A CONSISTENT WASTE OF TIME?

THE IMPACT OF CONSISTENT VERSUS INCONSISTENT PROFITABILITY RATIOS IN FACTOR PORTFOLIO CONSTRUCTION

EMIL STÅHL

LUDVIG LJUNGBERG

Master Thesis

Stockholm School of Economics

2025

A Consistent Waste of Time? The Impact of Consistent versus Inconsistent Profitability Ratios in Factor Portfolio Construction

Abstract

This thesis investigates whether the use of consistent versus inconsistent definitions of firm profitability affects the performance of profitability factor portfolios. A consistent profitability ratio matches the numerator and denominator in terms of financial perspective, for example using net income which reflects returns to equity holders together with equity. Using 60 years of stock market data and six profitability measures, including benchmarks from Fama-French and Novy-Marx, we evaluate the impact of consistency on factor portfolio returns. Despite the theoretical expectation that consistent measures should outperform inconsistent ones, our analysis finds no significant difference in performance between consistent and inconsistent measures. Furthermore, we find that the benchmark measures with stronger links to cash flows delivered higher but statistically insignificant returns. These findings suggest that consistency, while theoretically sound, may not be essential for constructing effective profitability factors.

Keywords:

Accounting ratios, Consistency, Factor models, Profitability

Authors:

Emil Ståhl (24950)
Ludvig Ljungberg (24933)

Supervisor:

Henrik Andersson, Affiliated Researcher, Department of Accounting

Master Thesis
Master Program in Accounting, Valuation and Financial Management
Stockholm School of Economics
Emil Ståhl and Ludvig Ljungberg, 2025

Acknowledgements

We would like to thank our supervisor Henrik Andersson for suggesting that we should explore the use of consistent and inconsistent accounting ratios in factor models, as well as for his valuable support throughout the process. Furthermore, we are grateful to all the friends and teachers we have met during our five years at the Stockholm School of Economics.

Stockholm, May 2025

Emil Ståhl Ludvig Ljungberg

Table of Contents

1. INTRODUCTION 5
2. LITERATURE REVIEW 7
2.1 Accounting Ratios for Profitability 7
2.2 Factor Models 9
2.3 Factor Performance 10
2.4 Profitability Definitions 11
2.5 Fama-MacBeth Two-step Regression 13
2.6 Contribution 13
3 METHOD 14
3.1 Portfolio Creation 14
3.2 Fama-MacBeth Model 16
4 DATA 18
4.1 Data Summary 18
4.2 Data Cleaning Process 18
5 RESULTS 20
5.1 Descriptive Statistics 20
5.2 Portfolio Performance 23
5.3 Fama-MacBeth Results 29
6 DISCUSSION 30
7 CONCLUSION 32
7.1 Limitations 32
7.2 Future Research 32
8 REFERENCES 34
9 APPENDIX 37

1. Introduction

In financial analysis and asset pricing, accounting ratios are widely used to assess firm performance and inform investment decisions (Johansson & Runsten, 2023). These ratios are often treated as objective inputs, but in practice they vary significantly in how they are defined and constructed. One important dimension is whether the numerator and denominator are matched in terms of financial perspective, a property that is often referred to as consistency. We examine whether different definitions of firm profitability impact the returns of a profitability factor portfolio. Fama and French (2015) introduced the five-factor model for explaining stock returns. One of the factors, RMW, looks at the profitability of firms and assigns a risk premium to more profitable firms relative to less profitable ones. In their paper, Fama and French use a profitability measure they refer to as "operating profit". For this study, we examine how different definitions of profitability measures in factor models affect the results of the profitability factor. We further investigate whether the use of "consistent" versus "inconsistent" measures influences performance. Consistent measures are such that the numerator and denominator are matched. For example, if equity is used as the denominator, then net income would be a consistent numerator since it represents earnings attributable to equity owners. An example of an inconsistent measure would be to use equity as the denominator and earnings before interest and taxes (EBIT) as the numerator, since that belongs to both equity and debt holders.

The motivation for this research question is twofold. First, the use of inconsistent measures is widespread in accounting and finance literature. Measures such as Net Income to Assets to describe Return on Assets (ROA) can be found in investing models, for example Piotroski (2000), and as controls in regressions, see Cheng et al. (2016), for example. By evaluating the performance of consistent versus inconsistent measures in a profitability factor setting, we aim to assess whether the use of inconsistent profitability ratios is problematic. If such measures yield inferior results in our setting, the findings could have implications for their use in other contexts.

Second, although it is not the aim of this study to propose refinements to the profitability component of the Fama-French five-factor model, our results could provide the basis for further research into alternative profitability measures to incorporate into the model. In this study, we isolate the profitability factor and will not control for the four other factors. However, if a particular definition of profitability consistently outperforms the Fama-French operating profitability measure, it might be valuable to evaluate its performance in the full five-factor model. Given that the Fama-French model is widely used to explain the returns of stock portfolios in both academic and practical applications, any improvement to its components could be meaningful.

The profitability factor in the Fama-French model is constructed as a long-short, self-financing portfolio. Each year, firms are ranked based on their profitability, and the portfolio takes a long position in the most profitable firms and a short position in the least profitable ones. The return of this portfolio over time represents the estimated risk premium associated with investing in more profitable firms.

In this study, we use 60 years of financial data on firms listed on the NYSE, AMEX, and NASDAQ stock exchanges to evaluate the profitability factor using six different definitions of profitability. We test two consistent measures, two inconsistent measures, and two measures from the academic literature, including the Fama-French operating profitability and Novy-Marx's (2013) gross profitability. The Fama-French and Novy-Marx measures will be used as benchmarks and will not be part of the consistent versus inconsistent comparison. Including the Fama-French measure allows us to validate our results, as our sample, time period, and methodology are like those used in Fama and French (2015).

The research question of this study is whether the use of consistent or inconsistent definitions of profitability affects the performance of the profitability portfolio. Our hypothesis is that consistent measures, given their theoretical justification, should outperform inconsistent ones. If there is a risk premium associated with investing in profitable firms, then a more consistent definition of profitability might better capture that risk. If such a result is observed, it could have broader implications for the application of profitability ratios in other research and practical contexts.

The main finding of this study is that there is no significant difference in performance between the consistent and inconsistent profitability measures. Across both the main analysis as well as the several robustness checks, we observe no consistent pattern of outperformance by either group. The best-performing measures are the two benchmark definitions from Fama and French and Novy-Marx. Additionally, we find that the profitability measures, both consistent and inconsistent, are highly correlated. This suggests that replacing an inconsistent profitability measure with a consistent one in a regression setting would likely not lead to notable improvements or meaningful changes in results. The results do not mean that all profitability measures are interchangeable in every setting. For instance, an EBIE to equity measure would not be appropriate in a residual income valuation model (Penman, 2013).

Given the lack of significant results in the portfolio analysis, we also conducted a Fama-MacBeth regression to estimate factor returns with potentially greater precision. This analysis did not produce any statistically significant coefficients either. Several robustness checks were performed, including more aggressive data cleaning through the exclusion of extreme values, testing alternative portfolio configurations, and analyzing the top and bottom portfolios separately. None of these tests materially altered the results.

2. Literature Review

2.1 Accounting Ratios for Profitability

Accounting can be broadly defined as the process of providing information about the economic activities of an accounting entity that is responsible for managing resources on behalf of others (Whittington, 2017). Businesses entities usually prepare financial statements at least once a year to give important financial information to owners and those who provide funding, as well as to governments for taxation purposes. These statements can impact matters such as remuneration for directors, share prices, dividends, taxation, and management decisions within the entity. This highlights the importance of financial statements, as they impact the economic welfare of individuals, business entities, and governments.

To evaluate the performance of firms, one can look at measures of profitability (Johansson & Runsten, 2023). Accounting ratios provide insight into both a firm's current performance and its potential future trajectory (Whittington, 1980). Absolute profit figures are not informative on their own (Johansson & Runsten, 2023). Instead, profitability must be evaluated in relation to another financial metric. A commonly used accounting ratio for this purpose is Return on Equity (ROE), which is calculated by dividing net income by the book value of shareholders' equity. This ratio is grounded in the proprietary theory, which views interest payments as expenses, not distributions of profit (Schroeder et al., 2023). A firm can be considered profitable if its ROE exceeds the required Return on Equity (Johansson & Runsten, 2023). ROE also facilitates comparisons of profitability across firms.

It is important to recognize that ROE is influenced by a firm's capital structure. For instance, if the proportion of debt relative to equity increases while net income remains constant, ROE will rise. To evaluate a firm's operational profitability independently of its financing decisions, Return on Assets (ROA) is more appropriate. According to Johansson & Runsten (2023), ROA is best measured as earnings before interest expenses (EBIE) divided by total assets. The ROA calculation combines operating and financing activities, making the ratio independent of the firm's financing mix (Penman, 2013). Interest income, which comes from financing activities, is included in the numerator, while the denominator, total assets, includes both operating and financial assets. As a result, ROA reflects a mix of returns from core operations and usually lower returns from excess cash or financial investments.

When constructing accounting ratios, consistency between the numerator and denominator is essential to accurately reflect what return the capital yields (Johansson & Runsten, 2023). ROE is consistent since the net income represents returns to equity holders, while equity represents the capital they have invested. Hence, ROE encapsulates

the return of investors' investments. Similarly, calculating ROA as EBIE divided by total assets is also consistent. EBIE reflects earnings available to both debt and equity holders, just as total assets are funded by both. In contrast, defining ROA as net income divided by total assets introduces inconsistency. This approach compares earnings attributable only to equity holders against a capital base financed by both debt and equity, leading to understated profitability. For highly leveraged firms, this can be particularly misleading, as it understates the return on total assets by incorporating interest-reduced earnings while still accounting for both debt and equity in the asset base.

In practice, however, academic studies often use inconsistent formulations of accounting ratios. For example, a wide range of numerators is used when calculating ROA. Barber & Lyon (1996) use operating profit, Piotroski (2000), and Cheng et al. (2016) use net income before extraordinary items, Erhardt et al. (2003) use net income, and Cheng (2008) uses earnings before interest and taxes. One explanation as to why inconsistent accounting ratios keep prevailing can be the availability of data. When collecting data from websites such as WRDS, data of firm's net income is widely available (WRDS, n.d.). This is not the case for other accounting information, such as EBIE. For a scholar who wants to use an accounting ratio as a control variable, for example, one hundred percent consistency may not be top of mind.

However, even accounting textbooks that do not have any data collection problems may define ratios inconsistently. Schroeder et al. (2023) define the ROA as net income divided by assets in their widely used accounting textbook, although they note that some financial analysts have recently started to deduct interest expense after tax in the numerator to make the ratio independent of the firm's capital structure. This definition would give ROA after tax, which differs from Johansson & Runsten's (2017) definition of ROA, which is before tax.

Although the implications of using consistent versus inconsistent financial ratios are not extensively studied in accounting research, similar issues have been explored in the context of valuation. When valuing firms, valuation multiples can be used (Goedhart et al., 2020). By expressing a firm's value as a multiple of a relevant financial metric, one can estimate the value based on how similar firms are priced in the market. Similarly to accounting ratios, valuation metrics can be consistent or inconsistent. An inconsistent metric might compare equity value to a pre-interest metric such as EBITDA, conflating equity and enterprise values.

Although this paper focuses on factor models and performance rather than valuation, insights from the literature about the impact of consistency in multiples on valuation accuracy could serve as empirical evidence of whether consistency in financial ratios matter. In "Does Consistency Improve Multiple Quality" (Chullen et al., 2013), the authors suggest that consistent definitions of valuation multiples result in significantly higher accuracy compared to inconsistent ones, based on a large sample of German firms.

This provides empirical support for the importance of consistency, an insight we aim to apply to the study of profitability ratios. This paper aims to contribute similar empirical evidence on the importance of consistency when it comes to profitability measures and accounting ratios.

2.2 Factor Models

Understanding the relationship between risk and expected return is fundamental when evaluating investments (Berk & Demarzo, 2024). The Capital Asset Pricing Model (CAPM) is the most widely used framework for explaining this relationship. According to CAPM, securities or asset classes with higher beta with respect to the market portfolio have higher expected returns. However, these higher expected returns are also linked to a greater risk of underperforming during bad times. Securities or asset classes with high betas typically perform worse in downturns compared to those with low betas. The assumptions behind CAPM include that investors are rational and hold efficient portfolios, that markets are perfect with no friction, and that all investors have identical expectations about returns and risks. Under these conditions, markets reach equilibrium, where expected returns align with each asset's systematic risk (measured by beta), and no asset is mispriced.

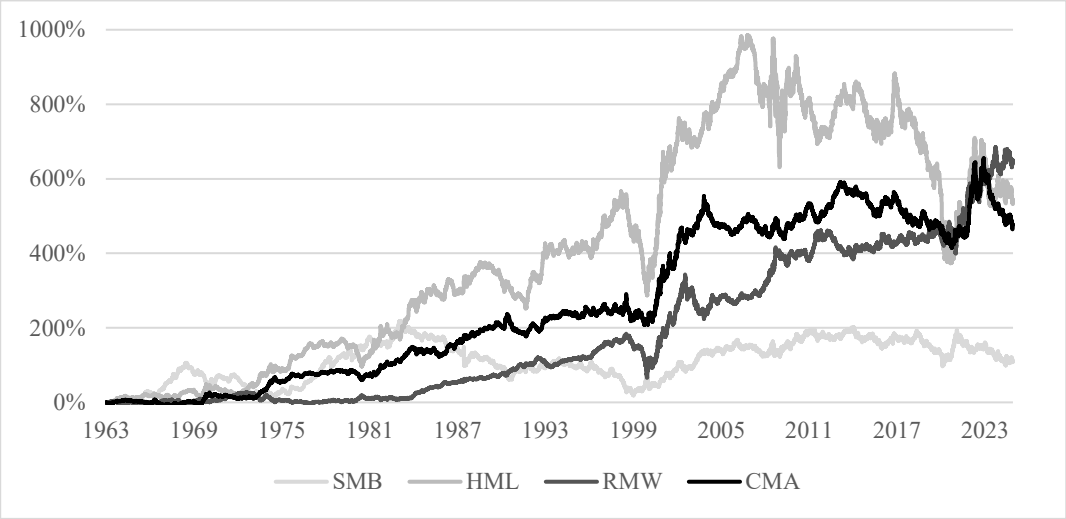
While CAPM includes only a single factor, the market risk factor, researchers have developed multi factor models to better capture the variation in stock returns (Berk & Demarzo, 2024). These models include multiple portfolios to capture risk rather than just one. They allow investors to decompose the risk premium into distinct risk factors, enabling them to tailor their portfolios based on their individual risk exposures and preferences. One such model is the Fama-French Three-Factor Model, introduced by Eugene Fama and Kenneth French in 1993, which extends the traditional CAPM by incorporating two additional factors to better explain stock returns (Fama & French, 1993). In addition to the market risk factor (β), the model includes a size factor (SMB, or small minus big), which captures the empirical observation that small-cap stocks tend to outperform large-cap stocks. It also introduces a value factor (HML, or high minus low), which reflects the historical tendency of value stocks with high book-to-market ratios to outperform growth stocks with low book-to-market ratios.

In 2015, Fama and French expanded their model into a Five-Factor Model by adding two additional factors. The profitability factor (RMW, or robust minus weak) suggests that firms with high profitability tend to outperform those with low profitability. The investment factor (CMA, or conservative minus aggressive) indicates that firms with conservative investment strategies achieve higher returns than those with aggressive strategies. The Five-Factor Model offers improved explanatory power by more accurately capturing firm-specific characteristics. However, the value factor becomes less significant in the extended model, as its explanatory role overlaps with the newly added profitability and investment factors.

2.3 Factor Performance

All five Fama-French factors delivered positive performance between 1963 and 2009 (Blitz, 2020). However, during the decade from 2010 to 2019, only the market and profitability factors produced positive returns. The size and investment factors generated returns close to zero, while the value factor experienced significantly negative performance. This pattern of underperformance is like what was observed during parts of the 1990s, highlighting that factor investing outcomes can vary over time and are not constant.

Graph 1. Cumulative returns of factors in the Fama-French 5 factor portfolio (excluding market risk factor). Source: Kenneth French's (2025) data library.



The graph above illustrates the cumulative returns of investing in each factor portfolio (SMB, HML, RMW, and CMA), except for the market risk factor, from July 1st 1963 to December 31st 2024. When calculating factor returns, Fama and French first split firms into two size groups, small and large, based on market capitalization (French, 2025). Within each group, they form two portfolios based on the target factor, going long on the group with high values and short on the group with low values. The factor return is the average return of the two long portfolios minus the average return of the two short portfolios. These are self-financing long-short strategies, requiring no initial net investment. The graph highlights how outcomes from factor investing can vary significantly over time. We can observe that the RMW factor has delivered higher cumulative returns than the SMB, HML, and CMA factors over the sample period.

The efficient market hypothesis (EMH) gives that security prices always reflect the publicly available information (Fama, 1970). This rules out the consistent ability to earn returns in excess of equilibrium levels through trading based only on publicly available information. The assumption behind the EMH is that either investors are rational, or that any irrational behavior is either random and cancels out, or is offset by rational arbitrageurs who correct mispricing (Shleifer, 2000). It could be tempting to dismiss the

EMH based on Graph 1, where we see that investing in certain factors would provide positive returns. However, Fama and French (1993) interpret factors measures of fundamental riskiness of stocks. This view is often referred to as the risk-based view, arguing that factor premiums compensate investors for taking on risks that tend to surface during economic downturns or other adverse conditions (Ang, 2014). During bad times, like in the financial crisis, strategies based on factor investing tend to perform poorly. For value stocks, this underperformance may be attributed to their tendency to be financially distressed, highly leveraged, or exposed to uncertainty in future earnings (Chen & Zhang, 1998). Hence, factor risk premiums exist to compensate investors for these risks (Ang, 2014). If there is a risk factor associated with firm profitability, then one would expect that a correctly defined profitability measure would better capture this underlying risk than an inconsistent alternative. Firms with low profitability are typically more volatile, as they tend to be financially distressed and exhibit higher levels of firm-specific risk (Barinov, 2023).

Alternatively, the behavioral view suggests that factor returns result from systematic investor biases or persistent mispricing that are not fully corrected by arbitrage (Ang, 2014). Supporters of this view argue that certain types of assets are systematically overlooked due to investor psychology. For instance, value stocks are often seen as less exciting compared to growth stocks. This leads investors to bid up the prices of growth stocks, making them overpriced, while value stocks remain underpriced.

2.4 Profitability Definitions

The use of profitability measures to explain and predict stock returns is common. Piotroski (2000) introduced the F-score, a metric ranging from zero to nine, based on nine financial signals such as Return on Assets (ROA), changes in leverage, and changes in asset turnover. His analysis showed that taking a long position in firms with high F-scores and a short position in those with low scores generated significantly higher returns than investing in a broad portfolio of high book-to-market firms.

However, Piotroski's measure of profitability lacks consistency, as the numerator and denominator do not align. He defines ROA as net income (before extraordinary items) divided by total assets. This mixes an equity-based numerator with a denominator that represents both equity and debt, which introduces inconsistency.

Fama and French (2015) introduces the profitability factor with a risk premium associated with investing in companies with high "operating" profitability. They define the operating profitability as:

$$Operating\ Profitability = (Revenue - COGS - SG\&A - Interest\ Expense)/Equity$$

In contrast to Piotroski, this measure is consistent. The profit that is left after removing COGS, SG&A, and Interest expense belongs to the equity owners, which is what is used

in the denominator. The name of the measure is still a bit misleading since it is not a traditional measure of "operating" profitability. Typically, an operating margin would use revenue as the denominator (Berk & Demarzo, 2024).

Novy-Marx (2013) offers another consistent approach with his gross profitability measure, defined as:

$$Gross\ Profitability = (Revenue - COGS) / Total\ Assets$$

This measure is consistent because it ensures that both the numerator and denominator reflect values available to all investors (equity and debt holders), unlike Piotroski, who mixes different financial signals. Novy-Marx also finds that gross profitability is a strong predictor of stock returns, reinforcing the idea that more profitable firms tend to generate higher future returns.

The inconsistency in profitability definitions across these studies begs the question if the choice of measure significantly influences the results of factor-based investing strategies. A more consistent approach, such as Novy-Marx's gross profitability or Fama-French's operating profitability, may offer a cleaner and more reliable way to capture the relationship between profitability and stock returns, compared to a theoretically m'ore inconsistent approach such as Piotroski's.

For some purposes, it is clearly important to use consistent profitability measures, for example, in valuation models (Penman, 2013). In a Residual Income Valuation (RIV) model, residual income is calculated as:

$$(Return\ on\ common\ equity - Required\ return\ of\ equity) * Book\ value\ of\ common\ equity$$

If ROE was defined incorrectly, for example, by using earnings before interest expenses in the numerator rather than net income, the residual income calculated would not be the true excess profit for the shareholders, as it would include interest expenses. Using that definition of residual income in the valuation model would therefore not provide a correct estimation of the value of equity. When it comes to factor models, however, it has not been studied whether using consistent accounting ratios in factor models would yield better results. This study will look deeper into whether such a relationship exists.

Beyond consistency, studies have investigated whether cash flow measures outperform other measures. Ball et al. (2016) find that cash flow-based measures of operating profitability outperform operating profitability in factor models. They find that any increase in profitability driven solely by accruals shows no association with stock returns. Fama and French (2018) similarly find that models which include cash profitability, defined as operating profit minus accruals, have higher maximum squared Sharpe ratios. This supports the idea that cash profitability better captures cross-sectional return patterns.

2.5 Fama-MacBeth Two-step Regression

The Fama-MacBeth two-step regression method (Fama & MacBeth, 1973) is used in empirical asset pricing research, particularly for estimating risk premia and testing the cross-sectional relation between expected returns and risk factors. The method separates the estimation of factor loadings (betas) and risk premia into two distinct steps. In the first step, time-series regressions are run for each asset to estimate their exposures to underlying risk factors, in our case profitability. In the second step, cross-sectional regressions are performed in each time period, regressing realized returns on the estimated betas to obtain time-varying estimates of risk premia. The average of these second-stage coefficients is then interpreted as the estimate of the factor risk premia. In this study, the Fama MacBeth method is used as a complementary approach to the portfolio sorting analysis for estimating the returns to the profitability factor.

2.6 Contribution

With the use of Fama-French style profitability factor portfolios based on different profitability measures, we hope to contribute to the existing literature by examining whether the use of consistent profitability measures improves the empirical results compared to using inconsistent definitions. While prior research has extensively explored the role of profitability in asset pricing and firm performance evaluation, little attention has been given to the potential distortions caused by inconsistencies in how profitability is measured. By conducting our analysis within a specific case, we hope to provide direct evidence that consistent profitability definitions lead to improved results.

Beyond our specific case, the implications of our findings may extend to a wide range of accounting and finance research where profitability measures play a crucial role. Controlling for profitability is a common occurrence in accounting research models and if consistency in definitions of profitability improves results in our context, it is possible that similar improvements could be observed across other studies relying on profitability ratios.

3 Method

This study employs two complementary empirical approaches to examine how different definitions of profitability affect factor portfolio performance. First, six different long-short portfolios are formed, one based on each profitability measure. Their annual returns and risk characteristics are then recorded over a 60-year sample period. Second, the Fama-MacBeth (1973) two-step regression methodology is applied to estimate the relationship between firm-level exposure to these profitability factors and subsequent stock returns. Together, these methods allow us to test whether theoretically consistent profitability measures produce systematically stronger results than inconsistent ones.

3.1 Portfolio Creation

The six profitability measures were selected to capture a range of definitions commonly used in both academic literature and financial practice. ROE and ROA are classified as "consistent" due to alignment between their numerators and denominators. EBIT to Equity and Net Income to Assets are "inconsistent" by construction. In addition, two well-established profitability measures from the literature are included to serve as benchmarks: The Operating Profitability measure introduced by Fama and French (2015) and the Gross Profitability measure proposed by Novy-Marx (2013). These benchmark measures provide a basis for comparison when evaluating the performance of the consistent and inconsistent measures. All profitability measures are calculated using opening balances for the denominator. ROE for example is calculated as net income for year t divided by equity for year t-1.

The financial data source, Compustat, does not provide earnings before interest expenses (EBIE) directly. While EBIE can theoretically be approximated by adding interest income to EBIT, or removing interest expense from profits before tax, interest data was either unavailable or unreliable for a large portion of the sample. As a result, EBIT was used as a proxy in the ROA measure where EBIE would have been the more theoretically appropriate choice, due to its greater availability and consistency across firms. For this reason EBIT is also used in the inconsistent measure, EBIT to Equity.

Table 1. List of included measures and their definitions.

Measure	Definition	Abbreviation
ROE	Net Income / Equity	ROE
ROA	EBIT / Total Assets	ROA
Ebit to Equity	EBIT / Equity	EBIT / Eq.
Net Income to Assets	Net Income / Total Assets	NI / As.
Novy-Marx Ratio	Gross Profit / Total Assets	N-M
Fama-French Ratio	(Sales - COGS - SG&A - Interest Expense) / Equity	F-F

To evaluate the performance of these 6 profitability measures, we construct long-short portfolios based on firm-level rankings. Each year, using the latest publicly available accounting data, all companies in the sample are ranked by their score on the selected profitability measure. Portfolios are formed on July 1st each year, using accounting data from March 31st at the latest, allowing for a three-month lag to ensure the data is publicly available.

The baseline portfolio strategy follows a structure similar to the Fama-French RMW (Robust Minus Weak) factor. Firms are sorted according to the selected profitability measure, and the portfolio takes a long position in the top 30% of firms (those with the highest profitability) and a short position in the bottom 30% (those with the lowest profitability). Portfolios are held for one year and rebalanced annually based on updated financial data.

This approach differs somewhat from the original Fama-French methodology, which often employs more complex portfolio sorting procedures, such as additional sorting on size or forming multiple intersecting portfolios. In contrast, this study uses a single profitability sort across the entire sample without first dividing firms based on size or other firm characteristics.

To test the robustness of our findings, we also analyse portfolios constructed using alternative cutoffs:

- Long top 20% / Short bottom 20%
- Long top 40% / Short bottom 40%
- Long top 50% / Short bottom 50%

All portfolios are constructed to be zero-net-investment, meaning the long and short positions are weighted equally. The portfolios are value-weighted within both the top and bottom parts based on market capitalization, following the methodology of Fama and French (2015). This approach reduces portfolio volatility, as smaller firms tend to have

more volatile returns. Market capitalization is calculated using the number of shares outstanding and the stock price as of June 30th the last trading day before portfolio formation.

To assess whether consistent profitability measures significantly outperform inconsistent ones, we would perform time-series tests on the return differentials between the respective long-short portfolios. For instance, to compare the returns of the ROE portfolio and the Net Income to Assets portfolio, the annual difference in returns between the two over the 60-year sample period is calculated. The returns difference is then tested to determine whether it is statistically different from zero. This approach allows us to assess whether any observed outperformance of the consistent measures is not only economically meaningful but also statistically significant.

Our objective is to evaluate whether the construction of profitability ratios, specifically whether they are consistent or inconsistent, affects the performance of profitability based factors. To isolate this effect, we do not include other risk factors such as size, value, or momentum, as our focus is not on explaining the full cross sectional variation in returns but rather on assessing the informational quality of different profitability definitions.

3.2 Fama-MacBeth Model

To test the relationship between stock returns and firm-level profitability characteristics, a Fama-MacBeth two-step regression method (Fama & MacBeth, 1973) is used. This method is used to hopefully get better approximations of the factor returns/risk premiums, with a smaller confidence interval.

Step 1 is to estimate the factor loadings. For each stock in the sample, we estimate its exposure (beta) to the different profitability-based factor returns that were generated using the portfolios. A time-series regression is then run:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t} \quad (1)$$

Where:

$R_{i,t}$ is the return of stock i in year t,

$R_{f,t}$ is the risk-free rate in year t

β_i is the estimated sensitivity of stock i to the profitability factor

f_t is the factor return based on the profitability measure

This regression is repeated for each stock and each of the six profitability measures. The result is a set of stock-level betas that represent how sensitive each stock is to the different profitability-based factor returns. To ensure more accurate beta estimation, companies with fewer than five years of available financial data and return observations are excluded.

In the second step, another regression is used to examine whether these estimated betas can explain future stock returns. For each year t, a cross-sectional regression of stock excess returns on the betas estimated in Step 1 is run:

$$R_{i,t} - R_{f,t} = \lambda_{0,t} + \lambda_{1,t} * \beta_i + \nu_{i,t} \quad (2)$$

Where:

$R_{i,t}$ is the return of stock i in year t,

$R_{f,t}$ is the risk-free rate in year t,

$\lambda_{1,t}$ captures the risk premium associated with the profitability factor in year t,

β_i is the estimated sensitivity of stock i to the profitability factor,

$\nu_{i,t}$ is the residual term

This regression is run once per year across all stocks, and the resulting $\lambda_{1,t}$ estimates are then averaged over time to determine whether there is a statistically significant and persistent relationship between a stock's exposure to a profitability measure and its realized return.

4 Data

4.1 Data Summary

The sample includes companies listed on the NYSE, AMEX, and NASDAQ stock exchange from 1964-2023. Data is downloaded from Compustat and CRSP, with a filter applied to include only share codes 10 and 11, which represent ordinary common stocks. From Compustat, we obtain the most recent financial statement information, while CRSP is used to collect data on stock returns and market capitalization. Firms from the financial sector are excluded since the profitability measures are not applicable. Firms in the utilities sector are also excluded. This is done by applying a filter on the SIC-codes, which are used to identify the type of firm. SIC-codes of 6000-6799 are for financial companies and 4900-4999 for utilities. The period of the return data is July 1st, 1964, to June 30th, 2024. Since the denominator uses opening balances, it means that the sample includes income statement and balance sheet data from 1963 to 2023.

A risk-free rate is used in some parts of the analysis and is then approximated using the 1-month U.S. treasury bill rate. This data is downloaded from Kenneth French's data library (French, 2025).

4.2 Data Cleaning Process

The original financial statement dataset included 438,640 income statement observations and 430,783 balance sheet observations, with each observation representing a firm-year pair. In the first stage of data cleaning, duplicate entries were removed as well as firm-years for which data was available in only one of the two datasets. This reduced the dataset to 406,170 observations.

In the second stage, observations lacking the necessary financial statement components required to calculate the accounting ratios used in the analysis were excluded. Specifically, observations were dropped if any of the following variables were missing: gross profit, EBIT, net income, total assets, or equity. No filters were applied for firm size or liquidity, meaning both small- and large-cap firms are included in the sample. However, firms with negative or zero equity/assets were excluded to ensure ratio definitions remain meaningful. After these steps, 298,261 firm-year observations remained.

The initial stock market returns dataset consisted of 2,732,826 monthly return observations. Observations with missing, zero, or negative market capitalization were removed, resulting in 2,153,058 observations. To compute annual returns, firm-years with fewer than 12 months of return data were removed, resulting in 2,007,564 observations.

The monthly data was then aggregated to calculate annual returns, yielding 167,297 firm-year observations.

Finally, the financial statement and returns datasets were merged, retaining only those firm-years with data available in both sources. The resulting merged dataset contained 127,233 observations. For the year 1963, there was only one observation left. Since it does not make sense to do this strategy one year for only one firm, this observation was removed, resulting in a final dataset of 127,232 observations.

While the dataset is comprehensive, several limitations apply. First, despite extensive data cleaning, some unwanted observations may remain, for example, financial firms not properly excluded due to misclassification. Second, extreme values in financial ratios, particularly those with small denominators, could influence firm rankings. However, robustness checks suggest these do not materially affect the results. Third, the limited availability of certain accounting variables, such as interest income, restricted the scope of profitability measures tested. In particular, the initial plan to use EBIE as the numerator in ROA had to be revised, and EBIT was used instead due to broader data availability.

5 Results

5.1 Descriptive Statistics

The dataset comprises a total of 127,232 firm-year observations. Table 2 displays how these are distributed across six distinct time periods, each spanning ten years. The largest number of observations is found in the period 1994–2003 (33,789), while the period 1964–1973 has the fewest (9,196). These differences in the number of observations do not imply differences in portfolio weighing across years, as the same net investment (zero) is applied in each year.

Table 2. Number of observations and unique firms per time-period.

Period	Total obs.	Unique firms
1964–1973	9,196	1,659
1974–1983	12,709	2,092
1984–1993	22,164	4,484
1994–2003	33,789	6,540
2004–2013	25,467	4,231
2014–2023	23,907	4,218

Table 3 presents summary statistics for the full sample, offering an overview of the distribution and typical values of annual returns and the various profitability ratios. Notably, the standard deviation of the ratios is relatively high, particularly for those that use equity as the denominator. This is largely due to the presence of firms with very low equity values, which can cause these ratios to take on extreme values, either very high or very low. This also helps explain the large differences between the mean and the median. In some cases, such as in the ROE portfolio, the extreme values pull the mean substantially below the median.

It is worth noting that in our investment strategy, we go long in the 30% of stocks with the highest ratio values and short in the 30% with the lowest, with portfolio weights based on market capitalization. As a result, it does not matter whether an observation has an extremely high ROE or a moderately high ROE. As long as it is within the top 30%, it will be included in the long portfolio and weighted according to its market capitalization. To test the potential impact of extreme values, we also reran the investment strategy with these values removed (see Appendix A), and there were no significant changes.

Table 3. Summary statistics for the entire dataset. Values in %.

Variable	Mean	25th pct	Median	75th Pct	Std.Dev.
Annual Return	14.2	-23.4	4.1	35.1	74.8
ROE	-149.4	-3.6	10.1	19.3	36047.1
ROA	-5.9	1.3	8.7	15.6	3185.8
Ebit to Equity	-12.8	-1.6	4.6	9.5	3368.2
Net Income to Assets	-99.2	2.7	19.2	34.3	33925.9
Novy-Marx ratio	35.2	21.9	38.2	59.5	3115.0
Fama-French ratio	-84.8	7.4	24.3	40.3	3308.2

Table 4 presents the annual cut-off values for assigning firms to the top and bottom 30% portfolios based on various profitability measures, expressed in percentages. In each year, firms were sorted by a given profitability measure, and the top and bottom 30% were selected for long and short portfolios. The cut-off values shown in the table represent the profitability thresholds that separated the top 30% from the rest, and the bottom 30% from the rest. Since these thresholds vary from year to year, the table reports their distribution across time using the mean, 25th percentile, median, and 75th percentile. For example, the average annual ROE cut-off for inclusion in the top 30% portfolio was 17.0%, meaning that, on average, firms needed an ROE of at least 17.0% in a given year to qualify. Conversely, the average cut-off for the bottom 30% portfolio was 0.9%, indicating that, on average, firms with an ROE at or below this level were assigned to the bottom portfolio. The graph in Appendix A displays the cut-off values, as well as the median, for each year and ratio.

Table 4. Cut-off values for portfolios. Values in %.

Measure	Top/Bot	Mean	25th pct	Median	75th Pct
ROE	Top	17.0	15.2	17.1	19.0
	Bottom	0.9	-4.6	1.6	8.4
ROA	Top	14.4	11.9	15.0	17.1
	Bottom	3.9	1.5	3.8	7.7
EBIT / Equity	Top	30.9	26.5	31.6	34.6
	Bottom	8.4	3.0	7.8	17.5
Net Income / Assets	Top	8.4	7.3	8.7	9.4
	Bottom	0.4	-2.1	0.7	4.0
Novy-Marx Ratio	Top	54.0	49.1	54.4	60.6
	Bottom	26.8	23.7	27.8	30.4
Fama-French Ratio	Top	36.5	33.1	37.1	40.4
	Bottom	12.8	8.5	12.7	20.3

Table 5 reports summary statistics of stock market performance for firms within the three portfolios formed based on each profitability ratio. Portfolio 1 includes the top 30% of firms by profitability, Portfolio 2 the middle 40%, and Portfolio 3 the bottom 30%. For each portfolio, we present the average, median, and standard deviation of annual returns, as well as the same statistics for the underlying profitability measure. For example, the

table tells us that the average return of all firms sorted into the top portfolio based on ROE as the accounting ratio is 12.4%, while the average return of those sorted into the bottom portfolio is 15.2%. A consistent pattern across all ratios is that the standard deviation of annual returns is higher in the bottom portfolio compared to the top. Unlike the investment strategy, these results are not weighted by market capitalization. It is simply a summary of all observations in the dataset.

Table 5. Annual stock market returns for portfolios sorted by the six different accounting ratios. Values in %.

Sorted by	Portfolio	μ	Median	σ
ROE	Top	12.4	5.9	56.6
	Middle	14.9	6.6	60.9
	Bottom	15.2	-3.1	102.6
ROA	Top	12.4	6.2	55.4
	Middle	15.3	6.7	65.1
	Bottom	14.7	-3.7	99.7
Ebit to Equity	Top	13.2	6.4	56.4
	Middle	14.8	6.5	64.9
	Bottom	14.6	-3.6	99.4
Net Income to Assets	Top	12.2	5.6	57.2
	Middle	15.0	6.8	60.4
	Bottom	15.2	-3.3	102.6
Novy-Marx ratio	Top	14.3	4.4	74.3
	Middle	14.5	5.7	65.9
	Bottom	13.8	1.5	85.8
Fama-French ratio	Top	13.2	6.6	56.5
	Middle	15.2	6.7	65.4
	Bottom	14.0	-3.9	98.9

Table 6 is a correlation matrix of the 6 profitability measures. We observe that the correlations between the profitability ratios are consistently high, never falling below 97.8%. This is expected, as all six ratios are constructed from overlapping financial components. For example, all include revenue and cost of goods sold in the numerator. However, despite these high correlations, Table 7 reveals that portfolio assignments based on these ratios often differ substantially. For instance, while ROE and the Novy-Marx ratio are highly correlated (97.8%), only 51.0% of firms are placed in the same portfolio when sorted by these two ratios. This implies that the choice of profitability ratio matters in terms of portfolio construction and the resulting returns. As firms are either “in” or “out” of a portfolio based on thresholds, even small differences in ratio definition can cause large shifts in portfolio composition.

Table 6. Correlation matrix. Values in %.

	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
ROE	100.0					
ROA	97.9	100.0				
EBIT / Eq.	99.6	98.5	100.0			
NI / As.	98.1	99.9	98.4	100.0		
N-M	97.8	99.9	98.5	99.7	100.0	
F-F	99.2	98.1	99.9	98.0	98.1	100.0

Table 7. Portfolio consistency matrix. Values in %.

	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
ROE	100.0					
ROA	75.2	100.0				
EBIT / Eq.	79.0	79.5	100.0			
NI / As.	83.4	81.2	68.0	100.0		
N-M	51.0	54.5	51.5	52.9	100.0	
F-F	77.0	76.0	67.4	66.2	51.5	100.0

5.2 Portfolio Performance

Section 5.2 shows the results of the investment strategy. Six different long-short portfolios are formed, one based on each profitability measure. Every year firms are ranked according to the six profitability measures. The investment strategy is a long-short portfolio that goes long the 30% most profitable companies and short the 30% least profitable. These positions are held for one year at which point the firms are ranked again and new positions taken.

Table 8 shows the portfolio performance for the six profitability measures. The reported returns reflect the risk premium or factor return of the portfolio constructed using each profitability measure. For example, the first column, ROE, shows the average return, standard deviation, and 95% confidence interval of the portfolio formed by sorting firms based on ROE. The ROE itself is only used to form the portfolios; the reported returns are those of the resulting portfolios.

While none of the measures produce returns that are statistically significant at the 95% confidence level, some patterns can be observed. The consistent measures, ROE and ROA, generate small negative returns and perform the worst. EBIT to Equity has the highest return among the consistent and inconsistent measures. Net Income to Assets generates a small positive return. The two best-performing measures are the benchmark ratios from Novy-Marx and Fama-French, each generating returns above 3%, though still not statistically significant.

Table 8. Average annual percent returns and standard deviation for the portfolios of the 6 profitability measurements, using a 30/40/30 portfolio configuration. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

	ROE	**ROA**	**EBIT / Eq.**	**NI / As.**	**N-M**	**F-F**
Avg. return	-0.1	-0.2	0.9	0.4	3.0	3.1
St. Dev	16.0	17.1	18.0	15.3	14.7	16.2
Lower 95%	-4.1	-4.5	-3.6	-3.5	-0.7	-1.0
Upper 95%	4.0	4.2	5.5	4.2	6.7	7.2

To validate the results, a comparison can be made with the portfolio returns reported by French in his data library (French, 2025). Naturally, the only direct comparison available is with the Fama-French ratio. Table 9 presents the comparison between this study's results and those reported by Fama and French. In terms of average annual return, the figures appear comparable, 3.2% compared to 3.1%. However, the standard deviation reported by Fama and French is significantly lower. This lower volatility leads to statistically significant results at the 95% confidence level, while the results in this study are not statistically significant.

The discrepancy in standard deviation is considerable and may stem from several factors. There are some differences in the timing of portfolio formation and the financial data used, as this study does not fully replicate the Fama-French methodology. Additionally, Fama-French report monthly returns, while this study uses only annual returns. Differences in data cleaning procedures may also play a role, making the sample of firms not completely identical.

Table 9. Comparison of average annual percent portfolio returns and standard deviation between the results reported in the Kenneth French data library, and the results in this study. The lower and upper limit given a 95% confidence interval is also reported. Values in %.

Measure	Fama-French Results	Ljungberg-Ståhl Results
Average	3.2	3.1
St. Dev.	8.9	16.2
Lower 95%	1.0	-1.0
Upper 95%	5.5	7.2

Table 10 shows the performance of the top and bottom portfolios individually. The returns of the top portfolios are relatively similar across all measures, including the benchmarks. However, the returns of the bottom portfolio differ more widely. This suggests that much of the variance in the long-short portfolio performance stems from the bottom part.

Table 10. Average annual percent returns for the top and bottom portfolios individually, as well as the combined long-short return.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Top portfolios return	12.3	12.5	12.5	12.2	13.6	12.6
Bottom portfolios return	12.4	12.6	11.6	11.8	10.5	9.5
Long–short return	-0.1	-0.2	0.9	0.4	3.0	3.1

To investigate the difference between the top portfolios and bottom portfolios we look at them separately. Tables 11 and 12 show the results of a strategy that goes long the top/bottom portfolios respectively, and short the risk-free rate.

Table 11 reports the results of self-financing portfolios that go long the top 30% most profitable firms and short the risk-free rate. Here, all six measures produce similar average returns (7.7% to 8.1%) and standard deviations (17.9% to 18.2%), with the Novy-Marx ratio being a slight outlier with higher return (9.1%) and volatility (19.7%).

Table 11. Average annual percent return and standard deviation for a self-financing portfolio that goes long the 30% most profitable companies and short the risk-free rate. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	7.8	7.9	8.0	7.7	9.1	8.1
St. Dev.	18.0	18.1	18.0	18.2	19.7	17.9
Lower 95%	3.2	3.4	3.5	3.1	4.1	3.6
Upper 95%	12.3	12.5	12.6	12.3	14.0	12.7

Table 12 reports the reverse: portfolios that go long the 30% least profitable firms and short the risk-free rate. Results vary more here. The consistent measures yield higher average returns on the bottom portfolios, which drags down their performance in the standard long-short strategy. The Fama-French and Novy-Marx measures have the lowest average returns in this configuration, especially Fama-French at only 5.0%.

Table 12. Average annual percent return and standard deviation for a self-financing portfolio that goes long the 30% least profitable companies and short the risk-free rate. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	7.9	8.1	7.1	7.3	6.0	5.0
St. Dev.	25.7	27.1	27.1	25.3	18.4	26.2
Lower 95%	1.4	1.3	0.2	0.9	1.4	-1.6
Upper 95%	14.4	15.0	14.0	13.7	10.7	11.6

From Tables 11 and 12, it appears that most performance differences across profitability measures stem from the bottom portfolios. While the returns of the top portfolios are tightly clustered, returns of the bottom portfolios vary more. Still, there are no statistically significant differences between consistent and inconsistent ratios.

This study primarily uses a 30/40/30 portfolio configuration, going long the 30% most profitable companies and short the 30% least profitable, but alternative configurations are tested as a robustness check. Table 13 reports the average annual return and standard deviation across all six portfolios under these different configurations. The 20/60/20 setup is the most aggressive, investing only in the top 20% most profitable companies and short the bottom 20%. As the configuration shifts from 20/60/20 to 30/40/30, then to 40/20/40, and finally to 50/50, the number of companies in both the top and bottom portfolio increases, making the portfolios progressively less aggressive. A clear pattern emerges: standard deviations decrease as more companies are included, likely due to greater diversification. Similarly, the average return across the six profitability measures declines as the portfolios become less aggressive.

Table 13. Average percent return and average standard deviation across all 6 profitability measures for each portfolio configuration. Values in %.

Configuration	Avg. Return	Standard Deviation
20/60/20	2.1	19.9
30/40/30	1.2	16.2
40/20/40	1.0	12.5
50/50	0.8	9.7

Tables 14 to 17 present detailed results for the different portfolio configurations. For each profitability measure and configuration, average annual returns, standard deviations, and 95% confidence intervals are reported.

Table 14 shows the results for the most aggressive configuration, 20/60/20, where the portfolio takes a long position in the top 20% most profitable companies and a short position in the bottom 20%. Compared to the base-case 30/40/30 configuration in Table 15, significantly higher standard deviations are observed across all measures. This is likely due to underdiversification, as the 20/60/20 configuration includes fewer companies in both the top and bottom portfolio. In terms of returns, the two benchmark measures, ROA, and EBIT to Equity, deliver annual returns above 3 percent. In contrast, the two measures based on net income, ROE and Net Income to Assets, underperform, generating returns close to zero, with ROE even producing negative results.

Table 14. Average annual percent return and standard deviation for the portfolios of the 6 profitability measurements, using a 20/60/20 portfolio configuration. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	-0.5	3.2	3.4	0.3	3.3	3.1
St. Dev.	20.4	21.3	21.2	21.4	15.6	19.8
Lower 95%	-5.7	-2.1	-1.9	-5.1	-0.7	-1.9
Upper 95%	4.6	8.6	8.8	5.7	7.2	8.1

Table 15 shows the base case portfolio configuration of 30/40/30, which is the same as in Table 8. The benchmark measures perform the best, while the inconsistent measures slightly outperform the consistent measures. Both the consistent and inconsistent measures generate slightly negative returns.

Table 15. Average annual percent return and standard deviation for the portfolios of the 6 profitability measurements, using a 30/40/30 portfolio configuration. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	-0.1	-0.2	0.9	0.4	3.0	3.1
St. Dev.	16.0	17.1	18.0	15.3	14.7	16.2
Lower 95%	-4.1	-4.5	-3.6	-3.5	-0.7	-1.0
Upper 95%	4.0	4.2	5.5	4.2	6.7	7.2

Table 16 presents portfolio results for the 40/20/40 configuration. In this setup, the portfolio takes long positions in the top 40 percent of most profitable companies and goes short in the bottom 40 percent. Compared to the base case results in Table 15, this configuration exhibits lower standard deviation. Expanding the portfolio from 30 to 40 percent increases the number of companies included, resulting in higher diversification. The Novy-Marx and Fama-French measures show lower average returns relative to the 30/40/30 setup. In contrast, the consistent and inconsistent profitability measures yield equal or higher returns than in the base case, along with lower standard deviations.

Table 16. Average annual percent return and standard deviation for the portfolios of the 6 profitability measurements, using a 40/20/40 portfolio configuration. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	0.1	0.4	1.5	0.4	1.7	1.8
St. Dev.	12.6	11.9	13.1	12.1	12.7	12.5
Lower 95%	-3.1	-2.6	-1.8	-2.7	-1.5	-1.4
Upper 95%	3.3	3.4	4.8	3.4	4.9	4.9

Table 17 presents the portfolio results for the final and least aggressive configuration. The 50/50 portfolio splits the entire sample of companies in half based on profitability, investing in the more profitable half and shorting the less profitable half. This configuration results in maximum diversification, as no firm is excluded, which is reflected in the low standard deviation. However, it also leads to lower average returns for the two benchmark measures. In particular, the Fama-French measure generates an average annual return of only 1.7 percent in this configuration, compared to the 3.1 percent return observed in the 30/40/30 configuration.

Table 17. Average annual percent return and standard deviation for the portfolios of the 6 profitability measurements, using a 50/50 portfolio configuration. The lower and upper limits of a 95% confidence interval are also reported**.** Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	0.1	0.8	1.7	-0.8	1.5	1.7
St. Dev.	9.4	9.4	10.6	8.7	11.3	9.1
Lower 95%	-2.2	-1.6	-1.0	-3.0	-1.4	-0.6
Upper 95%	2.5	3.2	4.4	1.4	4.4	3.9

Across all four portfolio configurations the two benchmark measures, Novy-Marx ratio and Fama-French ratio, consistently produce the highest average returns. The Fama French ratio in particular performs well across all specifications, with average annual returns ranging from 1.7 percent to 3.1 percent and relatively moderate standard deviations.

Among the consistent measures, ROA generally performs the same or better than ROE, although the return difference is not statistically significant. For example, in the 20/60/20 strategy, ROA delivers an average return of 3.2 percent, compared to negative 0.5 percent for ROE.

The inconsistent measures show more mixed results. EBIT to Equity stands out with strong performance across all configurations. For instance, it yields an average return of 3.4 percent in the 20/60/20 strategy, which is the best out of all the measures. Net Income to Assets, on the other hand, tends to underperform across most specifications and at times produces negative average returns.

When comparing consistent and inconsistent measures more broadly, no clear pattern emerges. However, the inconsistent measure EBIT to Equity is the top performer, generating the highest returns in all four configurations. Standard deviations for the four measures remain relatively similar to each other, regardless of the portfolio setup.

5.3 Fama-MacBeth Results

In this study, the Fama MacBeth method is used as a complementary approach to the portfolio sorting analysis for estimating the returns to the profitability factor. Table 18 reports the results of the Fama-MacBeth two-step regression. In the first step, stocks with fewer than five annual observations are excluded, reducing the sample to 116,638 observations.

Table 18. Output from the Fama-MacBeth 2-step regression.

Measure	Coef.	Std. Err.	t	P > (t)	Lower	Upper
ROE	-2.4%	2.1%	-1.14	0.257	-6.7%	1.8%
ROA	-3.5%	2.4%	-1.44	0.156	-8.4%	1.4%
EBIT / Eq.	-2.7%	2.7%	-1.01	0.318	-8.0%	2.6%
NI / As.	-3.2%	2.0%	-1.60	0.116	-7.3%	0.8%
N-M	0.7%	2.1%	0.32	0.752	-3.5%	4.8%
F-F	-3.0%	2.4%	-1.24	0.219	-7.8%	1.8%

Interestingly, the Fama-MacBeth two-step regression yielded no significant coefficients for any of the six profitability measures and, in several cases, produced signs that contradicted the portfolio-based results. One possible explanation is that the Fama-Macbeth method relies heavily on accurate beta estimation, which is challenging with annual return data and a relatively short time-series per firm. This suggests that, in our setting, portfolio sorting may be more effective at capturing the relationship between profitability and returns than regression-based approaches.

6 Discussion

The aim of this study is to investigate the importance of using consistent definitions of profitability in factor models, with a hypothesis that consistent measures should produce better results than inconsistent ones. Although the results do not show a significant difference in performance, important insights can still be drawn.

As shown in Table 6, the profitability measures exhibit a high degree of correlation. A firm with a high value on one measure, such as Net Income to Assets, is likely to score highly across the other measures as well, including those classified as consistent. This suggests that the use of inconsistent measures, such as Net Income to Assets, as control variables in regression models may not pose a significant issue. While such measures may be theoretically imperfect, the strong correlations observed in this study indicate that they perform comparably to more consistently defined profitability measures.

Although our findings may initially appear to contrast with Chullen (2013) and Johansson & Runsten (2023), the contexts are fundamentally different. Their work emphasizes the importance of consistency in accounting ratios for valuation and fundamental analysis, whereas our study focuses on the role of consistency within factor portfolio construction. Rather than contradicting their conclusions, our results suggest that consistency in ratio definitions may matter in some domains but appear to have limited relevance for profitability-based factor models.

Importantly, the high correlation observed between profitability measures in our analysis does not imply that these measures can be used interchangeably in all settings. For instance, substituting EBIT to Equity for ROE in a Residual Income Valuation (RIV) model would yield conceptually flawed results (Penman, 2013), despite the statistical similarity between the two measures.

Among all portfolio configurations, two profitability measures consistently outperformed the others: the Novy-Marx gross profitability measure and the Fama-French operating profitability measure. Of the two, the Fama-French ratio generally exhibited a slight performance advantage. Several factors may account for this outperformance. First, in developing the five-factor model, Fama and French likely evaluated numerous profitability definitions and ultimately selected the one yielding the most robust empirical results. Given that this study employs a similar sample of firms, over a comparable time period, and with a closely aligned methodology, it is unsurprising that the Fama-French measure performs particularly well in this context.

Second, both the Novy-Marx and Fama-French measures differ from the other four in that they are more closely tied to cash flows. The Fama-French measure serves as an equity-based cash flow proxy, excluding non-cash items such as depreciation and amortization. The superior performance of these measures aligns with the findings of Ball et al. (2016)

and Fama and French (2018), who report that profitability metrics more closely linked to cash flows tend to deliver stronger predictive power in asset pricing contexts.

One particularly noteworthy observation relates to where the return differences between profitability measures arise. Tables 10 to 12 show that the bottom portfolio, containing the least profitable companies, contributed the most to the variation in performance across measures. This is consistent with previous research showing that firms with lower profitability tend to exhibit higher market return volatility on (see, for example, Barinov, 2023 and French, 2025). When isolating the top portfolio (Table 11), the Novy-Marx measure slightly outperformed, but the remaining five measures produced nearly identical average returns. This suggests that the choice of profitability definition has a greater impact on which firms are excluded or shorted, rather than on the selection of high-performing firms.

From the summary statistics, it can be observed that a lot of extreme values for the ratios are included in the sample. For instance, one firm in a given year reported a Return on Equity (ROE) of negative 12 million percent. These extreme values might distort the results and can potentially explain the non-significant results. However, it is important to emphasize that the ratios are used solely for portfolio sorting, while the reported outcomes are based on annual returns. A firm with an ROE of negative 10 percent or negative 100 percent will be placed in the same portfolio, meaning such extreme values do not directly influence the return calculations. As a robustness check, additional results were generated using a sample in which observations with the most extreme values were excluded. These results can be found in Appendix B. The exclusion of extreme values did not result in any significant change in results, which was expected for the reasons discussed above.

Following the lack of significant results from the portfolio analysis, a Fama-MacBeth two-step regression was used to see if a better approximation of the factor return, with a lower confidence interval, could be found. However, this regression also produced no statistically significant coefficient for any of the profitability measures. This may be due to limitations in estimating factor exposures using annual data, especially because of the short return history for many of the firms.

7 Conclusion

This study sets out to examine whether consistent definitions of profitability yield better performance in factor portfolio construction compared to inconsistent ones. Using 60 years of stock market data and six profitability measures, including benchmark measures from Fama-French and Novy-Marx, we found no significant evidence that consistent measures outperform inconsistent ones in terms of return or risk-adjusted performance. While theoretical arguments favor consistency, the high correlation among all profitability measures suggests that in practical applications, such as portfolio sorting, the choice between consistent and inconsistent definitions may have limited impact. Notably, the two benchmark measures consistently performed best across specifications.

7.1 Limitations

The analysis in this study is based entirely on annual returns, rather than quarterly or monthly returns, which limits the number of observations. While this is unlikely to affect the portfolio sorting results, it may reduce the statistical power of the Fama MacBeth regressions. Using monthly return data could potentially have produced more precise estimates of the factor risk premia by increasing the number of time periods and improving the reliability of the regression outputs.

Another potential limitation is the risk of errors in the financial data downloaded from Compustat. We excluded financial and utility firms using SIC-code filters; however, some firms lacked a SIC-code entirely, which means that a small number of these firms may have remained in the final sample despite our exclusions. Additionally, while we cross-checked parts of the dataset and did not observe any obvious inaccuracies, it is still possible that some financial statement items contain reporting errors or inconsistencies.

Ideally, we would have been able to use EBIE to Assets as the definition of ROA, as this definition ensures consistency between the numerator and denominator. However, due to data limitations, it was not possible to obtain reliable EBIE figures for many firms in the sample. We thus settled on using EBIT to Assets as the definition.

7.2 Future Research

Future research could explore whether using EBIE instead of EBIT in the ROA calculation yields different results, provided that the necessary data can be collected.

Furthermore, it would be interesting to investigate in future research whether the Fama-French ratio continues to outperform other ratios. We suspect that Fama and French selected this specific ratio for their model because it produced the strongest results over the time period they examined. By testing different profitability measures over a

subsequent period, one can assess whether there is something inherently robust about their chosen definition, or if it simply happened to perform best within their sample during that particular time period.

8 References

Ang, A. (2014). *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press: New York. Digital version accessed through SSE Library. https://hhs.primo.exlibrisgroup.com/discovery/fulldisplay?context=L&vid=46SSOE_INST:46SSOE_view&docid=alma991001404178906056

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. (2016). "Accruals, Cash Flows, and Operating Profitability in the Cross Section of Stock Returns". *Journal of Financial Economics, 121(1), 28-45*. https://doi.org/10.1016/j.jfineco.2016.03.002

Barber, B. M. & Lyon, J. D (1996). "Detecting Abnormal Operating Performance: The Empirical Power and Specification of Test Statistics". *Journal of Financial Economics, 41(3), 359-399*. https://doi.org/10.1016/0304-405X(96)84701-5

Barinov, A. (2023). "Profitability Anomaly and Aggregate Volatility Risk". *Jornal of Financial Markets, 64*. https://doi.org/10.1016/j.finmar.2022.100782

Berk, J. B. & DeMarzo, P. M. (2024). *Corporate Finance*. Sixth edition. Pearson Education Limited: Martin.

Blitz, D. (2020). "Factor Performance 2010-2019: A Lost Decade?". *Portfolio Management Research, 11(2), 1-9*. https://www.robeco.com/files/docm/docu-robeco-factor-performance-2010-2019-a-lost-decade.pdf

Chen, N. & Zhang, F. (1998). "Risk and Return of Value Stocks". *The Journal of Business, 71(4), 501-535*. https://doi.org/10.1086/209755

Cheng, Q., Lee, J., & Shevlin, T. (2016). Internal governance and real earnings management. *The Accounting Review, 91(4), 1051-1085*. http://dx.doi.org/10.2308/accr-51275

Cheng, S. (2008). "Board Size and The Variability of Corporate Performance". *Journal of Financial Economics, 87(1), 157-176*. https://doi.org/10.1016/j.jfineco.2006.10.006

Chullen, A. (2013). "Does Consistency Improve Multiple Quality?". *Journal of Business Economics*, *85(6)*. http://dx.doi.org/10.1007/s11573-015-0768-2

Erhardt, N. L., Shrader C. B., & Werbel, J. D. (2003). "Board of Director Diversity and Firm Financial Performance". *Corporate Governance: An International Review, 11(2), 102-111*. https://doi.org/10.1111/1467-8683.00011

Fama, E (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25(2), 383-417*. https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (1993). "Common Risk Factors in the Returns on Stocks and Bonds". *Journal of Financial Economics, 33(1), 3-56*. https://doi.org/10.1016/0304-405X(93)90023-5

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., & MacBeth, J. D. (1973). "Risk, Return, and Equilibrium: Empirical Tests". *Journal of Political Economy, 81(3), 607-636*. https://www.jstor.org/stable/1831028

French, K. (2025). "Current Research Returns". *Kenneth R. French*. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html

Goedhart, M. H., Koller, T., & Wessels, D. (2020). *Valuation: Measuring and Managing the Value of Companies*. Wiley: Hoboken. Digital version accessed through SSE Library. https://hhs.primo.exlibrisgroup.com/discovery/fulldisplay?context=L&vid=46SSOE_INST:46SSOE_view&docid=alma991001420279706056

Johansson, S. & Runsten, M. (2023). *The Profitability, Financing, and Growth of the Firm*. Studentlitteratur AB: Lund.

Novy-Marx, R. (2013). "The Other Side of Value: The Gross Profitability Premium". *Journal of Financial Economics, 108(1), 1-28*. https://doi.org/10.1016/j.jfineco.2013.01.003

Penman, S. H. (2013). *Financial Statement Analysis and Security Valuation*. Fifth edition. McGraw-Hill Company: Singapore.

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, *38, 1-41*. https://doi.org/10.2307/2672906

Schroeder, R. G., Cathey, J. M., & Clark, M. (2023). *Financial Accounting Theory and Analysis*. Fourteenth edition. John Wiley & Sons: Hoboken.

Shleifer, A. (2000). *Inefficient Markets*. Oxford University Press: Oxford.

Whittington, G. (1980). "Some Basic Properties of Accounting Ratios". *Journal of Business Finance and Accounting, 7(2), 219-232*. https://doi.org/10.1111/j.1468-5957.1980.tb00738.x

Whittington, G. (2017). *Value and Profit: An Introduction to Measurement in Financial Reporting*. Cambridge University Press: Cambridge.

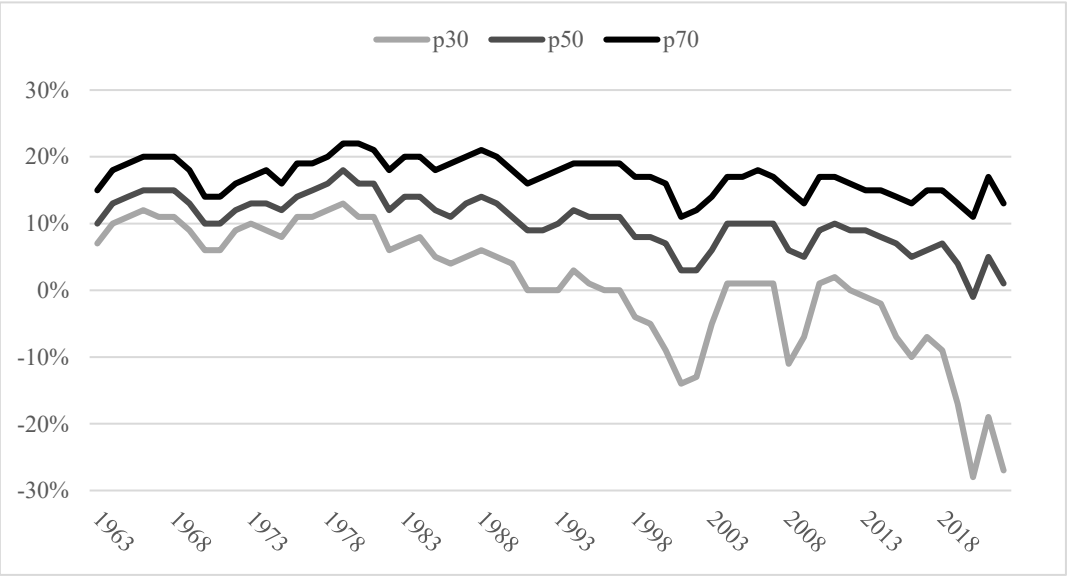
WRDS (n.d.). "Compustat – Capital IQ". *Wharton University of Pennsylvania*. https://wrds-www.wharton.upenn.edu/login/?next=/pages/get-data/compustat-capital-iq-standard-poors/

9 Appendix

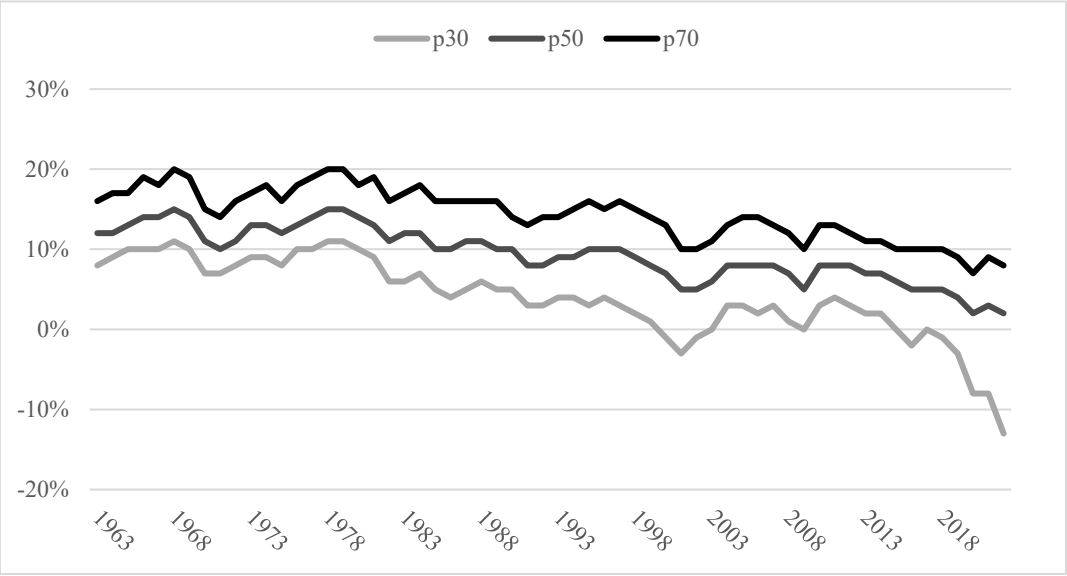
Appendix A: Additional Graphs

Graphs 2-7 show the 30th, 50th, and 70th percentiles for each accounting ratio and year. The 30th and 70th percentiles are shown since those were the cut-off values for the portfolios. The firms below the 30th percentile were placed in the bottom part of the portfolio, while the firms above the 70th percentile were placed in the top part of the portfolio.

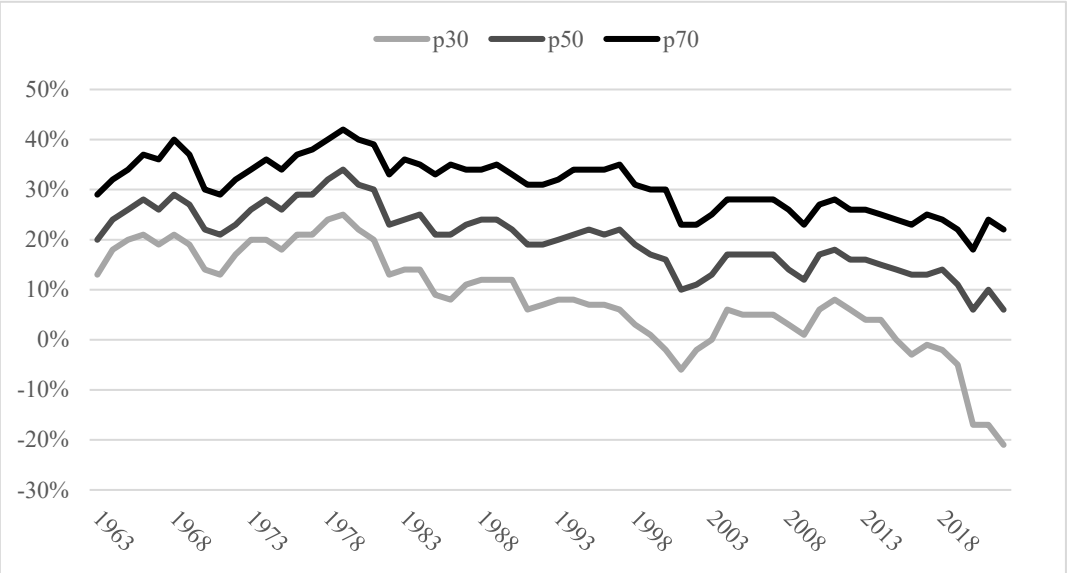
Graph 2. Values for ROE.



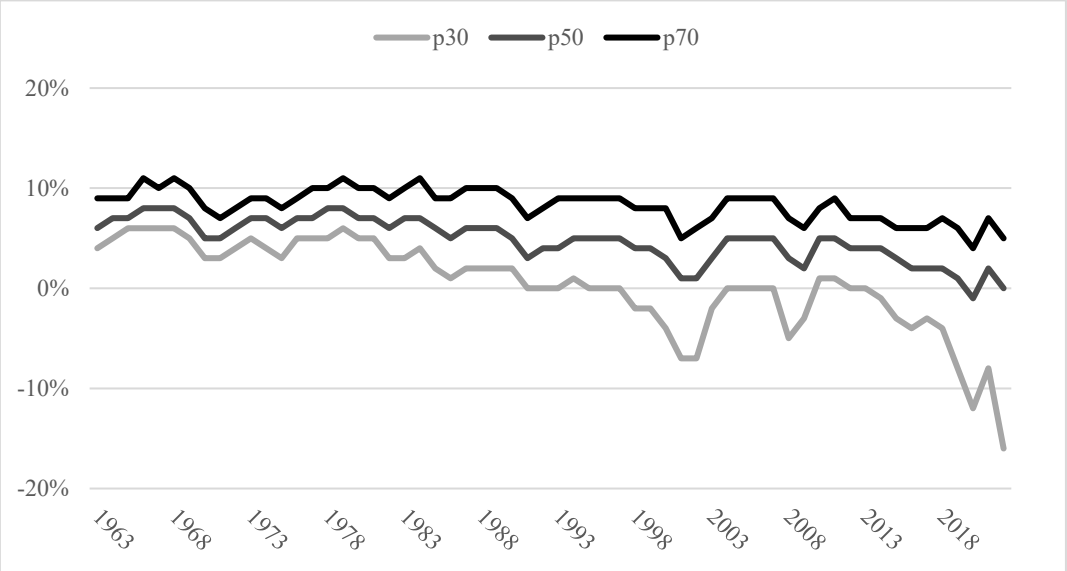
Graph 3. Values for ROA



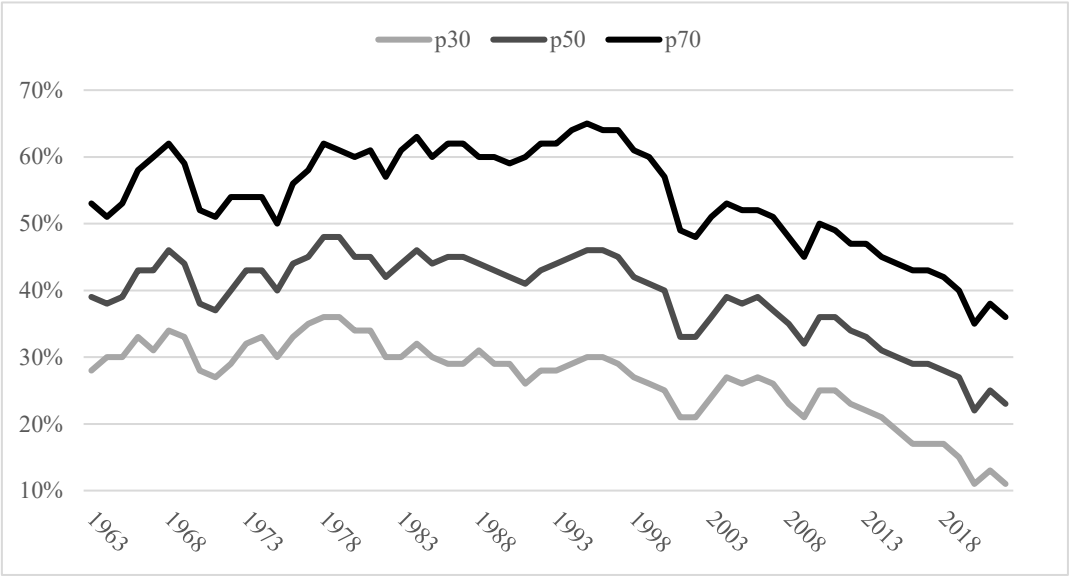
Graph 4. Values for EBIT to Equity.



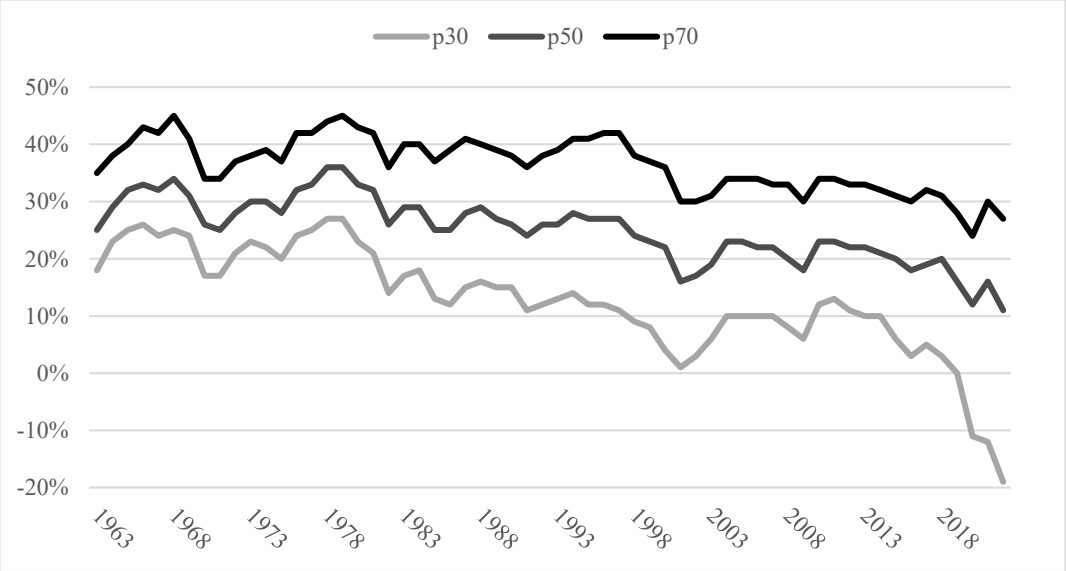
Graph 5. Values for Net Income to Assets.



Graph 6. Values for Novy-Marx Ratio.



Graph 7. Values for Fama-French Ratio.



Appendix B: Effect of Excluding Outliers

From the summary statistics section it is clear that the sample includes a lot of companies with extreme values for the profitability measurements. Table 19 and 20 below show the new results after removing some of these outliers.

Table 19. Data cleaned by removing the top 1% and bottom 1% of all profitability observations. Average annual percent returns and standard deviation for the portfolios of the 6 profitability measurements, using a 30/40/30 portfolio configuration. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	-0.6	-0.6	0.5	-0.5	2.8	3.0
St. Dev.	15.7	16.9	16.9	14.7	14.3	15.8
Lower 95%	-4.6	-4.9	-3.7	-4.2	-0.8	-1.0
Upper 95%	3.4	3.7	4.8	3.2	6.4	7.0

Table 19 shows the results after removing the top 1% and bottom 1% observations for each profitability measure. This sorting is only done for one profitability measure at a time. The first column shows the returns for the roe portfolio after removing the companies with the 1% highest and 1% lowest roe in the sample. The process is then repeated for each measure, meaning that the sample will not be the same for all measures anymore.

By comparing the results in Table 19 with those in base-case Table 8 the effects of this data cleaning are visible. The average return for every measure is lower, with only a small decrease in standard deviation.

Table 20. Data cleaned by removing the top 5% and bottom 5% of all profitability observations. Average annual percent returns and standard deviation for the portfolios of the 6 profitability measurements, using a 30/40/30 portfolio configuration. The lower and upper limits of a 95% confidence interval are also reported. Values in %.

Measure	ROE	ROA	EBIT / Eq.	NI / As.	N-M	F-F
Average	-0.9	-0.9	0.2	-1.8	2.4	2.1
St. Dev.	15.0	15.6	15.4	14.7	14.1	14.1
Lower 95%	-4.7	-4.8	-3.7	-5.6	-1.2	-1.5
Upper 95%	2.9	3.1	4.1	1.9	5.9	5.6

Table 20 shows the results after doing the same sorting as in Table 19 but now removing the top 5% and bottom 5% observations for each profitability measure. Comparing these results with those in Table 19 and Table 8 all the average returns are lower with only a small decrease in standard deviation.

Appendix C: Usage of AI Tools in Thesis Writing

ChatGPT by OpenAI was used in the creation of this thesis for two main purposes: assistance with coding and support in text and language editing. No other AI tools were used.

For coding, ChatGPT helped us generate STATA code, as well as identifying errors in our coding. When we were uncertain about which commands to use, we described the desired outcome to the tool and received code suggestions. When errors occurred, we used ChatGPT to diagnose the issue and propose solutions. This significantly reduced the time we spent on coding. However, we were aware of the risks of using the tool in this way. First, the tool could misinterpret our prompts and provide code that does not work for our purposes. Second, it could provide incorrect code. Running such code and using the output for our thesis would compromise its accuracy. We minimized this risk by checking ChatGPT's suggestions against STATA manuals and carefully validated our data. For example, when merging the balance sheet, income statement, and annual returns datasets, we verified several observations to ensure accuracy across years. Additionally, the data used in our descriptive statistics was cross-verified with our Excel files.

In terms of language editing, ChatGPT was also a big help. For parts where we thought the text could be improved, we requested suggestions from the tool. This helped enhance phrasing, grammar, and overall clarity. At times, the tool's edits removed important points. We thus always compared the original and revised versions carefully before finalizing any text. Furthermore, the tables in the thesis were partially created with the help of ChatGPT. Instead of manually entering large amounts of numerical data, we provided screenshots of tables generated in STATA and asked the tool to format them in the way we wanted. This saved us time and minimized the risk of human error. All tables were cross-checked against the original STATA output to ensure accuracy.