schemes and (2) heuristic-based weighting schemes. Optimization-based weighting schemes include minimum variance portfolios, maximum decorrelation portfolios, maximum diversification portfolios, risk parity portfolios, maximum deconcentration portfolios, and maximum Sharpe ratio portfolios. According to the same author, heuristic-based weighting schemes include fundamental indexing, equal weighting, and diversity weighting.

Equal-weighted portfolios do not require any estimates of returns, volatility, or correlations, thus not having the estimation problem. Moreover, they are easy to understand (Bender et al., 2019). Naive equal-weighted portfolios assume that both expected stock returns and covariances are cross-sectionally constant and, therefore, that it is optimal to invest equally in each available stock (Lyle & Yohn, 2021).

With minimum variance portfolios, weights are selected to minimize the variance of the portfolio, irrespective of the expected rate of return (Baker & Haugen, 1996; Bender et al., 2019). Lyle and Yohn (2021) refer to them as Covariance portfolios. In these portfolios, expected returns are ignored and the risk of the portfolio is minimized. The underlying assumption is that expected stock returns are cross-sectionally constant but covariances differ cross-sectionally (Lyle & Yohn, 2021). So, they are equivalent to mean-variance portfolios if expected returns across all assets are the same (Bender et al., 2019).

A special case of minimum variance is maximum decorrelation. These portfolios weigh each asset taking into account that the variance of the portfolio has to be minimized, assuming all securities have the same volatility. They are equivalent to a mean-variance portfolio if expected returns and volatilities are equal (Bender et al., 2019).

Expected return portfolios assume that expected stock returns differ cross-sectionally but that the covariances are cross-sectionally constant. These portfolios are closest to those often used in the fundamental analysis literature in which risk is ignored (Lyle & Yohn, 2021).

In the maximum diversification scheme, the portfolio weights of securities are selected to maximize the ratio of weighted-average asset volatilities to portfolio volatility. In this case, they are equivalent to a mean-variance portfolio if expected returns are proportional to volatilities (Bender et al., 2019).

Risk parity or equal risk contribution portfolios are characterized by weights based on the contribution to portfolio risk. Each asset must have the same contribution to it. It is equivalent to a mean-variance portfolio if all assets have the same Sharpe ratio and the correlations between them are constant (Bender et al., 2019).

A maximum deconcentration approach is a modified method of equal weighting that maximizes the effective number of stocks. It is equivalent to a mean-variance portfolio if expected returns and volatilities are equal and correlations are constant (Bender et al., 2019).

Finally, in maximum Sharpe ratio portfolios, weights are selected to maximize the Sharpe ratio of the portfolio where, in practice, the forecast returns are often assumed to be proportional to security volatilities (Bender et al., 2019).

2.3 Portfolio construction based on stocks characteristics

According to Lyle and Yohn (2021) and Richardson et al. (2010), the most common approach used to demonstrate the predictive ability of fundamentals-based characteristics is to cross-sectionally rank stocks based on a specific characteristic and form equal-weighted or value-weighted portfolios of the stocks integrated into the extreme deciles or quintiles. Specifically, it is common practice to take a long position in the top extreme quantile and a short position on the bottom extreme quantile. This is known as a zero-cost investment strategy.

As explained earlier, equal-weighted and value-weighted portfolios are not formed based on an objective function nor consider risk. So they are not optimized (Lyle & Yohn, 2021). There has been increasing research on creating optimized portfolios taking into account asset characteristics, including Factor-Based approaches and Parametric Portfolio Policy approaches. According to Lyle and Yohn (2021), they offer performance gains over naive equal-weighted and value-weighted portfolios.

Factor-Based portfolios are based on the belief that factors drive returns and risk (Bender et al., 2019). In fact, equity investors increasingly view portfolios as not only a collection of securities but also a package of exposures to the factors which drive security returns (Clarke et al., 2016). For instance, Bender et al. (2019) examine three price-based characteristics – book-to-market, size, and momentum – and show that portfolio returns come from factors previously identified by the financial literature.

In this setting, risk factors with the potential to explain stock returns are identified. Then, estimates of expected stock returns are obtained by multiplying the stock's sensitivity to each factor by the estimated rate of return on that factor. The covariance matrix is obtained from the covariance of the factor returns, factor sensitivities or factor exposures, and idiosyncratic return variances (Lyle & Yohn, 2021). There are relevant properties that make this approach appealing: it reduces the number of estimates required in the covariance matrix and eliminates the reliance on expected return estimates (Lyle & Yohn, 2021).

Brandt et al. (2009) propose a Parametric Portfolio Policy model that directly determines the portfolio weight in each asset as a function of the asset's characteristics. Then, the coefficients of the function are found by optimizing the average utility of the portfolio's return over the sample.

While mean-variance portfolio optimization endogenously determines the weight function given expected returns and covariances, the Parametric Portfolio Policy approach exogenously assumes a parametric weight function of firm characteristics (Lyle & Yohn, 2021). This model has several advantages. First, it eliminates the need to directly estimate expected returns and the covariance matrix (Lyle & Yohn, 2021). It avoids completely the auxiliary and difficult step of modeling the joint distribution of returns and characteristics and focuses directly on the portfolio weights (Brandt et al., 2009). It is also computationally simple and it reduces the dimensionality of the problem (Brandt et al., 2009; Lyle & Yohn, 2021). Finally, it can be easily modified and extended by using different objective functions or capturing the effect of transaction costs (Bender et al., 2019).

2.4 Portfolio performance metrics

Several metrics have been used to assess portfolio performance. Some of the most traditional measures include the Treynor and the Sharpe ratios. There are also the alpha and information ratio associated with multifactor models, such as the Fama and French factor models (with three-, five- and six-factor) and the Carhart four-factor model.

The Treynor ratio is a portfolio performance measure derived from the Capital Asset Pricing Model (CAPM). It measures the excess return over systematic risk and is represented by Equation (1), where $\overline{r_p}$ is the average return on the portfolio, $\overline{r_f}$ is the average risk-free rate, and β_p is the beta of the portfolio.

$$T_p = \frac{\overline{r_p} - \overline{r_f}}{\beta_p} \tag{1}$$

The Sharpe (1966) ratio adjusts returns to the standard deviation of a portfolio instead of the beta. The author assumes that, under riskless lending and borrowing, the optimal portfolio is the one with the highest excess return over standard deviation. This ratio is also known as a reward-to-variability measure and is defined by Equation (2), where σ_p is the standard deviation of the portfolio.

$$SR_p = \frac{\overline{r_p} - \overline{r_f}}{\sigma_p} \tag{2}$$

Moving to Factor models, the simplest one is the single-index model. Here, the alpha is the Jensen (1968) measure. It is an absolute measure and represents the average incremental rate of return from possible deviations of the portfolio returns from the CAPM. If it is positive, the portfolio performance is above the market portfolio performance, otherwise, the portfolio performance is below the market portfolio performance. Jensen's alpha is usually estimated as the intercept in the time-series regression presented in Equation (3), where $r_{m,t}$ is the market return in period t and $\varepsilon_{p,t}$ the error term of the regression.

$$r_{p,t} - r_{f,t} = \alpha_p + \beta_p(r_{m,t} - r_{f,t}) + \varepsilon_{p,t} \tag{3}$$

Several issues related to CAPM have been discussed by academics and it was suggested that other variables besides the market and risk-free rates also explain returns (Elton & Gruber, 2011). Multifactor models emerged due to the empirical failures of the CAPM (Fama & French, 2018) and the development of the Arbitrage Pricing Theory (APT) led to a generalization of Jensen's measure. According to Elton et al. (2006), the use of a multifactor model allows for a more appropriate measuring and diagnosis of portfolio performance.

Fama et al. (1993) developed a three-factor model with market, size and book-to-market as the main factors. They found that these factors are useful to explain stock returns. The model is presented in Equation (4). It introduces the Small-minus-Big (SMB) and High-minus-Low (HML) factors, which are the difference in returns between a portfolio of small stocks and a portfolio of large stocks and the difference in returns between a portfolio of high book-to-market stocks and a portfolio of low book-to-market stocks, respectively.

$$r_{p,t} - r_{f,t} = \alpha_p + b_{p1}(r_{m,t} - r_{f,t}) + b_{p2}(SMB_t) + b_{p3}(HML_t) + \varepsilon_{p,t} \tag{4}$$

Carhart (1997) adds to the previous model the variable momentum (MOM), which represents the difference in the returns of a portfolio of past winners and a portfolio of past losers (Equation (5)). This variable was introduced since Jegadeesh and Titman (1993) concluded that there is a tendency for a good or bad performance of stocks to persist over several months.

$$\begin{aligned} r_{p,t} - r_{f,t} = {} & \alpha_p + b_{p1}(r_{m,t} - r_{f,t}) + b_{p2}(SMB_t) + b_{p3}(HML_t) + b_{p4}(MOM_t) \\ & + \varepsilon_{p,t} \end{aligned} \tag{5}$$

Motivated by the dividend discount model, Fama and French (2015) find evidence that variations in average returns related to profitability and investment are explanatory of portfolio returns. Taking this into account, the authors introduce to the three-factor model two other variables: a profitability factor (RMW) defined as the difference between the returns on diversified portfolios with robust and weak profitability and an investment factor (CMA) defined as the difference between the returns on diversified portfolios of stocks of low and high investment firms (Equation (6)).

$$\begin{aligned} r_{p,t} - r_{f,t} = {} & \alpha_p + b_{p1}(r_{m,t} - r_{f,t}) + b_{p2}(SMB_t) + b_{p3}(HML_t) + b_{p4}(RMW_t) \\ & + b_{p5}(CMA_t) + \varepsilon_{p,t} \end{aligned} \tag{6}$$

Fama and French (2018) also use a six-factor model, adding a Momentum factor to the five-factor model (Equation (7)). According to these authors, the momentum factor was added to the five-factor model, despite the absence of theoretical justification.

$$r_{p,t} - r_{f,t} = \alpha_p + b_{p1}\left(r_{m,t} - r_{f,t}\right) + b_{p2}(SMB_t) + b_{p3}(HML_t) + b_{p4}(RMW_t) + b_{p5}(CMA_t) + b_{p6}(MOM_t) + \varepsilon_{p,t} \tag{7}$$

The information ratio is another measure that can be determined using factor models. It is calculated by dividing the alpha by the standard deviation of the residual. So it can be interpreted as the excess return over the benchmark portfolio per unit of portfolio risk that is unrelated to the benchmark portfolio (Bender et al., 2019; Lyle & Yohn, 2021). The expression for this ratio is presented in Equation (8), where $\hat{\sigma}$ is the standard deviation of excess returns from the benchmark (Goodwin, 1998). As Clarke et al. (2016) point out, it should not be used alone to characterize portfolio performance since it does not account for the optimal amount of active risk.

$$IR_p = \frac{\overline{r_{p,t} - r_{m,t}}}{\hat{\sigma}} \tag{8}$$

3. Methodology

In this section, the methodology used to form the fundamental portfolios as well as other portfolios for performance comparison is explained. The first component of this analysis is the estimation of expected returns and the covariance matrix. I follow Lyle and Yohn's (2021) approach and estimate fundamentals-based expected returns and a full covariance matrix of individual stock returns. Those are then used to form mean-variance optimized fundamental portfolios in Markowitz's (1952) style.

After the construction of the portfolios of main interest for this research, alternative approaches, namely Expected returns, Minimum Variance and Equal-weighted portfolios, are constructed in order to examine performance differences. Finally, some extensions are introduced to the model including the consideration of different periods and cardinality constraints.

3.1 Modeling Expected Returns and Covariances

In order to calculate the expected returns needed for the mean-variance framework, I follow Lyle and Yohn (2021) and construct a model that directly links fundamentals-based ratios to expected returns. In the same line as them, rolling five-year time periods with monthly data are used for the estimation and portfolios are revised monthly.

In order to estimate the expected returns, Lyle and Yohn (2021) present a model similar to Equation (9), which is linear in firm characteristics. It includes as explanatory variables the inverse of the market value of equity ($1/M_{i,t}$), the book to market value of equity ratio ($B_{i,t}/M_{i,t}$), the earnings before extraordinary items to market value of equity ratio ($E_{i,t}/M_{i,t}$), the change in net operating assets to market value of equity ($\Delta NOA_{i,t}/M_{i,t}$) and the change in financial assets to market value of equity ($\Delta FIN_{i,t}/M_{i,t}$).

$$R_{i,t+1} = \beta_{i,0} + \beta_{i,1}\frac{1}{M_{i,t}} + \beta_{i,2}\frac{B_{i,t}}{M_{i,t}} + \beta_{i,3}\frac{E_{i,t}}{M_{i,t}} + \beta_{i,4}\frac{\Delta NOA_{i,t}}{M_{i,t}} + \beta_{i,5}\frac{\Delta FIN_{i,t}}{M_{i,t}} + \Omega_{i,t}\varepsilon_{i,t+1} \quad (9)$$

It is important to note, however, that returns have monthly data and the independent variables of the model are only updated quarterly at the end of the month in which they are reported. So, for the model used to create the estimates for expected returns, a variable of lagged three-month returns is added, to guarantee variability of estimates even when the accounting ratios remain constant (Equation (10)).

$$R_{i,t+1} = \beta_{i,0} + \beta_{i,1}\frac{1}{M_{i,t}} + \beta_{i,2}\frac{B_{i,t}}{M_{i,t}} + \beta_{i,3}\frac{E_{i,t}}{M_{i,t}} + \beta_{i,4}\frac{\Delta NOA_{i,t}}{M_{i,t}} + \beta_{i,5}\frac{\Delta FIN_{i,t}}{M_{i,t}} + \beta_{i,5}R_{i,t-3} + \Omega_{i,t}\varepsilon_{i,t+1} \quad (10)$$

In what concerns outliers, in order to ensure that each predictor is close to mean zero with a stable distribution over time, it is cross-sectionally standardized, by converting it into a percentile rank, dividing by 99, and subtracting 0.5, in line with Lyle and Yohn (2021) and Green et al. (2011).

Regarding the estimation, five years of monthly returns historical data are collected for each stock. In order to ensure reasonable estimates for the covariance matrix (stock return volatility and pairwise correlations), I then remove penny stocks with prices below 1 dollar and/or negative book value, which is common practice (Hou et al., 2021). Note that Lyle and Yohn (2021) refer that they remove penny stocks and stocks with negative book value as they clean the data and not in each estimation period. However, to avoid look-ahead bias, I opt for this procedure.

Then, the coefficients are estimated by regressing one month-ahead stock returns on the fundamental variables. After the estimation of the model cross-sectionally, an initial estimate of expected returns is obtained. Then, following Lyle and Yohn (2021), I form deciles of those expected returns and re-estimate the model within each decile. Finally, for each estimation period, the estimated in-sample coefficients are applied to the most recent fundamental variables to generate the expected returns. Following Lyle and Yohn (2021) as well, these expected return estimates are winsorized at the 1% and 99% levels.

The covariance matrix of stock returns is formed with the residuals from the regressions. A nonlinear shrinkage estimator is applied (Ledoit & Wolf, 2020). According to Lyle and Yohn (2021), this approach overcomes issues related to estimating large-dimension covariance matrices and is easy to implement. I am using a different shrinkage estimation from Lyle and Yohn (2021) since this nonlinear shrinkage estimator from Ledoit and Wolf (2020) has two main advantages compared to the nonlinear shrinkage estimator from Ledoit and Wolf (2017), namely being faster with basically the same accuracy and accommodating covariance matrices of dimension up to 10,000 and more.

After estimating the fundamentals-based expected returns and the covariance matrix of individual stock returns, I form four different portfolios – equal-weighted portfolios, covariance portfolios, expected return portfolios and mean-variance optimal portfolios.

3.2 Alternative portfolio approaches

In order to compare the gains from the fully optimized fundamental portfolios, it is vital to understand if these gains arise from the use of the fundamentals-based expected returns, the covariance matrix, or both. So, I consider four different portfolio construction approaches with distinct characteristics in terms of expected returns and covariance: (1) Equal-weighted portfolios, (2) Covariance portfolios, (3) Expected return portfolios, and (4) Fully optimized fundamental portfolios. As referenced in the Literature Review (Section 2.2.2), these portfolios have different assumptions for what is cross-sectionally constant: expected returns, covariances, or both.

Considering this, Table 1 presents a summary of the different assumptions of the portfolios regarding expected returns and covariances. Taking this into account, (1) Covariance portfolios can be compared with Equal-weighted portfolios to analyze performance gains from cross-sectional differences in covariances, (2) the gains from cross-sectional differences in expected returns are obtained comparing Expected return portfolios relative to Equal-weighted portfolios, and, finally, (3) gains arising from cross-sectional differences in expected returns and covariances are found comparing Fully optimized portfolios relative to Covariance, Expected return, and naive Equal-weighted portfolios (Lyle & Yohn, 2021).

Table 1: Portfolio construction assumptions regarding the cross-sectional constancy/variability of expected returns and covariances

	Expected returns	Covariances
Equal-weighted portfolio	Constant	Constant
Covariance portfolio	Constant	Variable
Expected return portfolios	Variable	Constant
Fundamental portfolios	Variable	Variable

Regarding the methodology for the creation of these portfolios, the weights for the Equal-weighted portfolios are obtained by dividing one by the number of stocks in the portfolio, N. This is represented in Equation (11), where e is a $N \times 1$ vector of ones.

$$w_{EW} = \frac{1}{N} e \tag{11}$$

For the Covariance portfolios, I minimize portfolio variance and ignore expected returns (Equations (12) and (13)). This results in the optimal portfolio policy presented in Equation (14). Σ is a $N \times N$ covariance matrix, and $w = (w_1, w_2, \dots, w_N)^T$ is a $N \times 1$ vector of portfolio weights

$$\min_{w} \frac{1}{2} w^T \Sigma w \tag{12}$$

$$s.t. w^T e = 1 \tag{13}$$

$$w_{COV} = \frac{\Sigma^{-1} e}{e^T \Sigma^{-1} e} \tag{14}$$

For the Expected returns portfolios, I assume the covariance matrix to be an Identity matrix and maximize expected returns subject to this risk constraint (Equations (15), (16) and (17)). The consequent portfolio policy is expressed in Equation (18), where μ is a $N \times 1$ vector of expected returns

$$\max_{w} w^T \mu \tag{15}$$

$$s.t. w^T \Sigma w = e \tag{16}$$

$$w^T e = 1 \tag{17}$$

$$w_{ER} = \frac{\mu}{e^T \mu} \tag{18}$$

Finally, for the Fully optimized fundamental portfolios, I apply both the covariance matrix and expected returns previously explained, maximizing the expected return of the portfolio subject to a given portfolio variance (Equations (19), (20) and (21)).

$$\max_{w} w^T \mu \tag{19}$$

$$s.t. w^T \Sigma w = \Sigma_p \tag{20}$$

$$w^T e = 1 \tag{21}$$

This maximization of the expected portfolio returns per unit of portfolio volatility results in the optimal portfolio policy presented in Equation (22).

$$w_{FOP} = \frac{\Sigma^{-1}\mu}{e^{T}\Sigma^{-1}\mu} \quad (22)$$

I form both long-short and long-only portfolios since taking short positions may not be feasible and, even when feasible, implementation costs may be high (Lyle & Yohn, 2021). In the same line as Lyle and Yohn (2021), I rebalance the portfolios monthly and examine the out-of-sample performance of the portfolio over the period.

3.3 Portfolio performance comparison

In order to analyze the differences in portfolio performance of Equal-weighted portfolios, Covariance portfolios, Expected return portfolios, and Fundamental portfolios, I report (for all of them) key portfolio performance metrics, including the average returns, the standard deviation of returns, the Treynor ratio, and the Sharpe ratio. To better understand gains in performance across portfolios, I test the differences in Sharpe ratios for each pair of portfolios using the approach of Ledoit and Wolf (2008). This approach is based on studentized bootstrap inference and, because of that, does not present the common problems of other testing methods such as not being robust against tails heavier than the normal distribution, which is quite common with financial returns.

Then, I report the Fundamental Optimized portfolio alphas and Information ratios relative to, the CAPM, Fama-French three-, five- and six-factor models, and Carhart four-factor model.

3.4 Extensions

In order to further analyze gains in performance, I extend the methodology with: (1) different time periods, and (2) samples of different market capitalization firms.

Regarding the performance for different time periods, I first construct a timeframe exactly equal to one of the subperiods used by Lyle and Yohn (2021). This provides the possibility for a direct comparison of results. Secondly, I divide the sample time period into two non-overlapping nine-year periods (October 2003 – September 2012 and October 2012 – September 2021) and replicate the performance comparison methodology previously explained for each one.

According to Lyle and Yohn (2021), excluding smaller stocks from the sample reduces the likelihood of investors facing liquidity issues and higher transaction costs. So, I also form and compare performances of portfolios only containing only the 500, 200, and 100 largest stocks.

3.5 Portfolios on the extreme deciles of fundamentals-based expected returns

My next analysis is inspired by the original work of forming portfolios on the extreme deciles of stocks. However, I adapt this approach and form all portfolios (Equal-weighted, Value-weighted, Expected returns, Covariance and Fully Optimized portfolios) in the top decile of expected returns. This implies that the long-short portfolios in this analysis are not formed with a long position in the top decile and a short position in the bottom decile. They are not zero-cost strategies. Instead, I apply the maximization problems presented in Subsection 3.2 to the stocks in the top decile of expected returns each month.

This analysis has two main advantages. First, for an investor, it has fewer costs for an investor than the original zero-cost strategy. Second, it can give an insight into the predictive power of the models used. If the portfolios created using a sample of stocks from the top decile of expected returns have higher returns than the portfolios based on the entire sample, it indicates that the expected returns can to a certain degree predict ex-post returns.

4. Data

The data for this dissertation is mainly retrieved from the Refinitiv Eikon Datastream platform. The sample includes all US market-listed stocks. The necessary data for the study includes three different types of variables: (1) static data on stock characteristics, (2) stock market variables and (3) accounting variables. The static data on stock characteristics is used to create a comprehensive list of US stocks. The stock market variables are used to calculate returns. And, the accounting variables are used to compute the explanatory variables of the regression model. Regarding the data for computing the factor models, the risk-free rate and factor portfolios are collected from Kenneth French's data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

Data sources and data processing can have a very significant impact on empirical research conclusions (Landis & Skouras, 2021). Guidelines from Landis and Skouras (2021), specifically taking into account specificities for the US market, are used to derive high-quality equity data from Refinitiv Eikon Datastream data. Their approach improves data accuracy, filters problematic data and reduces survivorship bias and data staleness.

4.1 Data extraction

The stock universe should include not only stocks active at the moment of the study, but also stocks that have become inactive, for instance, because of mergers, acquisitions, or failure, but have some of their price history throughout the period under analysis. According to Baker and Haugen (1996), if a significant number of firms that have become inactive are systematically excluded, the data can suffer from survival bias. This bias is exacerbated since providers tend to add companies that are larger and more successful when the records are backfilled.

In order to create the initial stock universe, various Constituent Lists of stocks provided by Datastream are merged. Three types of lists are included: lists with currently active stocks (whose name starts with "FUSALL"), those including delisted stocks (whose name starts with "DEADUS") and Worldscope lists (whose name starts with "WSUS"). With this approach, a dataset of 137,521 stocks is obtained. However, this includes duplicates and stocks that are wrongly categorized in the constituent list.

The guidelines from Landis and Skouras (2021) invoke the extraction of information for all equities traded in all stock exchanges of the country using Datastream's Navigator GUI, instead of relying on this market specific constituent lists of equities provided by Datastream and Worldscope. However, this process is extremely time-consuming, especially when considering a country with a large stock market, such as the US.

For all the stocks on the above-mentioned lists, several variables were retrieved. namely: Type of Instrument (TYPE), Datastream Code (DSCD), Base Date (BDATE), Expanded Security Name (ENAME), Datastream Exchange Mnemonic (EXMNEM), Geographical Classification of company (GEOGN), ISIN (ISIN), Primary Indicator Flag (ISINID), Code Local (LOC), Currency (PCUR) and Security Type Code (TRAC). The description of each of these variables can be found in Appendix A, Panel 1.

4.2 Data Filters

In this subsection, there is a description of the several filters applied to the data obtained previously. They are mainly based on Landis and Skouras (2021) guidelines. It is important to keep in mind that, first, there is always a trade-off between too much and too little filtering and, second, the order in which the filters are applied can have an impact on the final result. These filters can be divided into two categories: stock filters and stockday filters. Stock filters exclude the entire history of an instrument whereas stockday filters exclude a specific day of a specific stock.

4.2.1 Filters based on static information

This subsection includes a description of the filters based on the static information explained above that will contribute to filtering the data. Table 2 is a summary of these filters. The filters are numbered by the order they are applied in this dissertation. The column "Variable" indicates the variable used to do the specific filter and the column "Accepted Values" refers to the specific values the variable can take to the stock being accepted. Finally, the column "Number of observations after the filter" presents the total number of stocks after applying the filter. It is important to note that whereas Landis and Skouras (2021) apply each filter independently of the others, the numbers in this column are cumulative.

Table 2: Stock filters based on static information description

Number	Variable	Accepted values	Number of observations after the filter
1	Type of Instrument (TYPE)	"EQ"	111,363
2	Datastream Exchange Mnemonic (EXMNEM)	"NAS", "NYS", "OTC", "ASE", "XSQ", "XBQ", "NMS", "BOS", "MID", "PSE", "PHL"	110,685
3	Security Type Code (TRAC)	"ORD", "ORDSUBR","FULLPAID", "UKNOWN", "UNKNOW", "KNOW", "NA"	102,166
4	Expanded Security Name (ENAME)	**Do not contain** "TRUST", "REPR", "RIGHT", "SERIES","NV", "IV TST", "REAL ESTATE INVESTMENT", "REALTY", "RLTY", "ROYALTY INVESTMENT", "ASSET INVESTMENT", "CAPITAL INVESTMENT", "ASSET MANAGEMENT", "CAPITAL MANAGEMENT", "INVESTMENT MANAGEMENT", "VENTURE CAPITAL", "FINANCIAL SHBI", "PROPERTY INVESTORS", "INCOME PROPERTY", "UNITS", "UNIT", "LIMITED PARTENERSHIP", "FUND", "EQUITY PARTNERS", "LIMITED VOTING", "SUB VOTING", "TIER ONE SUB", "VARIABLE VOTING", "NON VOTINGREIT", "RESIDENTIAL", "R E I T", "BENEFICIAL", "BENEFICIARY", "BENEFIT INTEREST", "BEN INTEREST", "SH BEN INT", "WARRANT", "WRTS", "L P", "L P INTEREST", "LP UT", "HOLDINGS LP", "PARTENERS UNIT", "PART INT", "UNIT PARTENERSHIP", "UNIT LIMITED", "MORTGAGE", "REAL ESTATE", "CERTIFICATE", "NO PAR VALUE", "HOLDING UNIT", "BACKED", "ST MIN", "CORTS", "TORPS", "TOPRS ", "SECURITIES TRUPS", "QUIPS", "STRATS HIGH YIELD", "TOTAL RETURN", "DIVERSIFIED HOLDINGS", "(SICAV)", "DEPOSITARY", "DEPOSITOR", "RECEIPT", "REP & SHARES", "GLOBAL SHARES", "ADR", "GDR", "EXPD.", "EXPIRED", "DUPLICATE", "CONVERTIBLE", "CNVRT.", "CONVRT.", "EXCH.", "DEBANTURE", "(DEB)", "NIL PAID", "STRUCTURED ASSET", "CALLABLE", "FLOATING RATE", "ADJUSTABLE", "REDEEMABLE", "PAIRED CTF", "CONSOLIDATED", "INSURED", "CAPITAL SHARES", "DEBT STRATEGIES", "LIQUIDATING", "LIQUID UNIT", "L UNIT", "- LASD", "ACQUISITION", "CAP UNIT", "INCOME UNIT", "PREFERRED" **Nor** "(NYS)", "(NAS)", "(ASE)", "(OTC)", "(XSQ)", "(XQB)"	68,966
5	Geographical Classification of company (GEOGN)	"UNITED STATES"	67,811
6	Currency (PCUR)	"U$"	67,809
7	Code Local (LOC) **and** Primary Indicator Flag (ISINID)	No duplicated codes – if duplicated remove if ISINID is not P, as long as there exists one stock with this LOC that does have ISINID equal to P	35,678
8	Datastream Code (DSCD)	No duplicated codes	35,198

The first filter applied is to make sure that all the observations are Equities (the Type of instrument is ”EQ”). Since the authors manually locate all instruments in the Category Equities and Type Equity, this filter is automatically done by them. However, by using the constituent lists, other types of securities are also found in the data, including, for instance, Unit Trusts (UT), American Depository Receipts (ADR) and Closed-end Funds (CF).

The next filter excludes stocks that are not traded on US Exchanges. According to Landis and Skouras (2021), some authors choose to exclude stocks listed in secondary exchanges, by assuming that stocks in secondary exchanges are likely to be small. However, a major limitation of this approach when applied to Datastream data, is that the platform only reports current exchange classifications. Consequently, excluding stocks from secondary exchanges would possibly cause the exclusion of stocks that were previously in a primary exchange and have been demoted due to poor performance (sample selection bias). The US has twelve exchanges plus the OTC market. The twelve Exchanges include: NYSE (NYS), NYSE MKT (ASE), NYSE ARCA (XC), NASDAQ (NAS), NASDAQ/NMS (NMS), OTC Bulletin (XBQ), Non-Nasdaq OTC (OTC), Boston (BOS), Chicago (MID), Pacific (PSE), Philadelphia (PHL), and BATS (E1)

Then, I exclude all instruments that cannot be classified as common stocks. To do so, I exclude all stocks with security type code (TRAC) taking any value other than "ORD", "ORDSUBR", "FULLPAID", "UKNOWN", "UNKNOW" and "KNOW" and “NA”. I maintain all the values with unknown content since, first, they correspond to a large part of the sample and it is very unlikely that all the securities classified as unknown would not be common equity and, second, this happens frequently in delisted stocks, which can cause an unsuspecting user to filter out delisted stocks instead of noncommon stocks, biasing the filtered sample.

In order to filter even more securities to only include common stocks, I use Filter 5. This filter searches for specific text strings in the extended name of the security (ENAME) which identify a stock that is non-common. This varies by country, so Table 2 presents the ones used for the US Market. Another common issue that can be solved using the extended name is the presence of Cross-listed stocks. Stocks that are listed in another country usually have on their name a string that suggests it is the local listing of a stock primarily traded elsewhere, for instance having the code of the US exchange.

Another relevant filter is the exclusion of stocks that are not domiciliated in the US. I remove any security for which the geographical classification (GEOGN) is not the United States. Similarly, in the next filter, I remove all securities traded in a currency (PCUR) other than the US dollar.

The last two filters in this category remove the duplicated observations. First, I use the Local Code (datatype LOC), by excluding all stocks which have a nonunique Local code and their ISINID is not P, as long as there exists one stock with that local code that has ISINID equal to P. Second, I remove duplicated Datastream codes. After the application of all the above filters, the sample included 35,198 US common stocks.

4.2.2 Filters based on return index information

For the 35,198 stocks previously found, I retrieve the daily Return Indices, Prices, Unadjusted Prices and Adjustment Factors between December 1984 and September 2021. The sample starts in December 1984 since Landis and Skouras (2021) found that Datastream data for the US is more reliable after this date. I extend the period to the most recent information available at the beginning of this study.

Although, as explained in the Methodology section, the models are based on monthly data, I retrieve daily data since, according to Landis and Skouras (2021), this makes it possible to identify data problems that would be difficult to recognize with monthly data.

When retrieving this data, I use the Datastream DPL function and round the values to the largest available number of decimal places, which is six. Especially with small values, the standard rounding of Datastream (two or four decimal places) would have a large impact on returns calculations.

The filters based on the times series data are summarized in Table 3. They include stockday filters and stock filters (third column of the table).

Although the last section's filters lead to a sample of 35,198 stocks, only 31,171 of them had return index data available. I retrieve not only the RI series, but also the RI#S and RI#T series. In Datastream, adding the "#T" after the symbol of the series displaying N/A after the series goes dead, rather than padding the last real value. Adding "#S" removes padded values for non-trading days.

Table 3: Stock and stockday filters based on return index information description

Number	Filter	Stock filter vs Stockday filter	Number of observations after the filter
1	Return index availability	Stock filter	31,171
2	Remove returns beyond the last trading date	Stockday filter	29,906
3	Identify stocks with more than 95% of zero returns	Stock filter	27,642
4	Implausibility filter	Stock filter	27,561
5	Outlier errors	Stockday filter	27,561
6	Holidays	Stockday filter	27,561
7	Remove stocks with few observations 1250	Stock filter	20,166
8	High volatility	Stock filter	16,217
9	Low volatility	Stock filter	16,217
10	Staleness	Stockday filter	16,217
11	Adjustment inconsistencies	Stockday filter	16,217
12	Nonsense values	Stockday filter	16,217
13	Remove stocks with few observations 1250	Stock filter	15,713
14	Remove stocks from Utilities and Financial industries	Stock filter	12,619

Datastream delisting dates typically occur later than the last date for which return index data is available. This means that many stocks appear with padded values for return indexes towards the end of their series even when the series has been truncated at their delisting dates. I retrieve de delisting date (TIME) for each stock. I compare the last non-zero return date of the RI, RI#S and RI#T series and the delisting date. If the delisting date is not available or if there is a difference of more than 10 days between the last non-zero return and the delisting date, I remove the tenth and subsequent padded daily observations. Otherwise, I use as the end date of the series the date retrieved directly from Datastream. Although this is a stockday filter, the entire history of 1,265 stocks was removed, as all their return indexes had the same value.

In the third filter, I remove stocks for which the returns are zero in more than 95% of their sample. Then, I apply an "implausibility filter" which removes stocks of which more than 98% of the non-zero daily returns are either non-negative or non-positive. According to Landis and Skouras (2021), the underlying problem is that Datastream bases its return index calculation on a dividend yield factor, suggesting that a dividend was paid out every day for which data is available.

In order to reduce outliers in returns caused by errors in adjustments, I apply the following methodology: (1) if the daily return on a date is greater than 100% and the daily return on the next day is lower than -50% or (2) if the daily return on a date is lower than -50% and the return on next day is greater than 100%, then both days are eliminated.

Another set of days that should be removed from the sample is holidays and days when markets are closed. I identify and remove these days by analyzing if for a day the percentage of non-zero returns is below 20%. After this, I also confirm if the identified dates are correct by searching the exchanges' calendars. Taking into account that, as explained in the Methodology section, at least five years of data for each stock are needed, I require 1250 valid daily observations.

The next two stock filters are related to volatility. A daily standard deviation of more than 40% is a regular occurrence in Datastream. This can be caused, for instance, by missing observations that are filled with padded return indexes, large errors in adjustments for corporate events or extreme illiquidity which leads to very rare price updates. I remove all stocks of standard deviation above this value. Landis and Skouras (2021) also eliminate stocks with a daily standard deviation of less than 0.01 bps. This filter had no impact on my sample at this stage.

Still related to the staleness caused by padded prices, if 30 consecutive return indexes are identical, all subsequent observations are eliminated until the next change. This has a significant impact on the number of daily observations.

The next filters are related to prices. Sometimes, there can be an inconsistency between adjusted prices and the prices implied from unadjusted prices and adjustment factors. I filter out cases where there is a discrepancy of more than 5% between the two. I filter out all days for which the price is more than 5% different from the unadjusted price times the adjustment factor. I also make sure there are no stockdays for which unadjusted prices contain non-sense values, specifically zero or negative values.

After the application of all the filters, I again remove stocks with less than 1250 daily observations. The sample is composed of 15,713 stocks. For these stocks, a series of monthly returns is calculated. Also, for these series, I apply the outliers filter, which removed 26,666 observations, and remove 233 stocks with less than 60 months of data.

Finally, since accounting is different for firms in the financial and utilities industries due to heavy regulations (Hou et al., 2021; Lyle & Yohn, 2021; Palazzo, 2012), I also remove them from the sample, guaranteeing comparability across results. I exclude from the dataset stocks with SIC codes between 4900 and 4999 and between 6000 and 6999. This reduces the number of stocks for the study to 12,619.

4.3 Accounting variables

In order to reproduce the study, data on accounting variables need to be updated every three months. Datastream provides data from Worldscope on various accounting variables, both annually and interim. However, coverage is significantly different: for some variables only annual data is available; for other variables, although interim data is available, the number of companies covered is drastically lower than the annual counterpart. Another relevant issue is the lack of data for the early years. Interim data is only available from 1998, which implies that, although return index data is available from December 1984, this study can only be conducted after 1998.

Table 4 contains a brief summary of all the variables retrieved from Datastream: the code, name and if the data used in the study is interim or annual. Interim variables have the same datatype as annual variables, but include an "A" at the end. As per previous sections, a brief description of the variables is presented in Appendix A. Each of the next subsections explains how each of the explanatory variables used in Equation (9) is obtained.

Table 4: Refinitiv Datastream Static Datatypes, description and availability

Datatype	Name	Interim vs Annual
WC02999	Total assets	Interim
WC03351	Total liabilities	Interim
WC04101	Deferred Income Taxes & Investment Tax Credit	Annual
WC03451	Preferred Stock	Interim
WC01551	Net Income Before Extraordinary Items/Preferred Dividends	Interim
WC02001	Cash & Short-Term Investments	Interim
WC03051	Short-Term Debt & Current Portion of Long-Term Debt	Interim
WC03251	Long-Term Debt	Interim
WC03426	Minority Interest	Interim
WC03501	Common Equity	Interim
WC02008	Short-term investments	Annual
WC02258	Long-Term Receivables	Annual
WC02250	Other Investments	Annual

4.3.1 Book value of equity

In order to obtain the Book value of equity in each moment of time, I use Equation (23), following Fama and French (2006), Hou et al. (2014), and Palazzo (2012). Deferred Income Taxes & Investment Tax Credit and Preferred Stock are assumed to be zero if not available. I also tested the possibility of using the datatype "shareholders equity" instead of calculating Total Assets minus Total Liabilities. For most observations, the values coincided, however, I decided to apply the calculation presented below. The Worldscope variable Deferred Income Taxes & Investment Tax Credit (WC04101) has no interim series associated, which means that it is only available with an annual frequency.

$$\begin{aligned} &Boook\ Value\ of\ Equity \\ &= Total\ assets\ (WC02999) - Total\ Liabilities\ (WC03351) \\ &+\ Deferred\ Income\ Taxes\ \&\ Investment\ Tax\ Credit\ (WC04101) \\ &-\ \ Preferred\ Stock\ (WC03451) \end{aligned} \tag{23}$$

4.3.2 Earnings before extraordinary items

Earnings before extraordinary items are simply obtained from retrieving the Worldscope variable Net Income Before Extraordinary Items/Preferred Dividends, whose code is WC01551.

4.3.3 Net operating assets

Net operating assets are defined as the difference between all operating assets and all operating liabilities (Hirshleifer et al., 2004). According to Hou et al. (2014), operating assets are total assets minus cash and short-term investment (Equation (25)). Operating liabilities are total assets minus short-term debt, minus long-term debt, minus minority interest, minus preferred stocks, and minus common equity (Equation (26)). Variables Short-Term Debt, Long-Term Debt, Minority Interest and Preferred Stock are considered zero if the value is missing.

$$Net\ Operating\ Aseets = Operating\ Assets - Operating\ Liabilities \tag{24}$$

$$\begin{aligned} &Operating\ Assets \\ &\qquad = \ Total\ Assets\ (WC02999) \\ &\qquad -\ Cash\ and\ Short\ Term\ Investment\ (WC02001) \end{aligned} \tag{25}$$

$$
\begin{aligned}
&Operating\ Liabilities \\
&\qquad = Total\ Assets\ (WC02999) \\
&\qquad - Short\ Term\ Debt\ (WC03051) \\
&\qquad - Long\ Term\ Debt\ (WC03251) \\
&\qquad - Minority\ Interest\ (WC03426) \\
&\qquad - Preferred\ Stock\ (WC03451) \\
&\qquad - Common\ Equity\ (WC03501)
\end{aligned} \tag{26}
$$

4.3.4 Financial assets

Although Lyle and Yohn (2021) do not specify what they consider to be financial assets in the model, I use Net financial assets, since this is the common approach in literature. Karpoff and Lou (2010) assert that the change in net financial assets is the change in short-term investments and long-term investments less the change in short-term debt, long-term debt, and preferred stock. Hou et al. (2014) describe this relation as net financial assets being financial assets minus financial liabilities. Financial assets are short-term investments plus long-term investments. In Refinitiv Datastream, long-term investments correspond to the variables Long-term Receivables and Other Investments. Financial liabilities are equal to Long-term Debt plus Short-term Debt plus Preferred Stocks (Equation (27)). As done with net operating assets, missing values in Short-term Debt, Long-term Debt, Short-term Investments, Long-term Investments, and Preferred Stocks are set to zero, as long as they are not all missing.

$$
\begin{aligned}
&Net\ Financial\ Aseets \\
&\qquad = [Short\ term\ investments\ (WC02008) \\
&\qquad + Long\ term\ Investments\ (WC02258\ +\ WC02250)] \\
&\qquad - [Long\ term\ debt\ (WC03251) \\
&\qquad + Debt\ in\ current\ liabilities\ (WC03051) \\
&\qquad + Preferred\ Stock\ (WC03451)]
\end{aligned} \tag{27}
$$

5. Empirical results

This section provides a description and analysis of the results found in this study. It starts with an analysis of the quality of the model suggested in Equation (9). Then, the results for the base methodology of this research are presented, followed by the analysis for different time periods, cardinality constraints and the portfolios in the top decile.

5.1 Precision of the model

Table 5 presents the fixed effects estimation results of several models. For all the models the dependent variable is the actual return at time t +1 multiplied by 100. In models 1 to 5, each of the independent variables in Equation (9) is taken as the only explanatory variable. Model 6 combines all these variables in one model.

Table 5: Regression analysis: Dependent variable is $100R_{t+1}$

$1/M_t$ is the inverse of the market value of equity, B_t/M_t is the book to market value of equity ratio, E_t/M_t is the earnings before extraordinary items to market value of equity ratio, $\Delta NOA_t/M_t$ is the change in net operating assets to market value of equity, and $\Delta FIN_t/M_t$ is the change in financial assets to market value of equity.
T-statistics in parenthesis.
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$\frac{1}{M_t}$	21.963*** (1.677)					21.969*** (1.298)
$\frac{B_t}{M_t}$		7.165*** (1.026)				0.501 (0.909)
$\frac{E_t}{M_t}$			2.608*** (0.362)			3.748*** (0.346)
$\frac{\Delta FIN_t}{M_t}$				0.690*** (0.074)		0.261*** (0.073)
$\frac{\Delta NOA_t}{M_t}$					-1.119*** (0.101)	-0.882*** (0.091)
Num.Obs.	462,925	462,925	462,925	462,925	462,925	462,925
R2	0.174	0.167	0.165	0.164	0.164	0.176
R2 Adj.	0.168	0.162	0.159	0.158	0.159	0.171

The overall results are similar to the ones found by Lyle and Yohn (2021). In Model 1, I find a positive and significant relation between the inverse of the firm size and returns. This relationship persists in Model 6. This is consistent with the previous literature that finds that small firms tend to have higher average returns than big firms.

In Model 2, higher book-to-market ratios are, on average, associated with higher stock returns. This is also in line with previous findings that high book-to-market (value) firms have higher average returns than low book-to-market (growth) firms. However, this association is no longer significant when the other variables are added to the model.

Model 3 also presents a common association in literature: more profitable firms tend to grow faster and have higher returns. The relationship between returns and one-month lagged earnings over market value is positive and statistically significant in both Models 3 and 6.

In what concerns the relationship between the change in financial assets over the market value and future returns, it appears to be positive and statistically significant. This result is consistent with the one found by Lyle and Yohn (2021) in the same setting.

Finally, there is a negative and significant association between the change in net operating assets over the market value and future returns in Models 4 and 6. This result is in accordance with the literature. As noted in the literature review section of this dissertation, the change in net operating assets is the broad measure of accruals and research tends to show a negative relation between accruals and future returns.

The adjusted R-squared for all the models is between 0.158 and 0.171, having Model 6 the highest value. However, I emphasize that this evidence that certain characteristics explain returns is not inconsistent with market efficiency.

5.2 Comparing nested portfolios

5.2.1 Average returns, standard deviations, Sharpe ratios and Treynor ratios

Table 6 Panel A presents values for average returns, standard deviations, Sharpe ratios and Treynor ratios for the portfolios created in the period between July 2003 and September 2021. The Equal-weighted portfolios have an average monthly return of approximately 1.3% and a standard deviation of 5.5%. The Sharpe ratio for these portfolios is 21.5%.

The Covariance portfolios provide, on average, a return of 1.1% if short sales are allowed and 1.2% if not. The standard deviation is also slightly higher for the long-only portfolios (4.4% vs 4.3%). The Sharpe ratios for long-short and long-only Covariance portfolios are 23.3% and 25.4%, respectively.

Table 6: Nested Portfolio analysis

EW = Equal-weight portfolio; COV = Covariance portfolio; ER = Expected returns portfolio; FOP = Fully optimized portfolio
Mean and Standard deviation are shown as a percentage.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel A: Portfolio summary statistics – Average returns, standard deviations, Sharpe ratios and Treynor ratios

		Long-short			Long-only		
	EW	COV	ER	FOP	COV	ER	FOP
Mean	1.279	1.092	2.350	1.559	1.220	1.459	1.682
SD	5.486	4.279	10.025	7.705	4.417	5.409	4.775
SR	0.215	0.233	0.225	0.190	0.254	0.252	0.332
TR	0.010	0.011	0.015	0.031	0.012	0.011	0.017

Panel B: Differences in Sharpe ratios

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW
Long-Short	0.017	0.010	-0.007	-0.042	-0.035	-0.025
Long-Only	0.039	0.036**	-0.002	0.077**	0.080**	0.116**

The long-short Expected return portfolios provide a higher average return (about 2.4%). However, this is also accompanied by a larger standard deviation (10.0%). This is translated into a Sharpe ratio of 22.5%. In the long-only scenario, the average return is 1.5% with a standard deviation of 5.4%. The associated Sharpe ratio is 25.2%.

Finally, Fully optimized portfolios provide an average return of 1.6% and 1.7% in the long-short and long-only scenarios, respectively. The standard deviation is higher for the long-short portfolios (7.7% vs 4.8%). This translates into a Sharpe ratio of 19.0% when short sales are allowed and 33.2% when short sales are not allowed.

Treynor ratios are very similar across portfolios, except for the long-short Fully optimized portfolios, which have the highest Treynor ratio, at 0.031. Although the standard deviation for this portfolio is high, the beta is relatively low. This explains the difference between the Sharpe ratio and Treynor ratio ranking of portfolios being so different at first glance. Lyle and Yohn (2021) also concluded that Fully optimized portfolios tend to have lower betas.

Table 6 Panel B presents the differences in Sharpe ratios for each pair of portfolios. Regarding the long-short portfolios, there are no significant differences in Sharpe ratios. Using this measure, there is no evidence that any of the portfolios has a better performance compared to the others. However, with the constraint of no short selling, Expected returns portfolios have a higher Sharpe ratio than Equal-weight portfolios. This indicates that considering expected returns cross-sectional differences when constructing a portfolio improves performance. On the other hand, there is no evidence that Covariance portfolios have superior performance relative to Equal-weight portfolios. Expected returns portfolios and Covariance portfolios have no significant differences in Sharpe ratios. Finally, Fully optimized portfolios have superior performance relative to all other portfolios, indicating that both the covariances and expected returns can be associated with performance gains.

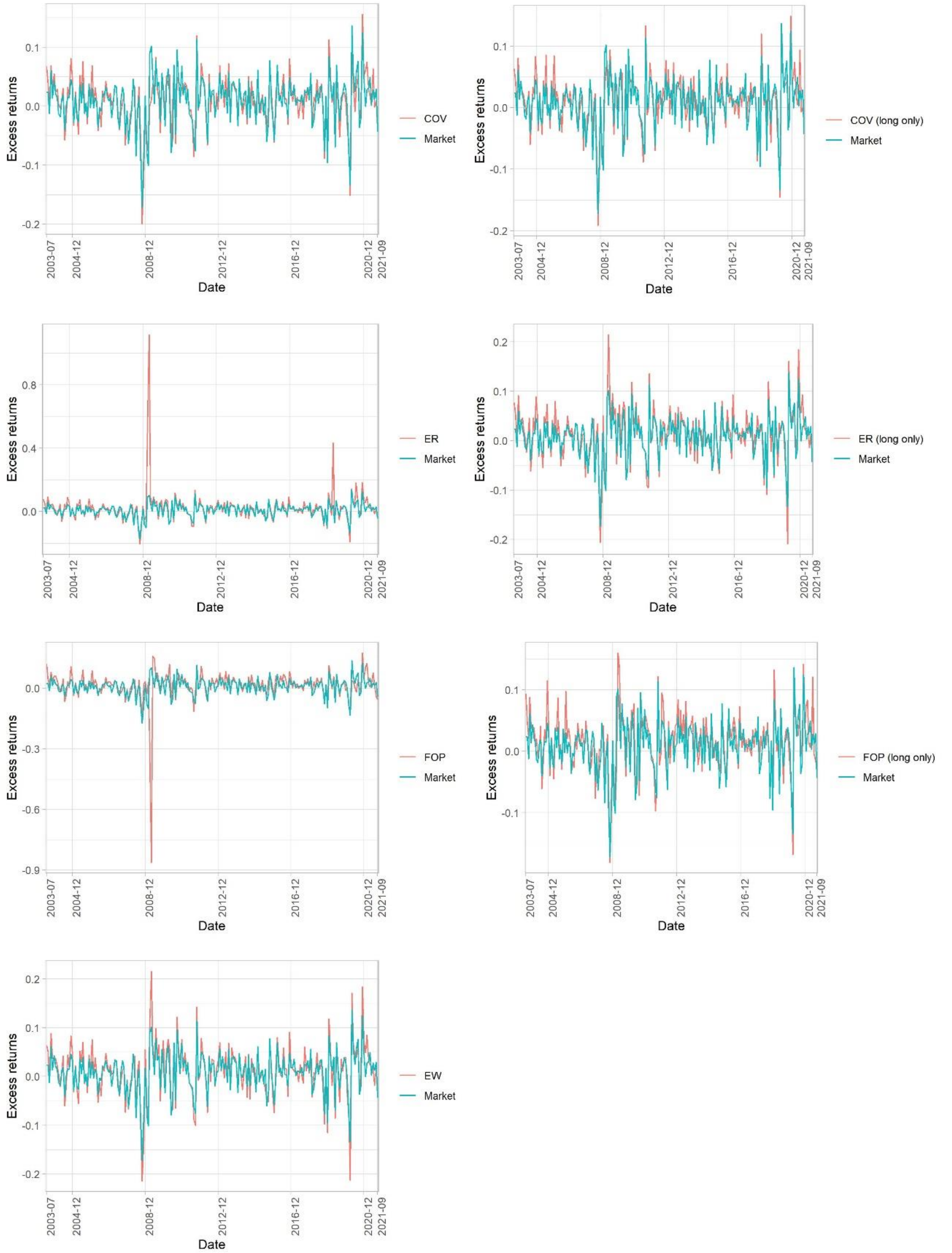
The biggest differences relative to the original results of Lyle and Yohn (2021) are the higher standard deviation of the long-short Expected returns and Fully optimized portfolios and the much lower average return on the Fully optimized portfolios. Figure 1 gives an insight into the causes of these deviations. The Expected returns portfolios series presents a set of unexpected high returns, especially the one in April 2009 of more than 110%. In the same month, the Fully optimized portfolio had a return of approximately -86%.

In order to understand the cause of these unexpectedly high and low portfolio returns, I chose a subset of the biggest and lowest monthly stock returns in April 2009. For these stocks, I used another database (Yahoo Finance) and confirmed that the Return Indices data retrieved from Datastream is correct. After that, I also searched for events that could explain big changes in Prices for each of these firms. For some of them, there were key events during this month, for example, the announcement of positive results on a clinical trial by a pharmaceutical company, the announcement of an acquisition of more hotel units by a resort company, or the announcement of the expansion of a company to a new country. However, there is no clear answer to the question of which was the cause for these unexpected returns: the volatility of the period, data issues, or both.

To a certain degree, these issues could be diminished by imposing constraints on weights. For instance, Lyle and Yohn (2021) construct portfolios with weights constrained to between -2.5% and 2.5% and 0% and 2.5% for long-only portfolios. I do not follow this approach as I have delimited my analysis to quadratic programming and imposing this kind of constraint involves other classes of optimization algorithms. With such a large number of assets, using these algorithms would be extremely time-consuming and there would always be the possibility of not achieving convergency when applying the model.

Figure 1: Portfolio Excess returns versus Market excess returns: full sample period

EW = Equal-weight portfolio; COV = Covariance portfolio; ER = Expected returns portfolio; FOP = Fully optimized portfolio



5.2.2 Multifactor models: alphas and information ratios

Other measures of performance include alphas and information ratios based on factor models. Figure 1 presents the times series of out-of-sample excess returns of each of the portfolios created versus the market excess return. The alpha in the single-factor model is a measure of the differences in these series across time.

Table 7 presents a summary of the alphas and information ratios of the Fully optimized portfolios found using different factor models. For the portfolios with the assumption of no short selling, the alpha values are always positive and statistically significant for the 0.1% level and range between 0.007 and 0.008. The associated Information ratios range between 0.290 and 0.396.

Comparing these results with the original work of Lyle and Yohn (2021), the signal and significance of alphas are identical. However, both the value of alphas and information ratios are below the original results. It is important to note, however, that the authors do this analysis based on portfolios with weights constrained between -2.5% and 2.5% (or 0% and 2.5% for long-only portfolios) and, as stated previously, I do not impose these constraints.

Regarding the portfolios with no restrictions on short sales, the alphas are positive in all models, however with different statistical significance levels (10% and 5% depending on the model). The information ratios are lower than the ones for the previous case, ranging between 0.121 and 0.154. Lyle & Yohn (2021) achieved positive statistically significant alphas for these portfolios across all models, and the alpha's values and information ratios were superior to the ones found for long-only portfolios.

Table 7: Factor models: alphas and Information ratios of Fully optimized portfolios

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	0.010*	0.141	0.007***	0.290
3 factor	0.010+	0.137	0.008***	0.370
4 factor	0.008+	0.121	0.008***	0.377
5 factor	0.011*	0.154	0.008***	0.388
6 factor	0.009+	0.139	0.008***	0.396

5.3 Comparing nested portfolios: January 2008 to December 2017

In order to better understand if the differences in results with the original study are driven by differences in the data sample or caused by the use of different time periods, I provide an analysis of the results for the only exact subperiod provided by the authors that can be replicated with my data – from January 2008 to December 2017. However, it is important to mention that also in this case the authors only provide the results for portfolios with weights constrained to between -2.5% and 2.5% (or 0% and 2.5% for long-only) and I do not impose these restrictions.

5.3.1 Average returns, standard deviations, Sharpe ratios and Treynor ratios

As presented in Table 8, the results are similar to the ones found in the entire sample period case. All the Sharpe ratios decreased in this example, especially the Sharpe ratio for Fully optimized portfolios. This was expected since the biggest contributor month (April 2009) to the poor results for this portfolio was included in the subperiod. Once again, the biggest outlier in the Treynor ratio comparison are the long-short Fully optimized portfolios due to a low beta.

For Lyle and Yohn (2021), this subperiod portfolios produced lower returns (except the long-only Covariance portfolios) and higher standard deviations (except the Fully optimized portfolios) compared to the entire sample. In consequence, the Sharpe ratios for this period were smaller than the ones found for the entire period, as in this study.

This comparison indicates that the data used in each study and the slight differences in methodology have an impact on the magnitude of expected returns, standard deviations and Sharpe ratios obtained. However, it also shows that when faced with specific market environments, the trends in portfolio statistics are similar.

Table 8: Nested Portfolio analysis (January 2008 to December 2017)

EW = Equal-weight portfolio; COV = Covariance portfolio; ER = Expected returns portfolio; FOP = Fully optimized portfolio

Mean and Standard deviation are shown as a percentage.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel A: Portfolio summary statistics – Average returns, standard deviations, Sharpe ratios and Treynor ratios

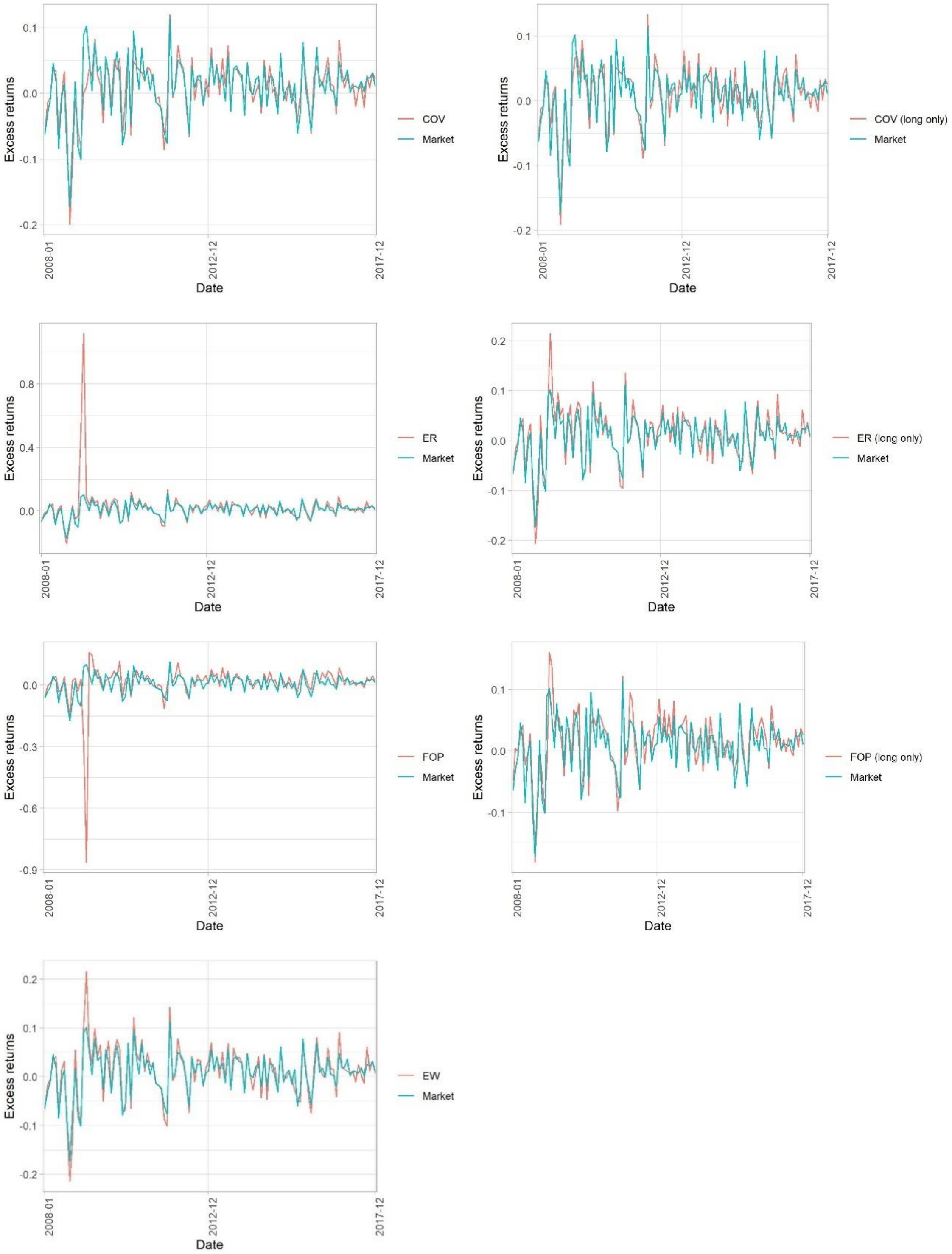
		Long-short			Long-only		
	EW	COV	ER	FOP	COV	ER	FOP
Mean	1.117	0.867	2.489	1.290	1.040	1.325	1.572
SD	5.657	4.249	12.073	9.525	4.459	5.549	4.809
SR	0.193	0.198	0.204	0.133	0.228	0.234	0.322
TR	0.009	0.010	0.015	0.062	0.011	0.011	0.017

Panel B: Differences in Sharpe ratios

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW
Long-Short	0.005	0.011	0.006	-0.065	-0.071	-0.060
Long-Only	0.034	0.041**	0.007	0.094*	0.087*	0.128*

Figure 2: Portfolio Excess returns versus Market excess returns: 2008-2017

EW = Equal-weight portfolio; COV = Covariance portfolio; ER = Expected returns portfolio; FOP = Fully optimized portfolio



5.3.2 Multifactor models: alphas and information ratios

Similarly to Figure 1, Figure 2 gives an overview of the time series of the excess returns of each portfolio versus the market, but, in this case, it focuses on the period between 2008 and

2017. Regarding the alphas and information ratios for this period, the long-only Fully optimized portfolios maintain their positive and significant alphas and information ratios in the same ranges as before. However, for the long-short portfolios, the alpha is only statistically significant and positive in the six-factor model at a 5% significance level (Table 9).

Table 9: Factor models: alphas and Information ratios of Fully optimized portfolios (January 2008 to December 2017)

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	0.011	0.116	0.008***	0.332
3 factor	0.010	0.108	0.008***	0.373
4 factor	0.010	0.125	0.008***	0.377
5 factor	0.014	0.161	0.009***	0.394
6 factor	0.015*	0.199	0.009***	0.396

5.1 Comparing nested portfolios: Dividing the sample into two 9-year periods

To analyze possible differences in performance across other periods, I now divide the entire period into two subperiods of nine years each: from October 2003 to September 2012 and from October 2012 to September 2021. As the entire period was eighteen years and three months, July, August and September 2003 were suppressed from this analysis.

5.1.1 Average returns, standard deviations, Sharpe ratios and Treynor ratios

In the first period under analysis, Sharpe ratios are generally low, especially for the long-short Fully optimized portfolios (5.9%). On the other hand, the highest Sharpe ratio, 22.1%, is found in the long-only Fully optimized portfolios. The first period is marked by lower average returns and higher standard deviations (Table 10). These results were expected since the 2008 financial crisis occurred during it.

In this subperiod, there are no significant differences in Sharpe ratios between portfolios except in two cases: (1) long-only Expected return portfolios have higher Sharpe ratios than Equal weighted portfolios and (2) long-only Fully optimized portfolios have higher Sharpe ratios than long-only Expected return portfolios. These findings suggest that taking into account differences in Expected returns in portfolio construction conducts in gains in performance.

Table 10: Nested Portfolio analysis over time

EW = Equal-weight portfolio; COV = Covariance portfolio; ER = Expected returns portfolio; FOP = Fully optimized portfolio
Mean and Standard deviation are shown as a percentage.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel A: Portfolio summary statistics – Average returns, standard deviations, Sharpe ratios and Treynor ratios

Panel A1: October 2003 to September 2012

		Long-short			Long-only		
	EW	COV	ER	FOP	COV	ER	FOP
Mean	1.144	0.794	2.570	0.741	0.890	1.283	1.288
SD	5.881	4.536	12.715	10.077	4.712	5.811	5.183
SR	0.170	0.144	0.191	0.059	0.158	0.196	0.221
TR	0.008	0.007	0.014	0.027	0.008	0.009	0.012

Panel A2: October 2012 to September 2021

		Long-short			Long-only		
	EW	COV	ER	FOP	COV	ER	FOP
Mean	1.345	1.318	2.073	2.247	1.479	1.552	1.969
SD	5.125	4.025	6.521	4.138	4.125	5.026	4.322
SR	0.253	0.315	0.310	0.531	0.346	0.299	0.444
TR	0.011	0.014	0.017	0.029	0.016	0.013	0.021

Panel B: Differences in Sharpe ratios

Panel B1: October 2003 to September 2012

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW
Long-Short	-0.026	0.021	0.047	-0.084	-0.131	-0.111
Long-Only	-0.012	0.026*	0.038	0.062*	0.025	0.051

Panel B2: October 2012 to September 2021

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW
Long-Short	0.062	0.058	-0.004	0.216**	0.219	0.277**
Long-Only	0.093*	0.046**	-0.047	0.097**	0.144**	0.190**

The second period finds better performance results: average returns are higher for all portfolios, except for long-short Expected returns portfolios, and standard deviations decreased. This results in overall higher Sharpe ratios, especially for Fully optimized portfolios. In this case, the average return for long-short portfolios is 2.2% and for long-only portfolios is 2.0%. the standard deviations are 4.1% and 4.3% respectively. The Sharpe ratios are above 53% and 44% respectively. This was also expected since the major detractor month for the performance of Fully Optimized portfolios (April 2009) is not included in the period.

For this period, Equal-weight portfolios have significantly lower Sharpe ratios than all other portfolios and, in opposition, Fully optimized portfolios have significantly higher Sharpe ratios than the other portfolios, except for long-short Fully optimized portfolios versus Expected return portfolios. Contrarily to all the prior analyses, the results for this period present a higher Sharpe ratio for the long-short than for the long-only Fully optimized portfolios.

Although Lyle and Yohn (2021) find that the performance gains of Fully optimized portfolios are persistent through time, these results show that the superior performance of these portfolios is dependent on the time period, especially for long-short portfolios.

5.1.2 Multifactor models: alphas and information ratios

The analysis of the alphas and Information ratios corroborates the previous analysis (Table 11). For the first period, for the long-short portfolios, the alphas are not statistically significant and the Information ratios are low. For the long-only portfolios, the alphas are positive and statistically significant at least at a 5% level and the information ratios are between 0.239 and 0.287.

For the second period, alphas are positive and statistically significant at a 0.1% level in all models (the exception is the long-only portfolio in the single factor model with significance at the 1% level) and are higher for the long-short portfolios. The Information ratios varied between 0.331 and 0.628.

Table 11: Factor models over time: alphas and Information ratios of Fully optimized portfolios

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Panel A: October 2003 to September 2012

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	0.005	0.049	0.006*	0.239
3 factor	0.004	0.045	0.006*	0.245
4 factor	0.003	0.033	0.006*	0.249
5 factor	0.012	0.129	0.007**	0.285
6 factor	0.012	0.168	0.007**	0.287

Panel B: October 2012 to September 2021

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	0.013***	0.441	0.008**	0.331
3 factor	0.015***	0.616	0.010***	0.543
4 factor	0.015***	0.628	0.010***	0.553
5 factor	0.015***	0.614	0.010***	0.542
6 factor	0.015***	0.625	0.010***	0.552

5.2 Comparing nested portfolios: cardinality constraints

This subsection aims of identifying the impact on ex-post performance when the sample is constrained to the X biggest stocks, with X, in this study, being 100, 200 and 500. I note that I do not constrain weights between any specific interval.

Table 12: Cardinality constraints – Nested Portfolio analysis

EW = Equal-weight portfolio; COV = Covariance portfolio; ER = Expected returns portfolio; FOP = Fully optimized portfolio
Mean and Standard deviation are shown as a percentage.
* p < 0.05, ** p < 0.01, *** p < 0.001

Panel A: Portfolio summary statistics – Average returns, standard deviations, Sharpe ratios and Treynor ratios

Panel A1: 100 stocks

		Long-short			Long-only		
	EW	COV	ER	FOP	COV	ER	FOP
Mean	1.130	0.833	1.143	100.597	0.945	1.228	0.980
SD	3.984	3.214	4.930	1454.947	3.154	4.187	3.687
SR	0.259	0.229	0.212	0.069	0.269	0.270	0.240
TR	0.016	0.021	0.014	-0.026	0.020	0.016	0.017

Panel A2: 200 stocks

		Long-short			Long-only		
	EW	COV	ER	FOP	COV	ER	FOP
Mean	1.167	0.857	2.071	-0.812	0.926	1.250	0.942
SD	4.277	3.133	9.728	0.247	3.181	4.464	3.499
SR	0.250	0.243	0.203	-0.037	0.261	0.258	0.242
TR	0.011	0.014	0.019	-0.034	0.014	0.011	0.012

Panel A3: 500 stocks

		Long-short			Long-only		
	EW	COV	ER	FOP	COV	ER	FOP
Mean	1.244	0.945	2.207	0.387	1.040	1.372	1.417
SD	4.808	3.139	10.410	24.709	3.490	4.934	4.181
SR	0.239	0.270	0.202	0.012	0.270	0.259	0.316
TR	0.010	0.014	0.025	-0.006	0.013	0.011	0.015

Panel B: Differences in Sharpe ratios

Panel B1: 100 stocks

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW
Long-Short	-0.030	-0.047	-0.017	-0.159	-0.143	-0.190
Long-Only	0.009	0.011	0.002	-0.029	-0.031	-0.020

Panel B2: 200 stocks

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW
Long-Short	-0.007	-0.047	-0.040	-0.279	-0.240	-0.286
Long-Only	0.010	0.008	-0.002	-0.019	-0.017	-0.009

Panel B3: 500 stocks

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW
Long-Short	0.031	-0.036	-0.067	-0.258	-0.191	-0.226
Long-Only	0.031	0.020	-0.011	0.045	0.057	0.076*

5.2.1 Average returns, standard deviations, Sharpe ratios and Treynor ratios

Table 12 presents the results for average returns, standard deviations and Sharpe ratios and each panel corresponds to the optimizations considering a different number of stocks in the sample.

Firstly, addressing the unusually large values in the long-short Fully optimized portfolios for the biggest 100 stocks, this happens because of extremely high weights on some months for stocks that performed poorly in those months. These portfolios have a Sharpe ratio of only 6.9% and a negative Treynor ratio.

Secondly, it is noticeable that, except for the Fully Optimized portfolios, the Sharpe ratios obtained when considering only a small number of stocks are similar or even higher than the ones obtained with an entire sample.

Table 13: Cardinality constraints – Factor models: alphas and Information ratios of Fully optimized portfolios

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Panel A: 100 stocks

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	1.534	0.107	0.002+	0.120
3 factor	1.376	0.097	0.002	0.114
4 factor	1.420	0.100	0.002	0.109
5 factor	1.315	0.093	0.001	0.076
6 factor	1.359	0.096	0.001	0.070

Panel B: 200 stocks

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	-0.011	-0.046	0.002+	0.123
3 factor	-0.013	-0.052	0.002	0.106
4 factor	-0.021	-0.100	0.002	0.105
5 factor	-0.008	-0.035	0.002	0.084
6 factor	-0.017	-0.080	0.001	0.083

Panel C: 500 stocks

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	0.007	0.030	0.006***	0.277
3 factor	0.004	0.015	0.005***	0.271
4 factor	-0.004	-0.017	0.005***	0.268
5 factor	0.006	0.026	0.005***	0.259
6 factor	-0.001	-0.005	0.005***	0.257

Finally, across the three experiments, Fully optimized portfolios in the long-short form perform poorly compared to their counterparts; however, these differences are not statistically significant. The Sharpe ratio for the long-only portfolios increases as the number of stocks included also increases, being 31.6% with a sample of the biggest 500 stocks. This last portfolio, at a 5% significance level, has a higher Sharpe ratio than the Equal-weight portfolio. Lyle and Yohn (2021) also find that as the sample size increases, the Sharpe ratio of both long-only and long-short portfolios increases. However, the results presented in this section do not allow for the conclusion provided by the authors that the full optimization performance gains are robust even when eliminating small capitalization firms from the investment set.

5.2.2 Multifactor models: alphas and information ratios

The multifactor analysis points to non-significant alphas across the three samples and all models for the Fully optimized long-short portfolios, accompanied by low Information ratios (Table 13). In the Long-only portfolios case, alphas are positive and become statistically significant as the sample extends to 500 stocks and Information ratios increase to around 0.25.

5.3 Portfolios in the extreme deciles

This subsection provides performance results of the applicability of different portfolio creation methodologies to the expected returns top extreme decile stocks. As explained in the methodology section, I perform a slightly different type of analysis than the traditional approach. I reduce the sample to only the stocks with expected returns in the top decile and create different portfolios with these stocks. In this sense, I remark that the long-short portfolios are not created by taking a long position in the top decile and a short position in the bottom decile; they take both long and short positions on stocks on the top decile.

5.3.1 Average returns, standard deviations, Sharpe ratios and Treynor ratios

If expected returns and covariances were a good estimate of actual returns and covariances, we would expect that portfolios created with the top predicted performers would have a better performance than the counterparts based on the total sample. Table 14 confirms this expectation. The average returns are around 2% for all portfolios and the standard deviation ranged between 5%

and 7%. Regarding the Treynor ratios, the biggest values are associated with the Fully optimized portfolios, both long-short and long-only.

There is also a large improvement in Sharpe ratios when compared with the total sample values. Fully optimized portfolios have Sharpe ratios above 40%. The lowest Sharpe ratio found is for the Value-weight portfolio (28.9%). Comparing Sharpe ratios across portfolios, it is found that there are no significant differences in this statistic between Expected returns and Covariance portfolios, Covariance and Equal-weight portfolios, and Fully optimized and Covariance portfolios. The three portfolios that involve optimization have higher Sharpe ratios than the Value-Weight portfolio. Finally, both Expected return portfolios and Fully optimized portfolios have higher Sharpe ratios than the Equal-weight portfolios. These results are in line with Lyle and Yohn's (2021) conclusion that Fully optimized portfolios outperform equal-weight and value-weight portfolios of stocks in the extreme decile of expected returns.

Table 14: Top decile portfolios – Nested Portfolio analysis

Panel A: Portfolio summary statistics – Average returns, standard deviations, Sharpe ratios and Treynor ratios

EW = Equal-weight portfolio; VW = Value-weight portfolio; COV = Covariance portfolio; ER = Expected returns portfolio; FOP = Fully optimized portfolio
Mean and Standard deviation are shown as a percentage.
* p < 0.05, ** p < 0.01, *** p < 0.001

			Long-short			Long-only		
	EW	VW	COV	ER	FOP	COV	ER	FOP
Mean	2.038	1.990	2.063	2.108	2.350	2.140	2.108	2.294
SD	5.674	6.543	5.127	5.694	5.370	5.299	5.694	5.445
SR	0.342	0.289	0.384	0.352	0.420	0.386	0.353	0.404
TR	0.017	0.015	0.021	0.017	0.024	0.021	0.017	0.023

Panel B: Differences in Sharpe ratios

Portfolio	COV-EW	ER-EW	ER-COV	FOP-COV	FOP-ER	FOP-EW	VW-COV	VW-ER	VW-FOP
Long-Short	0.042	0.011*	-0.030	0.036	0.066*	0.078*	-0.094*	-0.064*	-0.130**
Long-Only	0.043	0.011*	-0.032	0.018	0.050	0.061*	-0.096**	-0.064*	-0.114**

5.3.1 Multifactor models: alphas and information ratios

The Fully optimized portfolios also see a big performance improvement when measured by alphas and information ratios (Table 15). Alphas are positive and statistically significant across all models for both long-only and long-short portfolios (between 0.014 and 0.015). The associated Information ratios are also higher, ranging between 0.37 and 0.44.

Table 15: Top decile portfolios – Factor models: alphas and Information ratios of Fully optimized portfolios

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	Long-short		Long-only	
	alpha	IR	alpha	IR
1 factor	0.014***	0.390	0.014***	0.372
3 factor	0.015***	0.435	0.014***	0.437
4 factor	0.015***	0.432	0.014***	0.443
5 factor	0.015***	0.443	0.014***	0.444
6 factor	0.015***	0.441	0.014***	0.452

6. Conclusion

This dissertation had the objective to test if incorporating expected returns and covariances estimated based on firm characteristics in Markowitz style optimization would provide any gains in performance. I found that these gains are dependent on several matters, such as the database used, the time period under analysis, and the number of stocks used for the construction.

If all the stocks of the US market are considered during a period between 2003 and 2021, there is no evidence that Fully optimized portfolios have a higher Sharpe ratio than all the other portfolios under analysis. However, if a restriction of no short sales is applied, Fully optimized portfolios have a higher Sharpe ratio than the nested portfolios. When performance is measured based on factor models, alphas are positive and statistically significant for both long-short and long-only fully optimized portfolios.

If the sample period is divided into two sections (October 2003 - September 2012 and October 2012 - September 2021), the first subperiod is characterized by generality low returns and Sharpe ratios and high standard deviations. In the second period, Fully optimized portfolios provide high Sharpe ratios and positive and statistically significant alphas. Their Sharpe ratios are higher than the nested portfolios ones.

When testing the possibility of obtaining improved performance by using a number of stocks with the highest capitalization, Fully optimized portfolios show poor results, with low Sharpe ratios. Only with 500 stocks and no short sales allowed the alphas are positive and statistically significant.

If the sample is divided into deciles by expected returns, and the portfolios in this analysis are constructed in the top extreme decile, the performance of Fully optimized portfolios is higher in terms of Sharpe ratios and positive significant alphas.

Generally, the authors' conclusion that Covariance portfolios outperform Equal-weight portfolios is not validated by this dissertation. However, in almost all cases under analysis, in a long-only setting, portfolios that incorporate only estimates of expected returns based on stock characteristics outperform Equal-weight portfolios, as in the original study. Another similar conclusion is that portfolio performance between Expected return portfolios and Covariance portfolios is not statistically different.

The most relevant difference between the results found in the two studies lies in the performance of long-short Fully optimized portfolios. In this dissertation, in the majority of cases, they do not significantly outperform all other portfolios under test – this only happens in the period between October 2012 and September 2021. Regarding the long-only portfolios, they generally perform better than the remaining portfolios and have positive alphas.

These findings suggest that mean-variance optimization based on expected returns and covariance matrix estimated with stock characteristics brings improvements to portfolio performance, specifically with restrictions on short sales. However, these results are dependent on the time period under analysis and the inclusion of small stocks in the investment universe, as there is no evidence that the higher performance is not driven by illiquid stocks or other anomalies.

There are several points worth mentioning regarding these results. First, regarding the database, although a consistent and in-depth data cleaning process was performed, it is likely that the data used still displays some differences when compared to data from CRSP and Compustat. As discussed previously, this can have an impact on the results. Additionally, I had to combine the information on accounting variables with annual and quarterly frequency which was not the case in the study of Lyle and Yohn (2021). Suggestions for future research could be (1) testing if the use of only annual data would improve results; (2) adding weight constraints to the portfolio construction; and (3) extending the analysis to other countries.

References

Allen, D., Lizieri, C., & Satchell, S. (2019). In Defense of Portfolio Optimization: What If We Can Forecast? *Financial Analysts Journal*, *75*(3), 20–38. https://doi.org/10.1080/0015198X.2019.1600958

Amenc, N., Goltz, F., Lodh, A., & Martellini, L. (2014). Towards smart equity factor indices: Harvesting risk premia without taking unrewarded risks. *Journal of Portfolio Management*, *40*(4), 106–122. https://doi.org/10.3905/jpm.2014.40.4.106

Baker, N. L., & Haugen, R. A. (1996). Commonality In The Determinants Of Expected Stock Returns. *Journal of Financial Economics*, *41*(3), 401–439.

Bender, J., Blackburn, T., & Sun, X. (2019). Clash of the titans: Factor Portfolios versus Alternative Weighting Schemes. *The Journal of Portfolio Management*, *45*(3), 38–49. https://doi.org/10.1525/bio.2012.62.11.8

Brandt, M. W., Santa-Clara, P., & Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, *22*(9), 3411–3447. https://doi.org/10.1093/rfs/hhp003

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509–1531. https://doi.org/10.1093/rfs/hhm055

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Clarke, R., De Silva, H., & Thorley, S. (2016). Fundamentals of efficient factor investing. *Financial Analysts Journal*, *72*(6), 9–26. https://doi.org/10.2469/faj.v72.n6.3

Elton, E. J., & Gruber, M. J. (2011). *Mutual Funds*. https://doi.org/https://dx.doi.org/10.2139/ssrn.2088418

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. (2006). Evaluation of asset managers. *Cadernos Do Mercado de Valores Mobiliários*, *24*, 10–33.

Fama, E. F., & French, K. R. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, *82*(3), 491–518. https://doi.org/10.1016/j.jfineco.2005.09.009

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252. https://doi.org/10.1016/j.jfineco.2018.02.012

Fama, E. F., French, K. R., Booth, D. G., & Sinquefield, R. (1993). Differences in the Risks and Returns of NYSE and NASD Stocks. *Financial Analysts Journal*, *49*(1), 37–41. https://doi.org/10.2469/faj.v49.n1.37

Goodwin, T. H. (1998). The information ratio. *Financial Analysts Journal*, *54*(4), 34–43. https://doi.org/10.2469/faj.v54.n4.2196

Green, J., Hand, J. R. M., Soliman, M. T., Green, J., & Soliman, M. T. (2011). Going , Going , Gone ? The Apparent Demise of the Accruals Anomaly. *Management Science*, *57*(5), 797–816. https://doi.org/10.1287/mnsc.llI0.1320

Hirshleifer, D., Hou, K., Teoh, S. H., & Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, *38*(1-3 SPEC. ISS.), 297–331. https://doi.org/10.1016/j.jacceco.2004.10.002

Hou, K., Mo, H., Xue, C., & Zhang, L. (2021). An Augmented q -Factor Model with Expected Growth. *Review of Finance*, *25*(1), 1–41. https://doi.org/10.1093/rof/rfaa004

Hou, K., Xue, C., & Zhang, L. (2014). A Comparison of New Factor Models. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2520929

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. https://www.jstor.org/stable/2328882

Jensen, M. C. (1968). The Performance of Mutual Funds in the Period 1945-1964. *The Journal of Finance*, *23*(2), 389–416. https://www.jstor.org/stable/2325404

Karpoff, J. M., & Lou, X. (2010). Short sellers and financial misconduct. *Journal of Finance*, *65*(5), 1879–1913. https://doi.org/10.1111/j.1540-6261.2010.01597.x

Landis, C., & Skouras, S. (2021). Guidelines for asset pricing research using international equity data from Thomson Reuters Datastream. *Journal of Banking and Finance*, *130*, 106128. https://doi.org/10.1016/j.jbankfin.2021.106128

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859. https://doi.org/10.1016/j.jempfin.2008.03.002

Ledoit, O., & Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies*, *30*(12), 4349–4388. https://doi.org/10.1093/rfs/hhx052

Ledoit, O., & Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Annals of Statistics*, *48*(5), 3043–3065. https://doi.org/10.1214/19-AOS1921

Lee, C. M. C., Myers, J., & Swaminathan, B. (1999). What is the intrinsic value of the dow? *Journal of Finance*, *54*(5), 1693–1741. https://doi.org/10.1111/0022-1082.00164

Lyle, M. R., & Yohn, T. L. (2021). Fundamental Analysis and Mean-Variance Optimal Portfolios. *The Accounting Review*, *96*(6), 303–327. https://doi.org/10.2308/tar-2019-0622

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. http://links.jstor.org/sici?sici=0022-1082%28195203%297%3A1%3C77%3APS%3E2.0.CO%3B2-1

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2010). *Out-of-Sample Equity Premium Prediction: Fundamental vs . Technical Analysis* (No. 3063). http://ink.library.smu.edu.sg/lkcsb_research/3063

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, *104*(1), 162–185. https://doi.org/10.1016/j.jfineco.2011.12.009

Richardson, S. A., Sloan, R. G., Solimanc, M. T., & Tuna, I. (2005). Erratum: (Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, *93*(3), 437–485. https://doi.org/10.1016/j.jacceco.2019.101235

Richardson, S., Tuna, I., & Wysocki, P. (2010). Accounting anomalies and fundamental analysis: A review of recent research advances. *Journal of Accounting and Economics*, *50*(2–3), 410–454. https://doi.org/10.1016/j.jacceco.2010.09.008

Sharpe, W. F. (1966). Mutual Fund Performance. *Journal of Business*, *39*(1), 119–138. https://www.jstor.org/stable/2351741

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, *71*(3), 289–315. https://doi.org/10.2469/dig.v27.n1.5

Appendices

A. Datastream and Worldscope Datatypes: Code, Name and Description

1. Static datatypes

Datatype	Name	Definition
TYPE	Type of Instrument	Type of instrument
DSCD	Datastream Code	Unique six-digit code allocated by Datastream
BDATE	Base Date	Date from which Datastream holds information about the issue
ENAME	Expanded Security Name	Expanded (unabbreviated) name of the quote
EXMNEM	Datastream Exchange Mnemonic	Datastream exchange code based on the ISO standard exchange code. The code is a three-character alpha mnemonic identifying the source of the default price datatypes for a given equity
GEOGN	Geographical Classification of company	Home or listing country of a security
ISIN	ISIN	International Security Identification Number
ISINID	Primary Indicator Flag	Returns either P or S – P indicates that the equity record is primary (domestic listing of the share), and S indicates that the equity record is secondary (foreign listing of a share)
LOC	Code - Local	Identification code based on the official local exchange code. It comprises up to 12 characters, prefixed by an alphabetic country code
PCUR	Currency	Currency in which the price of a security is quoted and displayed
TRAC	Security Type Code	Type of share as defined by the Thomson Reuters classification system
TIME	Time – Latest value	Date or time of the latest equity price data

2. Time series datatypes

AF	Adjustment Factor (Accumulated)	Datastream factor accumulated to stock base date.
RI	Total Return Index	A theoretical growth in value of a share holding over a specified period, assuming that dividends are re-invested to purchase additional units of an equity or unit trust at the closing price applicable on the ex-dividend date
UP	Unadjusted Price	Closing price which has not been historically adjusted for bonus and rights issues
P	Price – Trade	Official closing price. Unadjusted Price adjusted for subsequent capital actions
MV	Market value (Capital)	Share price multiplied by the number of ordinary shares in issue

3. Wordescope datatypes

WC02999	Total assets	Sum of total current assets, long-term receivables, investment in unconsolidated subsidiaries, other investments, net property plant and equipment and other assets.
WC03351	Total liabilities	All short- and long-term obligations expected to be satisfied by the company
WC04101	Deferred Income Taxes & Investment Tax Credit	Increase or decrease in the deferred tax liability from one year to the next resulting from timing differences in recognition of revenues and expenses for tax and financial reporting purposes
WC03451	Preferred Stock	Claim prior to the common shareholders on the earnings of a company and on the assets in the event of liquidation
WC04199	Deferred Income Taxes (Cash Flow)	Increase or decrease in the deferred tax liability from one year to the next resulting from timing differences in recognition of revenue and expenses for tax and financial reporting purposes
WC01551A	Net Income Before Extraordinary Items/Preferred Dividends	Income before extraordinary items and preferred and common dividends, but after operating and non-operating income and expense, reserves, income taxes, minority interest and equity in earnings
WC02001	Cash & Short-Term Investments	Sum of cash and short-term investments
WC03051	Short-Term Debt & Current Portion of Long-Term Debt	Portion of debt payable within one year including current portion of long-term debt and sinking fund requirements of preferred stock or debentures
WC03251	Long-Term Debt	All interest-bearing financial obligations, excluding amounts due within one year, net of premium or discount.
WC03426	Minority Interest	Portion of the net worth (at par or stated value) of a subsidiary pertaining to shares not owned by the controlling company and its consolidated subsidiaries
WC03501	Common Equity	Common shareholders' investment in a company
WC02008	Short–term investments	Temporary investments of excess cash in marketable securities that can be readily converted into cash
WC02258	Long-Term Receivables	Amounts due from customers that will not be collected within the normal operating cycle of the company
WC02250	Other Investments	Any other long-term investment except for investments in unconsolidated subsidiaries