Stockholm School of Economics
Department of Accounting & Financial Management
Bachelor Thesis in Accounting & Financial Management
May 2023

Can a modern value definition save a struggling investment strategy?

A study on the performance of the F-Score when adjusting book-to-market equity for intangibles

Hampus Johnson
24999@student.hhs.se

Max Rembe
24965@student.hhs.se

Abstract

Studies have found that the immediate expensing of intangible investments has understated the book-to-market metric, which has caused a substantial misclassification of value and glamor stocks in the new economy. This has coincided with a deteriorating performance of Piotroski's F-score, which aims to identify winning stocks among the value group by considering financial signals relevant to these stocks. We evaluate whether the artificial capitalization of intangible investments to the book-to-market metric would improve the performance of Piotroski's F-score through a less biased capture of value stocks, and if the resulting strategy manages to beat the S&P 500 index in the U.S. from 2000 to 2022. Our results show that adjusting the value screen can significantly increase the returns of a portfolio that invests in expected winners and that such a portfolio can outperform the market index. We also find evidence that the F-score is better able to separate winning from losing stocks after adjustments and that investing in identified winners is superior to investing in the underlying, complete book-to-market group.

Tutor: Ting Dong

Keywords: Value investing, Piotroski's F-Score, Intangible assets, Book-to-market equity

Acknowledgements: Special thanks to Assistant Professor Ting Dong for providing much-appreciated guidance and support. We also wish to thank friends and family for final remarks.

Table of contents

1. Introduction

The core principles of value investing are best summarized by Benjamin Graham's writing in *The Intelligent Investor,* which was first published in 1949. In this book, he defined value investing as an investment strategy that focuses on buying stocks that are trading at a discount to their intrinsic value, which may be identified by the firm's underlying fundamentals. In time, the stock price will revert to its intrinsic value as it is recognized by the market, driving profits for holding value stocks and losses for holding overvalued glamor stocks (Graham, 2003). The success of such strategies challenges the semi-strong form of the efficient market hypothesis, which posits that all publicly available information should already be reflected in stock prices (Fama, 1970).

One of the more common measures value investors have used to gauge the relationship between the intrinsic and fundamental value of a stock is the book-to-market equity ratio (B/M). Stocks with high B/M ratios are commonly referred to as value stocks while those with low B/M ratios are called glamor stocks. Value stocks have a lower market value relative to fundamentals, while glamor stocks have a higher market value relative to fundamentals. As such, the tendency for mean reversion has historically driven the gains of investing in value and shorting glamor stocks. Researchers have disagreed on the cause of these returns, attributing them either to investor mispricing or higher financial distress risk among value firms (Fama and French, 1993; Lakonishok et al., 1994).

Piotroski (2000) later supported the notion that value firms as a group are financially distressed. Specifically, he noted that the returns of investing in high B/M firms are driven by the strong performance of a minority of firms while tolerating the poor performance of the majority. In an attempt to separate firms with strong and weak outlooks, he designed an accounting-based metric with nine binary signals specifically aimed to assess the financial health of firms in this group. Piotroski demonstrated that the usage of this metric, which he called the F-score, would yield 33.5% annualized returns between 1976 and 1996 in the U.S. by investing in expected winners (high F-scores). He further noted that these returns were 23.5% higher than those for expected losers (low F-scores) in the value group over the same period (Piotroski, 2000). Covering a similar period, Mohanram (2005) confirmed that Piotroski's F-score was able to generate significant positive returns and that the strategy was best applied to value stocks.

However, later studies utilizing Piotroski's F-score have found that returns yielded from the strategy have decreased in periods after 1996 in the U.S. (Piotroski and So, 2012; Woodley et al., 2011). This has coincided with other research, which has found that there has been a decrease in returns from investing in high book-to-market (value) stocks in recent years. One of the most prominent explanations these studies put forward for the effect relates to the accelerating shift from corporate tangible to intangible investments in the economy. Under U.S. GAAP, internally generated intangible assets are immediately expensed, while tangible investments are capitalized and amortized. The market valuation of firms with higher levels of intangible investments will thus appear systematically higher in relation to book figures. As such, researchers argue that the book-to-market ratio in its basic form has become outdated, since it has an inherent bias of classifying stocks with more intangible investments as glamor stocks, even if such stocks would otherwise be classified as value stocks. To remedy the bias, these studies have artificially capitalized and amortized intangible investments to book figures to compute an intangible-adjusted book-to-market equity metric. The researchers subsequently demonstrated that usage of this metric improved the identification of value firms through fewer pricing errors, while also improving the returns of the value group as a whole. (Arnott et al., 2021; Eisfeldt et al., 2022; Lev and Srivastava, 2019).

This study aims to investigate whether Piotroski's F-score would benefit from a more updated definition of value stocks since the methodology was originally designed to address financial characteristics important for this group. To remedy biases from intangibles, value stocks are identified through the usage of a B^{INT}/M metric, which is computed through artificial capitalization of intangible investments to book-to-market equity. We then evaluate whether Piotroski's high F-score strategy generates higher returns when applied to the adjusted value group, and if the ability to separate between winning and losing stocks is improved. Finally, we also investigate whether the high F-score strategy has been able to outperform the S&P 500 index in the 21st century. Univariate tests are used to analyze the returns of F-score portfolios and the market index. As such, we do not aim to draw any conclusions on the risk-adjusted returns of portfolios, but rather their relative return performance. Furthermore, while we propose several explanations for our findings based on collected data and prior literature, it is left to future research to verify these possible explanations.

To the best of our knowledge, no prior study has evaluated the performance of Piotroski's F-score, or any similar value investing strategy, when adjusting the value screen by capitalizing intangibles. While this paper is limited to Piotroski's F-score, the intangible misclassification issue has adversely affected multiple investment methodologies. As such, we argue that the results of this study are of relevance to both private and institutional investors, as well as academics who study fundamental analysis of accounting information in relation to intangibles.

Research questions

This study aims to answer the following research questions:

> **Research question 1:** Is it possible to improve Piotroski's high F-score strategy by adjusting the book-to-market screening for intangibles?
>
> **Research question 2:** Is it possible to improve the F-score's ability to separate between winners and losers by adjusting the book-to-market screening for intangibles?
>
> **Research question 3:** Is it possible to beat the market index with Piotroski's high F-score strategy by adjusting the book-to-market screening for intangibles?

We demonstrate that intangible adjustments to Piotroski's methodology can significantly increase portfolio returns. This is in line with previous research which argues that adjusting for intangibles can increase the returns of the value effect (Arnott et al., 2021; Eisfeldt et al., 2022). The same strategy also beats the S&P 500 index during the 21st century, but the outperformance is mainly driven by comparatively strong performance in the first decade. Our results also show that the F-score's ability to separate between expected winners and losers is improved when adjusting the book-to-market screening for intangibles. Identified winners also perform sufficiently well to outperform a simpler strategy that invests in the complete underlying book-to-market group. For the unadjusted F-score methodology, neither the ability to separate winning and losing stocks, nor beat a high book-to-market strategy, is present. As such, we find evidence for significant improvements to Piotroski's F-score on all tested accounts when the book-to-market value screen is adjusted for intangibles. These findings support the importance of context for Piotroski's F-score demonstrated by prior studies (Mohanram, 2005; Piotroski and So, 2012).

2. Literature review and theory

In this section, we will present and review theories and research related to this paper. First, background is given to the efficient market hypothesis and some anomalies relevant to value investing. Thereafter, we give an overview of value investing and Piotroski's F-score strategy. This will be followed by an outline of literature covering the recent decline of the value effect and value relevance of accounting information, and how both are tied to the immediate expensing of intangible investments. In this subsection, we will also present studies that have attempted to adjust book-to-market equity for intangibles to better capture the value effect. The section concludes with the development of hypotheses tied to the findings of prior studies.

2.1. Underlying market theories

2.1.1. Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) was developed by Fama (1970) from the random walk theory, which suggests that 1) price changes of securities are independent, and 2) price changes are identically distributed. Under these assumptions, stock prices should move unpredictably, so that past stock prices or market trends cannot be used to predict future price movements. It further implies that stock prices should at any time fully reflect available information, and the publishing of new information should immediately be incorporated into stock prices. Based on these arguments, Fama (1970) proposed three different forms of market efficiency with varying restrictions: a weak form, semi-strong form, and strong form. In the weak form, past stock prices and patterns cannot be used to predict future stock prices, meaning that technical trading strategies cannot consistently outperform the market. In the semi-strong form, stock prices also conform to all publicly available information such as earnings announcements and news announcements, rendering fundamental analysis useless in predicting returns. In the strong form, stock prices reflect all available information including private or monopolistic information. As such, no level of analysis can predict future price movements. At the time, the author found strong empirical support for the weak form, mostly supportive evidence for the semi-strong form, and limited support for the strong form (Fama, 1970). Since the publication of his study, the investment community has generally accepted Fama's conclusions, and several supportive studies of notability have been conducted (Burton, 1973; Samuelson, 1973; Jensen, 1978).

In terms of pricing, the Efficient Market Hypothesis further suggests that the expected return of an asset should be proportionate to the risk exposure of investing in that asset. Any return above the expected return is called abnormal (Berk and Demarzo, 2019). A separate body of literature has thus been focused on developing asset pricing models to estimate the expected return. Two of the most influential ones are the CAPM model by Sharpe (1964) and Lintner (1965), and the three-factor model by Fama and French (1993), which builds upon the CAPM by adding an SMB (small-minus-big market capitalization) and HML (high-minus-low book-to-market equity) factor. These additional factors would proxy for unobserved systematic sources of risk, implying that the market is efficient and CAPM is misspecified (Fama and French, 1993).

2.1.2. Anomalies that challenge the theory of efficient markets

Although widely accepted by economists, the Efficient Market Hypothesis (EMH) has been challenged by the observation of anomalies that challenge the random walks of stock prices. While each anomaly by itself may not be enough evidence to reject EMH, when viewed as a whole, they may present a much stronger case to review acceptance of the theory (Jensen, 1978). Below we present some anomalies which are commonly brought up in tandem with value investing strategies. These include the size effect, post-earnings announcement drift, mean reversion, and the book-to-market effect.

Size effect

The size effect was first documented by Banz (1981) and Reinganum (1981), who noted that low market capitalization (small) stocks had historically generated higher risk-adjusted returns than high market capitalization (large) stocks, adjusted to CAPM. Both discussed the possibility that the model may be misspecified and that it may not necessarily be a source of market inefficiency. However, Roll (1981) argued that the lower trading frequency of small stocks resulted in risk measures obtained from short-interval return data to seriously understate the risk of holding a small firm portfolio. The effect was particularly apparent for daily data, but also significant for periods as long as one month. Other researchers have tried to specify the underlying risk captured by the size effect but with inconclusive explanations. For instance, Amihud and Mendelson (1986) argued that the size effect is largely tied to liquidity risk, whereas larger stocks are more easily traded. Zhang (2006) suggested that size proxies for 'information uncertainty', in that small stocks provide poorer information to investors. Whatever the case, research in the 21st century posits that it is no longer possible to earn abnormal returns from the size effect (Patel, 2012).

Post-earnings announcement drift

The post-earnings announcement drift was first documented by Ball and Brown (1968), who found that estimated cumulative "abnormal" returns continue to drift up for "good" news firms and down for "bad news" firms, even after earnings are announced. This contradicts the efficient market hypothesis, which suggests that all available information should already be reflected in stock prices. Bernard and Thomas (1989) later investigated two possible causes for the effect: 1) There is a delayed price response, either due to traders failing to assimilate information or transaction costs in exploiting it. 2) So-called abnormal returns are nothing but fair compensation for risks not captured by CAPM. The authors found evidence that was consistent with delayed price responses and inconsistent with CAPM misspecification. Furthermore, drifts were found to have a longer duration for small firms as compared to large firms. (Bernard and Tomas, 1989).

Mean reversion

Poterba and Summers (1988) were some of the first to provide evidence of mean reversals in stock prices. They demonstrated that there tends to be a positive autocorrelation in returns over the short-term and negative autocorrelation over longer horizons. The authors suggested that noise trading, i.e., stock trading where demand is determined by other factors than expected return, could possibly explain this occurrence. Such traders would push stock prices to divert from their fundamental values in the short term before they ultimately revert (Poterba and Summers, 1988). Mean reversion may challenge even the weak form of the efficient market hypothesis if implemented purely based on prior price information. However, researchers have also commonly connected mean reversion to other stock market phenomena, such as the book-to-market effect.

Book-to-market (value) effect

Book-to-market (B/M) refers to the ratio between the book value of equity to the market value of equity of a firm (Berk and Demarzo, 2019). The book-to-market effect, or value effect, refers to the tendency of firms with high book-to-market equity ratios to outperform the market. It was popularized by Fama and French (1992) who showed that there was a significant return spread between high and low B/M firms, which they attributed to additional risk exposure. Specifically, the authors argued that the average high B/M firm was financially distressed with poor earnings prospects, and that they experienced an involuntary leverage effect as future earnings were discounted at a higher cost of capital (Fama and French, 1992).

However, other researchers have argued that the book-to-market effect could be attributed to behavioral biases of investors, and that high B/M stocks are not fundamentally riskier. Lakonishok et al. (1994) showed that investors make judgment errors by extrapolating past performance in growth predictions, without appreciating the tendency for mean reversion. This would be characterized by excessive optimism for glamor stocks (low B/M) with strong past performance and excessive pessimism for value stocks (high B/M) with poor past performance. Excessive buying of glamor stocks thus leads to their overpricing, while overselling of value stocks leads to their underpricing. Value strategies would then exploit these expectational errors to yield higher returns when prices revert, without being fundamentally riskier. (Lakonishok, et. al, 1994).

2.2. Value investing and Piotroski's F-score

2.2.1. Principles of value investing

Benjamin Graham presents the core principles of value investing in his book *The Intelligent Investor* which was first published in 1949. He defined value investing as an investment strategy that focuses on buying stocks that are trading at a discount to their intrinsic value, which can be identified by assessing a firm's underlying fundamentals. This includes financial statements, earnings history, and other relevant firm data. Graham believed that markets are efficient in the long term, but that short-term discrepancies and prevailing market sentiments may distort stock prices. In time, however, the market will recognize the true value of a stock, which will drive profits for holding value stocks and losses for holding overvalued (glamor) stocks. As such, intelligent investors should keep a level head and ignore irrational signals from the market. By being disciplined, taking a long-term perspective, and only buying stocks at a significant discount to their intrinsic value, investors could thus achieve superior returns while minimizing downside risk. (Graham, 2003).

Later studies have supported the notion that higher returns may be earned by studying accounting fundamentals. Ball and Brown (1968) found that net income figures provided useful information for predicting future earnings, and that stock prices more accurately represented these figures closer to earnings announcements. Similarly, Sloan (1996) provided evidence that investors fail to appreciate the differing persistence cash and accruals components of earnings have in future periods. Specifically, he showed that investors taking a long-short position in firms with high/low relative cash flow components of earnings could earn abnormal returns. Ou and Penman (1989) took a more sophisticated approach and combined signals from multiple accounting figures into a single measure. They demonstrated that taking long-short positions for 2-year holding periods in firms with a strong/weak combined measure would yield returns of 12.5%. The authors concluded that while the market recognized some of the information contained in their measure upon publishing, it was slow to appreciate the information fully (Ou and Penman, 1989). Finally, some researchers have shown that investors may earn abnormal returns by focusing on financial ratios which take market valuations into consideration. These include buying firms with low Price-to-Earnings (P/E) or high B/M ratios, both strategies which were able to generate excess returns in relation to CAPM (Basu, 1977; Rosenberg et. al, 1998). Like previously mentioned anomalies, the fundamental analysis of accounting information may also challenge the efficient market hypothesis since no usage of public information should be able to predict future price movements under the semi-strong form (Fama, 1970). In line with the result of studies showing the potential success of fundamental-based investing, some economists have published simple but more encompassing investment strategies.

2.2.2. Piotroski's F-score methodology

Piotroski (2000) examined whether an investment strategy based on accounting fundamentals could shift the return distribution earned by an investor when applied to high book-to-market (B/M) firms. Prior research had shown that a high book-to-market investment strategy outperformed a low book-to-market investment strategy, and the return explanation had been divided into both market-efficient (Fama and French, 1993; Rosenberg et al., 1998) and a market mispricing explanation (Lakonishok, et. al, 1994). However, Piotroski documented that less than 44% of high B/M firms earn positive market-adjusted returns in the two years following portfolio formation. Thus, by considering variables that could discriminate between weak and strong prospects, an investor would be able to avoid suffering the lower returns of deteriorating firms. (Piotroski, 2000).

Piotroski (2000) focused on nine accounting signals across three categories of financial position: profitability, financial leverage/liquidity, and operating efficiency. At the firm level, a score of one was then assigned for each signal achieved and zero for those not achieved (see Table 1). The firm-specific F-score was then equal to the sum of scores reached that year. Firms with F-scores of 8-9 were considered to have the strongest outlook, while those with F-scores of 0-1 were considered to have the weakest outlook. (Piotroski, 2000).

Table 1. F-score metrics

Variable	Criteria	Explanation
F_ROA	$\frac{Net\ Income_t}{Assets_{t-1}} > 0$	Positive Net Income during the last fiscal year
F_CFO	$\frac{CFO_t}{Assets_{t-1}} > 0$	Positive Cash Flow from Operations during the last fiscal year
F_ΔROA	$\frac{Net\ Income_t}{Assets_{t-1}} > \frac{Net\ Income_{t-1}}{Assets_{t-2}}$	Increase in Return on Assets during the last fiscal year
F_ACCRUAL	$\frac{CFO_t}{Assets_{t-1}} > \frac{Net\ Income_t}{Assets_{t-1}}$	CFO larger than Net Income during the last fiscal year
F_$\Delta LEVERAGE$	$\frac{Long\ Term\ Debt_t}{Assets_t} < \frac{Long\ Term\ Debt_{t-1}}{Assets_{t-1}}$	Decrease in Long Term Debt during the last fiscal year
F_$\Delta LIQUID$	$\frac{Current\ Assets_t}{Current\ Liabilities_t} > \frac{Current\ Assets_{t-1}}{Current\ Liabilities_{t-1}}$	Increase in Current Ratio during the last fiscal year
EQ_OFFER	$Equity\ Offering_t = 0$	No equity issued during the last fiscal year
F_$\Delta MARGIN$	$\frac{Revenue_t - COGS_t}{Revenue_t} > \frac{Revenue_{t-1} - COGS_{t-1}}{Revenue_{t-1}}$	Increase in Gross Margin during the last fiscal year
F_$\Delta TURN$	$\frac{Revenue_t}{Assets_{t-1}} > \frac{Revenue_{t-1}}{Assets_{t-2}}$	Increase in Asset Turnover ratio during the last fiscal year

Piotroski (2000) argued that fundamental analysis of accounting information was more beneficial among value stocks, as compared to glamor stocks, since valuations for the latter are typically based on long-term forecasts of sales and cash flows. As such, inventors rely heavily on non-financial information to value these glamor stocks. By contrast, the author identified three main reasons why fundamental analysis was beneficial to differentiate among value (high B/M) firms. (1) Value stocks tend to be neglected, with a general lack of analyst forecasts and stock recommendations. (2) Poor recent performance and low access to informal information channels make financial statements the most accessible and reliable information source for these firms. (3) A tendency of financial distress among value firms places a larger valuation focus on leverage, liquidity, profitability trends, and cash flow adequacy. Subsequently, Piotroski designed the nine accounting signals to specifically address important factors for value firms. For instance, an increase in leverage (F_$\Delta LEVERAGE$) could be positive under the right conditions but is most likely negative in situations of financial distress. (Piotroski, 2000).

2.2.3. Performance of Piotroski's F-score

Piotroski demonstrated that the mean returns of a high book-to-market investor could be increased by 7.5% annually by delimiting themselves to stocks with F-scores of 8-9. Similarly, an investor who invested in winners (F-scores of 8-9) and shorted losers (F-scores of 0-1) could generate annual returns of 23.5% between 1976 and 1996 (Piotroski, 2000). Other researchers have later replicated Piotroski's strategy to cover more recent time periods. For instance, Woodley et al. (2011) studied the U.S. market between 1976 and 2008. They first confirmed Piotroski's findings that high F-score stocks outperformed other value stocks between 1976 to 1996. However, in the subsequent 1997 to 2008 period, these effects had reversed. More specifically, mean market-adjusted returns of high F-score stocks were lower than those of low F-score stocks and the complete value group by 26.5% and 23.7% respectively (Woodley et al., 2011).

Mohanram (2005) evaluated the contextual application of Piotroski's F-score between 1979 and 1999. He demonstrated that the F-score can separate winners from losers both among glamor and value stocks, but that the annualized return difference for the first (9.8%) paled compared to the latter (20%). The author also demonstrated that relatively fewer firms are classified in the tails of the F-score among glamor stocks, rendering it less useful. As such, the findings supported the contextual importance of Piotroski's F-score. (Mohanram, 2005).

Piotroski and So (2012) examined the performance of the F-score over the 1972 to 2010 period for firms in the top 30%, middle 40%, and bottom 30% book-to-market bracket. They found that a strategy that goes long on high B/M firms with F-scores of 7-9 (winners) and shorts low B/M firms with F-scores of 0-3 (losers) generated significant returns. The researchers redefined the short leg of the portfolio to cover glamor stocks since poor accounting fundamentals in this group should indicate that high valuations are especially unjustified. When delimiting themselves to buying winners and shorting losers in the high B/M group, they found that returns had decreased to 6% in the 1972-2010 period as compared to the 23.5% earned in Piotroski's first study. (Piotroski and So, 2012).

Finally, Walkshäusl (2020) studied the F-score's applicability in a more modern 2000-2018 period for developed non-US markets. He found that high F-score firms outperformed low F-score firms on average by 9.9% per year, and that positive results were significant for different B/M groups. However, the return performance between B/M groups was not analyzed further. (Walkshäusl, 2020).

2.3. Value effect, accounting relevance, and intangibles

2.3.1. Recent underperformance of the value effect

Maloney and Moskowitz (2020) stated that the HML book-to-market equity factor has had a flat performance in the US in the last two decades, with negative returns in the latter. The authors examined whether the interest rate environment could be attributed to such an effect. This theory suggests that falling bond yields since the 2010s should have had a more adverse effect on value stocks due to their tendency to be financially distressed. However, no robust link was found between the decline of the value factor and the interest rate environment. (Maloney and Moskowitz, 2020).

Lev and Srivastava (2019) noted that the value (HML) factor has lost much of its potency since the 1980s and only had a brief resurgence in the early 2000s before its ultimate demise in 2007. They argued that the increase in corporate intangible investments has led to substantial and growing accounting deficiencies which adversely affects value investing. First, since internally generated intangibles are immediately expensed under U.S. GAAP instead of capitalized (as with tangible assets), they will be absent from book values even if such investments are intended to support future profits. As such, firms investing heavily in intangible assets may falsely appear as overvalued according to the B/M ratio due to understated book equity levels. Second, the authors argued that valuation metrics based on earnings such as P/E ratios have become overstated for firms investing heavily in intangibles as such investments understate earnings figures. At the time of their publication, the rate of intangible investments was roughly twice the rate of investments in tangible assets in the U.S. corporate sector. (Lev and Srivastava, 2019).

Lev and Srivastava (2019) also provided evidence that there has been a substantial slowdown of mean reversion in value and glamor stocks since 2007, which historically has accounted for the gains from the HML (value) factor. They argued that a prolonged decline in bank lending following the financial crisis has had an indirect effect on value stocks. This stems from their heavy reliance on debt to finance investments, due to their inability to issue stock at low valuations. As such, value stocks have lacked the much-needed capital to improve operations and escape the low-valuation trap. Furthermore, the decline in consumer demand following the drop in the U.S. housing market has further driven down valuations and impeded the recovery of value firms. On the opposite end, glamor stocks have had increased stability since the financial crisis. Firms classified in this group are largely reliant on scalable intangible assets protected by patents and brands. Such business models are associated with more entrenched customer relationships and stronger barriers to entry. This has been rewarded by investors through easy access to capital, which has enabled these firms to remain at the top for an extended period. As such, the authors argue that the prolonged deterioration of firms in the long end and the stability of firms in the short end has deteriorated the profitability of the HML factor. Finally, the authors identified which value firms were most likely to escape the value trap from 2008 to 2017. These included value firms with higher intangible investments, solid financials (indicating successful business models), and those with the ability to raise debt financing. (Lev and Srivastava, 2019).

2.3.2. Declining value relevance of accounting information

Researchers have also studied the value relevance of accounting information beyond its effect on book-to-market equity. Srivastava (2014) found a decreasing value relevance of earnings quality measures between 1970 and 2009. He attributed this to an increasing intangible intensity among firms, driven by new firms after 1970 mostly entering knowledge-intensive industries such as business services, pharmaceuticals, and computers. The author showed that intangible-intensive firms (proxied for by SG&A intensity) are likely to display higher volatility in earnings and cash flows since intangible investments, as compared to tangible investments, carry higher uncertainty about future economic benefits. Similarly, the immediate expensing of internally generated intangibles intended to support future profits has led to a decreased matching of costs and revenues, as well as higher expense volatility. The author further argued that intangible-intensive firms have growth options whose changes in values are not reflected in financial statements. Taken together, these effects have reduced the value relevance of earnings, and most significantly for intangible-intensive firms (Srivastava, 2014). Similarly, Dontoh et. al (2004) found a declining association between market value and accounting information between 1983 to 2000, which was especially pronounced for intangible-intensive companies as proxied by low book-to-market ratios.

Other academics have provided a more nuanced view of the decrease in value relevance. Balachandran and Mohanram (2011) studied the link between accounting conservatism and value relevance of accounting information in the 1975 to 2004 period. The authors capitalized and amortized R&D and advertising expenses and used the magnitude of the resulting book figures to proxy for conservatism. While they noted that value relevance has decreased over time, no evidence was found to indicate that value relevance was lower for firms with increasing conservatism. Instead, they observed the most significant declines in value relevance of accounting information for firms where conservatism had not increased. The authors further evaluated whether value relevance would increase when using adjusted income statement and balance sheet items to forecast stock returns, but found only similar or negative effects post-adjustment (Balachandran and Mohanram, 2011). Similarly, Francis and Schipper (1999) found that firms in high-technology industries have not experienced a greater decline in value relevance of accounting information than those in low-technology industries. The authors proxied for intangible-intensity through book-to-market ratios and R&D spending (Francis and Schipper, 1999).

Finally, a separate body of literature has discussed the option of adjusting accounting items for internally generated intangibles. Barker et al. (2022) evaluated this question from an accountant perspective, basing much of their discussion around the matching principle and the role of uncertainty in financial statement recognition. They argued that mismatching in the income statement is inevitable, occurring both for the expensing of intangible assets or the employment of an amortization schedule. For the first, the income calculation is upset by the failure to differentiate expenditure that is intended to generate future revenues. For the second, multiple periods of reporting will be affected. Thus, poor amortization and impairments, when uncertainties in future benefits from intangible investments are too high, will also result in mismatching which is compounded in later periods. This will deteriorate the informativeness of the income statement. As such, the authors recommended conditional capitalization of intangibles, contingent on the level of uncertainty of the investments (Barker et al., 2022). Penman (2009) similarly argued that the accountant should abstain from capitalizing intangibles as it is a speculative art, but did not condone the tendency of investors to add back amortization charges to earnings figures.

2.3.3. Capturing the value effect through accounting adjustments

Owing to the consensus that accounting deficiencies have distorted book equity, several studies have attempted to adjust the HML book-to-market equity factor for the expensing of internally generated intangibles through capitalization and amortization schedules. Many of these stem from Peters and Taylor (2017), who measured intangible capital as the sum of *organization capital* and *knowledge capital*. According to the authors, organization capital comes from SG&A expenditures and includes a firm's human capital, brand, customer relationships, and distribution systems. Meanwhile, knowledge capital comes from R&D expenditures and includes patent development, software, innovation, and the like. The authors assumed that 30% of SG&A and 100% of R&D expenditures could be attributed to a firm's investments in intangible capital to support future profits. To estimate a firm's intangible capital at a point in time, past SG&A and R&D spending was accumulated using the perpetual inventory method. This assumes that the current intangible capital is equal to the prior period's intangible capital less depreciation, plus the current period's investments in internally generated intangibles (estimated as described above). A flat 20% depreciation rate was assumed for organization capital (SG&A) and industry-specific rates ranging between 10% to 40% were applied for knowledge capital (R&D). To estimate the initial stock of intangible capital at listing, the authors applied pre-IPO growth rates to yield values for R&D spending between the founding year and IPO year. (Peters and Taylor, 2017).

Using the method by Peters and Taylor (2017), Arnott et. al (2021) constructed an alternative iB/M measure to better proxy for value by replacing book equity with intangible-adjusted book equity. They found that between 1963 and mid-2020 the iHML factor outperformed the traditional HML factor in the U.S. by an average of 1.3% per year. Between 2007 - 2020, this performance gap was even wider at an average of 2.2%. While the authors found that their iHML factor was still unable to beat sales-to-price (S/P) or earnings-to-price (E/P) strategies post-2007, they argued that the relative valuation of the HML factor was in its most attractive valuation percentile in history at the time of writing. (Arnott et al., 2021).

Eisfeldt et. al (2022) similarly computed an intangible adjusted HML factor in the U.S. by using the perpetual inventory method. However, they did not treat SG&A and R&D expenditures separately in the capitalization and amortization process, and formed within-industry terciles for long-short positions using Fama and French's 12 industry classification. Specifically, they capitalized 100% of SG&A expenditures. Two main reasons were laid out to support these decisions. First, they noted that the B/M ratio's ability to predict stock returns is mostly driven by within-industry variation. Measuring value within industries thus reduces noise and exposure to unpriced risk, which increases the Sharpe ratio. Second, heterogeneous accounting practices increase the risk of industry under- or overweighting if industries are not accounted for in the value sort. R&D expenditures are sometimes broken out separately from SG&A expenses, and may even be reported under COGS, meaning that estimates for knowledge capital in Peters and Taylor's method (2017) will be understated when separate R&D expenditure items are missing. The authors found that their adjusted HML^{INT} factor correlated with the traditional HML factor by 76.2% which was sufficient to capture the value effect while having fewer pricing errors. They also showed that the adjusted HML^{INT} factor has significantly outperformed the traditional HML factor since 1975, but most excessively in the 2007 to 2018 period where the comparative alpha was 3.86%. The authors further noted that firms in the long leg of the HML^{INT} factor had superior fundamentals (productivity, earnings, and profitability) compared to firms in the long leg of the unadjusted HML factor. As such, they argued that intangible value firms may better be able to avoid value traps where market capitalization does not recover for high book-to-market firms. (Eisfeldt et al., 2022).

2.4. Hypotheses development

Below we summarize the main research findings underlying our study and then formalize our hypotheses.

(1) Early studies on Piotroski's F-score indicated that his strategy could generate significant returns. They also noted that the F-score could separate winners (high F-scores) from losers (low F-scores) and shift returns from a high book-to-market strategy (Mohanram, 2005; Piotroski, 2000). However, later U.S. studies have seen a decline in these abilities and in return performance (Piotroski and So, 2012; Woodley et al., 2011). This has coincided with research suggesting that investment strategies based on fundamental analysis, as well as the value effect, have deteriorated (Lev and Srivastava, 2019; Balachandran and Mohanram, 2011; Srivastava, 2014).

(2) The immediate expensing of intangibles has led to a substantial misclassification of value and glamor stocks, which has become more severe as the economy has increasingly shifted from tangible to intangible investments in recent decades (Lev and Srivastava, 2019). Piotroski's F-score was originally designed and best able to identify winners and losers among value stocks, as compared to other book-to-market groups (Mohanram, 2005; Piotroski, 2000).

(3) Adjusting the book-to-market screening by artificially capitalizing investments in internally generated intangibles (proxied for by SG&A and/or R&D expenses) leads to fewer pricing errors in Fama French's Three Factor Model, indicating an improved identification of value and glamor stocks. Furthermore, this identification of value stocks is associated with improved returns compared to the ones identified by an unadjusted book-to-market ratio. (Arnott et al., 2021; Eisfeldt et al., 2022; Lev and Srivastava, 2019).

These findings jointly build the foundation for the three hypotheses tested in this study. Specifically, we hypothesize that the recent deterioration of Piotroski's F-score may be tied to an imperfect identification of value stocks using the book-to-market metric, which research finds has become outdated. As such, accounting for intangible investments may improve the ability of Piotroski's F-score to separate winners from losers and increase the return performance of a high F-score investment strategy. Our null hypotheses are as follows:

> ***$H_0$1*** *: Piotroski's high F-score strategy is not improved when adjusting the book-to-market screening for intangibles*

> ***$H_0$2*** *: Piotroski's F-score is not better able to separate winners from losers when adjusting the book-to-market screening for intangibles*

> ***$H_0$3:*** *Piotroski's high F-score strategy does not beat the market index when adjusting the book-to-market screening for intangibles*

3. Method

3.1. Research design

This study investigates whether the performance of Piotroski's F-score in the U.S. stock market may be improved by adjusting the book-to-market value screen for intangibles. The hypotheses are tested by replicating the returns of two sets of two portfolios based on Piotroski's F-score. Each set contains one equal-weighted high F-score portfolio and one equal-weighted low F-score portfolio. The first set of portfolios screens for value stocks via the B/M metric, while the other utilizes the intangible-adjusted B^{INT}/M metric computed in line with Eisfeldt et al. (2022). The stock market is proxied for by the S&P 500 value-weighted index, which has a wide investor following and captures roughly 80% of available market capitalization (S&P Global, 2023). Hypotheses 1 and 3 are tested by comparing returns of the high F-score portfolio using a B^{INT}/M screening to the corresponding high F-score portfolio using a B/M screening and the S&P 500 index respectively. Hypothesis 2 is tested in two steps. First, the returns of the high F-score portfolio are tested against the low F-score portfolio when the B^{INT}/M screening is used. Second, we test whether the return difference between high and low F-score portfolios is greater when using the B^{INT}/M screening than the unadjusted B/M screening.

Portfolios are rebalanced on a yearly basis. Data necessary to compute book-to-market ratios and F-score components is based on fiscal data from December in year t. To avoid the look-ahead bias, stocks to the portfolios are selected at the start of May year t+1 (Piotroski, 2000). The portfolios are then held until the end of April year t+2 and are then rebalanced based on data from December in year t+1. This procedure starts in May 2000 and runs to the end of April 2022, yielding 264 monthly return observations per portfolio. The number of firms contained in each portfolio may naturally differ as an unequal number of firms qualify for selection.

Significant return differences between portfolios, and between the market index and the intangible-adjusted F-score portfolio, are tested for via univariate tests. We do not test for risk-adjusted returns. Hence, we do not draw conclusions on whether any potential excess return is abnormal, or risk-based. This method choice is further supported by the argument that the value factor, which is often included in modern asset pricing models, is misspecified due to the expensing of internally generated intangibles (Arnott et al., 2021; Lev and Srivastava, 2019). A more detailed outlay of our method is provided in subsequent subsections.

3.2. Sample selection

We obtain monthly stock returns from CRSP and annual accounting data from Compustat. Our sample starts with all firms listed on the main U.S. exchanges (NASDAQ, AMEX, NYSE) between May 2000 to April 2022. We exclude securities other than common shares and drop financial firms (SIC codes 6000-6999), regulated utilities (4900-4999), and firms categorized as public service, international affairs, or non-operating establishments (9000+) in line with Eisfeldt et. al (2022). Firm-year observations with insufficient data to calculate book-to-market equity, stock returns, and any one of the F-score components are excluded, as well as observations with negative book equity. Stock returns are adjusted for stock splits and cash distributions, and delisting returns are used to adjust stock returns when available. If the delisting return is missing and the delisting reason is performance related, we set the delisting return to -30% (Shumway, 1997). We obtain Fama-French 12 industry classifications from Kenneth French's data library (2023) and gather return data for the value-weighted S&P 500 index from CRSP. Our final sample consists of 5,772 unique firms and 457,088 firm-month observations. To estimate the intangible component of equity for firms in the early 2000s, we separately gather Compustat input data for the perpetual inventory method ranging back to December 1990 and subsequently match the model output to our main sample year-by-year.

3.3. Computation of book-to-market ratios

Traditional B/M ratio

We compute book equity per the definition of Davis, Fama, and French (2000).[1] The market value of equity is defined as the number of shares outstanding times the closing stock price of December month in year t, preceding portfolio formation in year t+1. The B/M ratio is then computed as book equity divided by the market value of equity.

Intangible-adjusted B^{INT}/M ratio

We follow Eisfeldt, et. al (2022) in the computation of the intangible-adjusted B^{INT}/M ratio. Each year, we capitalize 100% of the amount reported in the SG&A variable on Compustat to book equity as internally generated intangible capital (INT_{it}). Owing to accounting treatment

[1] Book equity is defined as stockholder's equity, plus balance-sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the value of preferred stock. If stockholder's equity is unavailable in Compustat, we measure it as the book value of common equity plus the par value of preferred stock, or the book value of assets minus total liabilities (in that order).

by the database, this also includes any R&D expenses reported separately by the firm. These expenses are then amortized in line with the perpetual inventory method:

$$INT_{it} = (1 - \delta_{SG\&A})INT_{it-1} + SG\&A_{it}$$

where $\delta_{SG\&A} = 0.2$, representing the yearly amortization rate. This calculation is applied to annual accounting data from the Compustat Annual Fundamentals file from December 1990 to December 2021. To complete the perpetual inventory method, an initial stock of intangible capital (INT_{i0}) must be computed at the first observation made for each firm by Compustat within our sample. In line with Eisfeldt, et. al (2022), we compute:

$$INT_{i0} = SG\&A_{i1}/(g + \delta_{SG\&A})$$

where $g = 0.1$, or approximately the historical average rise in SG&A expenses. In addition to adjustments for internally generated intangible capital, we deduct goodwill from book equity. This avoids double counting of intangibles and reduces the effect of M&A activity, making adjusted book equity more comparable between firms in our sample. Keeping the calculation of book equity the same as in the unadjusted B/M ratio, we can then estimate the intangible-adjusted book value of equity (B_t^{INT}) as:

$$B_{it}^{INT} = B_{it} + INT_{it} - Goodwill_{it}$$

We use the same approach to calculate the market value of equity as for the traditional B/M ratio to reach the intangible-adjusted B^{INT}/M ratio. Finally, we use the Fama-French 12 industry classification to group stocks according to B^{INT}/M within each industry. This addresses differences in accounting treatment for intangibles across industries, as well as the effect of differences in investment and amortization rates. (Eisfeldt, et al., 2022).

3.4. Creation of F-score portfolios

Piotroski (2000) designed his F-score metrics specifically to identify strong prospects among value stocks, which he classified as firms in the top B/M quintile (20%) of the stock market. We also apply the F-score strategy on value stocks but use two separate proxies: (1) the 30% highest B/M stocks per year, and (2) the 30% highest B^{INT}/M stocks per year within each of the

Fama-French 12 industries. The cutoff percentage is increased from 20% to 30% to reduce sensitivity to assumptions in the perpetual inventory method, as it may result in a significant reclassification of value stocks (Lev and Srivastava, 2019), while still being in line with value definitions by other researchers (Fama and French, 1993; Piotroski and So, 2012). In these two samples, stocks are then evaluated based on the nine metrics laid out by Piotroski (2000). The stock receives a score of 1 if the criterion of the individual metric is met and a score of 0 otherwise. The metrics and criteria are as follows:

1. $ROA : \frac{Net\ Income_t}{Assets_{t-1}} > 0$
2. $CFO : \frac{CFO_t}{Assets_{t-1}} > 0$
3. $\Delta ROA : \frac{Net\ Income_t}{Assets_{t-1}} > \frac{Net\ Income_{t-1}}{Assets_{t-2}}$
4. $ACCRUAL : \frac{CFO_t}{Assets_{t-1}} > \frac{Net\ Income_t}{Assets_{t-1}}$
5. $\Delta LEVERAGE : \frac{Long\ Term\ Debt_t}{Assets_t} < \frac{Long\ Term\ Debt_{t-1}}{Assets_{t-1}}$
6. $\Delta LIQUID : \frac{Current\ Assets_t}{Current\ Liabilities_t} > \frac{Current\ Assets_{t-1}}{Current\ Liabilities_{t-1}}$
7. $EQ\_OFFER : No\ equity\ issued\ during\ last\ fiscal\ year$
8. $\Delta MARGIN : \frac{Revenue_t - COGS_t}{Revenue_t} > \frac{Revenue_{t-1} - COGS_{t-1}}{Revenue_{t-1}}$
9. $\Delta TURN : \frac{Revenue_t}{Assets_{t-1}} > \frac{Revenue_{t-1}}{Assets_{t-2}}$

Each stock then receives a composite F-score ranging from 0-9 based on the sum of the nine individual scores. Piotroski's original strategy was based on investing in stocks with the strongest prospects, identified as having F-scores of 8-9. He further identified weak prospects as firms with F-scores of 0-1. However, we note that fewer observations end up in F-score tails in our period compared to Piotroski's (2000). Subsequently, we follow Piotroski and So (2012) and include firms with F-scores of 7-9 in high F-score portfolios and firms with F-scores of 0-3 in low F-score portfolios. This yields two high F-score and two low F-score portfolios:

1. Buy stocks with F-scores of 7-9 in the 30% highest B/M group
2. Buy stocks with F-scores of 0-3 in the 30% highest B/M group
3. Buy stocks with F-scores of 7-9 in the 30% highest B^{INT}/M group in each industry
4. Buy stocks with F-scores of 0-3 in the 30% highest B^{INT}/M group in each industry

We will refer to portfolio (1) as High F-score^{REG}, portfolio (2) as Low F-score^{REG}, portfolio (3) as High F-score^{INT}, and portfolio (4) as Low F-score^{INT} going forward.

3.5. Model description

Testing of the null hypothesis is done by comparing monthly returns between the portfolios, as well as between the High F-score^{INT} portfolio and the market index. Kothari and Warner (1997) argued that parametric long-horizon tests often indicate abnormal performance even if none is present, reducing the integrity of these tests. As such, we use both parametric and non-parametric tests in our study, in line with the methodology of Piotroski (2000). Specifically, we employ one-tailed paired-sample t-tests and Wilcoxon signed rank tests to test for differences in monthly portfolio returns. The t-test evaluates mean returns while the Wilcoxon test evaluates median returns, which makes the results more robust to outliers in return data. A null hypothesis is only rejected if both univariate tests support a rejection, which reduces the reliance on any one univariate test. Null hypothesis 2 is only rejected if both steps involved in the hypothesis indicate a significantly positive return difference. This is done to validate that a portfolio of high F-score firms (winners) significantly outperforms one of low F-score firms (losers) among high B^{INT}/M stocks, otherwise the ability to generate a significantly higher return difference when adjusting the book-to-market screening is made less meaningful. The significance level upon rejection is determined by the highest p-value among the tests performed.

3.6. Robustness tests

The robustness of our intangible adjusted book equity measure is assessed by varying assumptions entering the perpetual inventory method. These may materially impact which stocks are reclassified as value stocks according to the B^{INT}/M ratio. Following Eisfeldt et al. (2022) we conduct a sensitivity analysis on amortization rates from 10% to 50%, and on investment rates from 50% to 100%. We also test that any returns from the High F-score^{INT} portfolio are not purely driven by the within-industry book-to-market sorting. Lastly, we test whether any superior return performance for the High F-score^{INT} portfolio, or higher return difference between high and low F-score portfolios among high B^{INT}/M stocks, is robust for shorter time periods. This is done by splitting the sample period in two (2000-2011 and 2011-2022), which enables us to see if there is a lower performance in later sample years (Lev and Srivastava, 2019). The robustness tests employ the same univariate tests as already described.

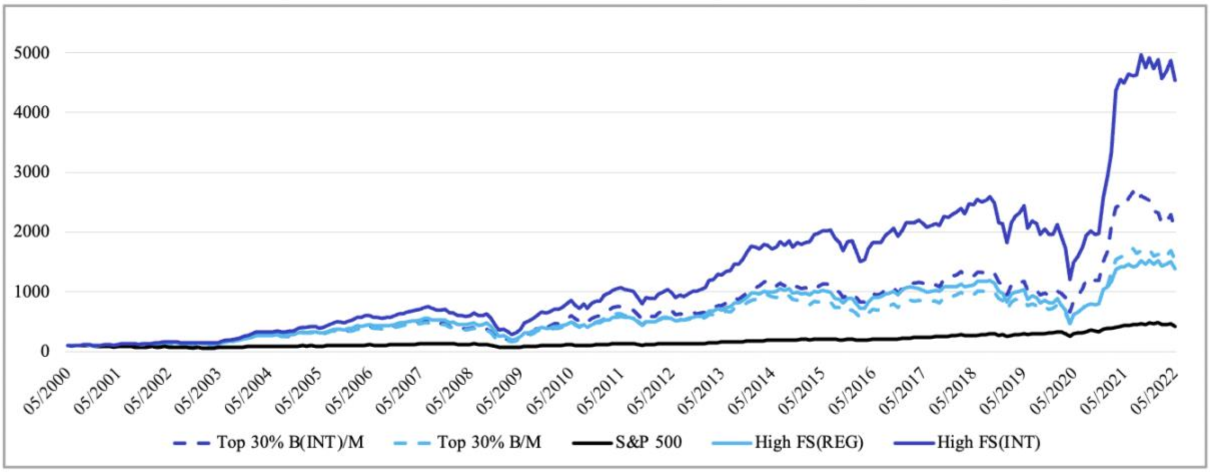
4. Results

In this section, we first present descriptive statistics for our sample before presenting the results of our hypothesis tests and concluding with further analysis and robustness checks. The results suggest that the High F-score^{INT} portfolio has outperformed both the S&P 500 index and the High F-score^{REG} portfolio in the 21st century. In addition, the F-score is more efficient at distinguishing between winners and losers when adjusting the book-to-market screening for intangibles. As such, all three null hypotheses are rejected.

4.1. Descriptive statistics

Figure 1 shows the indexed returns for the S&P 500 index, the high F-score portfolios, and the returns of the complete high book-to-market groups from May 2000 to April 2022. We first note that an early investment in either high F-score portfolio would in the long run have yielded returns above the S&P 500 index. There appears to be a correlation between both high F-score and book-to-market portfolios. While the High F-score^{INT} portfolio has been consistently above the High F-score^{REG} portfolio and book-to-market portfolios, the return gap is most prevalent post-2009. Interestingly, the High F-score^{REG} portfolio ends up accumulating a lower return than its underlying book-to-market group over the entire 22-year period.

Figure 1. Accumulated returns indexed to May 1st, 2000



Panel A of Table 2 offers a more in-depth analysis of annual returns. The CAGR for both the High F-score^{INT} and High F-score^{REG} portfolios declined from 24.0% and 17.4% to 14.1% and 8.2% respectively from the 2000-2011 to 2011-2022 period. Further, only the High F-score^{INT} portfolio manages to match the returns of the S&P 500 index in the 2011-2022 period. Notably, the return gap to the underlying book-to-market group is positive for the High F-score^{INT} portfolio (18.9% – 14.6%) whereas it is negative for the High F-score^{REG} portfolio (12.7% – 13.2%) over the total 22-year period. This is also true for both subperiods. Panel B of Table 2 covers the annual return difference between high and low F-score portfolios for both the B^{INT}/M and B/M value screening. Over the entire 22-year period, there is a wider return difference between high and low F-score portfolios when intangible adjustments are made to the book-to-market ratio (5.8%) compared to when they are not made (0.0%). The return gap is wider for the B^{INT}/M screening during both the 2000-2011 and the 2011-2022 periods compared to the B/M screening, which even has a negative return gap during the second period (-1.1%).

Table 2. One-year returns for the S&P 500, high F-score and B/M portfolios (%)

Panel A: High F-score portfolios

	00/01	**01/02**	**02/03**	**03/04**	**04/05**	**05/06**	**06/07**	**07/08**	**08/09**	**09/10**	**10/11**	**CAGR (00-11)**
High FS^{INT}	17.9	36.4	-5.1	112.2	20.3	53.6	19.0	-14.6	-38.5	127.4	25.6	24.0
High FS^{REG}	14.7	34.8	-13.6	97.1	16.9	46.0	17.4	-13.1	-44.5	91.5	19.8	17.4
High B^{INT}/M	-5.0	32.2	-8.8	138.9	5.6	47.2	16.2	-22.1	-37.9	153.9	24.1	20.1
High B/M	-0.3	31.5	-8.8	122.7	8.6	41.2	15.0	-21.4	-41.1	132.0	23.9	18.1
S&P 500	-14.6	-10.1	-12.9	26.9	7.5	20.6	14.8	-3.3	-35.2	41.8	18.9	2.7

	11/12	**12/13**	**13/14**	**14/15**	**15/16**	**16/17**	**17/18**	**18/19**	**19/20**	**20/21**	**21/22**	**CAGR (11-22)**	**CAGR (00-22)**
High FS^{INT}	-6.9	28.5	34.0	17.6	-9.4	17.1	14.9	-0.4	-38.8	201.4	0.8	14.1	18.9
High FS^{REG}	-5.3	25.7	41.8	3.3	-11.1	13.3	8.0	-6.9	-40.1	128.9	-1.9	8.2	12.7
High B^{INT}/M	-9.2	11.7	46.8	0.3	-14.4	18.9	10.8	-7.2	-28.3	188.9	-17.3	9.4	14.6
High B/M	-8.1	10.3	44.3	-6.0	-18.4	20.3	13.0	-8.1	-31.9	167.4	-3.1	8.6	13.2
S&P 500	1.1	15.8	19.3	11.0	-2.0	17.7	12.4	11.2	-2.3	50.3	-4.4	11.0	6.7

Panel B: High – Low F-score Portfolios

	00/01	**01/02**	**02/03**	**03/04**	**04/05**	**05/06**	**06/07**	**07/08**	**08/09**	**09/10**	**10/11**	**AVG (00-11)**
High-Low^{REG}	24.0	1.6	-6.9	-20.1	9.8	-1.1	11.7	23.7	-4.3	-25.3	-1.9	1.0
High-Low^{INT}	27.4	0.4	2.2	-20.1	23.1	2.1	7.8	20.5	2.9	-17.5	3.1	4.7

	11/12	**12/13**	**13/14**	**14/15**	**15/16**	**16/17**	**17/18**	**18/19**	**19/20**	**20/21**	**21/22**	**AVG (11-22)**	**AVG (00-22)**
High-Low^{REG}	6.0	22.3	-6.6	15.5	15.0	-6.1	-5.4	-2.8	-15.0	-29.0	-5.9	-1.1	0.0
High-Low^{INT}	7.9	25.7	-13.9	24.0	19.0	0.5	3.7	7.2	-16.1	-2.8	20.8	6.9	5.8

Notes: This table presents the 12-month buy-and-hold raw returns for the S&P 500 index, High FS^{INT}, High FS^{REG}, High B^{INT}/M, and High B/M portfolios over the period May 2000 - April 2022. The High B^{INT}/M portfolio consists of firms above the 70th percentile of B^{INT}/M values in each industry, while the High B/M portfolio consists of firms above the 70th percentile of B/M values irrespective of industry. *FS* is used as an abbreviation for *F-score*. All numbers are denoted in percent (%).

Table 3 provides data for different quality and valuation metrics, as well as return data, for the different high and low F-score portfolios over the May 2000 to April 2022 period. Firms contained in the high F-score portfolios are generally larger than their low F-score counterparts, both in terms of asset values and market capitalization. Further, both high F-score portfolios have lower B/M and B^{INT}/M values. Firm characteristics also differ between portfolios formed on the intangible-adjusted book-to-market screening compared to the unadjusted screening. Intangible-adjusted portfolios have higher B^{INT}/M values and lower market capitalizations. Furthermore, the ratio between the number of High F-score^{INT} firms and Low F-score^{INT} firms is lower than between the number of High F-score^{REG} firms and Low F-score^{REG} firms.

Table 3. Descriptive statistics for firms included in the high and low F-score portfolios

2000-2022	**High FS^{INT} (obs = 1695)**			**Low FS^{INT} (obs = 5698)**			**High FS^{REG} (obs = 1991)**			**Low FS^{REG} (obs =5153)**		
Variables	**mean**	**median**	**std. dev.**	**mean**	**median**	**std. dev.**	**mean**	**median**	**std. dev.**	**mean**	**median**	**std. dev.**
Monthly returns	0.0172	0.0198	0.0735	0.0123	0.0100	0.0860	0.0127	0.0164	0.0730	0.0129	0.0118	0.0822
Size metrics												
Market cap.	1135.9	123.2	5430.6	359.7	57.0	2048.7	3624.3	293.0	15901.0	817.6	76.2	6047.3
Assets	1964.5	206.9	9009.8	1014.7	93.6	8053.5	5882.7	547.8	21320.1	1998.8	172.2	10652.5
Value metrics												
B^{INT}/M	3.803	2.594	4.273	5.111	3.207	7.137	2.887	1.765	3.927	5.073	2.940	7.912
B/M	1.068	0.801	0.985	1.289	0.894	1.721	1.225	0.974	0.862	1.653	1.193	1.763
Quality metrics												
ROA	0.031	0.039	0.433	-0.191	-0.103	0.507	0.033	0.038	0.333	-0.132	-0.089	0.745
ΔROA	0.069	0.038	0.247	-0.054	-0.052	1.004	0.046	0.026	0.339	-0.068	-0.071	1.398
CFO / Assets	0.114	0.102	0.187	-0.103	-0.017	0.305	0.111	0.096	0.152	-0.063	-0.013	0.226
EQ-offer	0.456	0.000	0.498	0.196	0.000	0.397	0.469	0.000	0.499	0.146	0.000	0.353
ΔCurrent ratio	0.236	0.130	1.201	-0.518	-0.201	4.112	0.232	0.115	1.344	-0.651	-0.248	4.280
ΔAsset turnover	0.169	0.083	1.184	-0.117	-0.049	0.478	0.088	0.059	0.291	-0.175	-0.081	0.502
ΔGross margin	0.213	0.013	8.840	-1.422	-0.016	77.714	0.210	0.013	8.170	-6.301	-0.029	152.560
ΔLTD / Assets	-0.025	-0.013	0.082	0.017	0.000	0.103	-0.022	-0.015	0.076	0.022	0.002	0.097
Accruals	0.083	0.059	0.310	0.089	0.084	0.419	0.078	0.057	0.308	0.069	0.082	0.759

Notes: This table summarizes key value investing metrics related to the F-score and monthly returns over the entire period for both high F-score and low F-score portfolios. Observations correspond to the total number of firm-year observations per portfolio. *FS* is used as an abbreviation for *F-score*. All variables except ratios and returns are denoted in Million USD.

4.2. Hypothesis testing

Null hypothesis 1: *Piotroski's high F-score strategy is not improved when adjusting the book-to-market screening for intangibles*

Table 4 shows the results for the one-tailed paired t-test and the one-tailed Wilcoxon signed rank test for the mean and median monthly return difference between the High F-score^{INT} and High F-score^{REG} portfolios for the May 2000 - April 2022 period. The mean and median monthly return differences were 0.45% and 0.35% respectively, and both differences are significant at the 1% level. Since both tests indicate that adjusting the book-to-market screening for internally generated intangibles increases returns of a high F-score strategy, we reject the first null hypothesis at the 1% level.

Table 4. High F-score^{INT}'s performance compared to the High F-score^{REG} portfolio

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{INT}	0.0172	0.0045	0.0198	264
High FS^{REG}	0.0127	0.0045	0.0164	264
High FS^{INT} – High FS^{REG}	0.0045	0.0013	0.0035	264
T-statistic (Z-value)	3.592***	–	(3.972)***	–
P-value	0.0002	–	0.0001	–

Notes: This table presents the results for the paired sample t-test (T-statistic) and the Wilcoxon signed rank test (Z-value) on monthly returns over the May 2000 - April 2022 period for the two high F-score portfolios. *FS* is used as an abbreviation for *F-score* in the table. Significance levels are indicated by: *p<0.1; **p< 0.05; ***p<0.01.

Null hypothesis 2: *Piotroski's F-score is not better able to separate winners from losers when adjusting the book-to-market screening for intangibles*

Table 5 displays the results for univariate tests on the return difference between different long-short portfolios for the May 2000 - April 2022 period. Panel A shows the results for the return difference between the High F-score^{INT} and Low F-score^{INT} portfolios. Panel B shows the results for the return difference between an intangible-adjusted long-short portfolio (High-Low F-score^{INT}) and the equivalent unadjusted long-short portfolio (High-Low F-score^{REG}). In Panel A, the mean and median monthly return differences between the High F-score^{INT} and Low F-score^{INT} portfolios corresponded to 0.49% and 0.98%. The differences are significant at the 5% and 1% levels respectively. In Panel B, the mean and median monthly return differences between the two long-short portfolios corresponded to 0.51% and 0.52%. Both differences are significant at the 1% level. Since all tests indicate that adjusting the book-to-market measure for internally generated intangibles improves the ability of Piotroski's F-score to separate winners from losers, we reject the second null hypothesis at the 5% level.

Table 5. F-score's ability to separate winners from losers among high B^{INT}/M firms

Panel A: High F-score^{INT} vs. Low F-score^{INT}

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{INT}	0.0172	0.0045	0.0198	264
Low FS^{INT}	0.0123	0.0053	0.0100	264
High FS^{INT} – Low FS^{INT}	0.0049	0.0027	0.0098	264
T-statistic (Z-value)	1.837**	–	(3.314)***	–
P-value	0.0337	–	0.0004	–

Panel B: High-Low F-score^{INT} vs. High-Low F-score^{REG}

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{INT} – Low FS^{INT}	0.0049	0.0027	0.0098	264
High FS^{REG} – Low FS^{REG}	-0.0002	0.0024	0.0046	264
(High-Low)^{INT} – (High-Low)^{REG}	0.0051	0.0014	0.0052	264
T-statistic (Z-value)	3.669***	–	(4.285)***	–
P-value	0.0001	–	0.0000	–

Notes: This table presents the results for the paired sample t-test (T-statistic) and the Wilcoxon signed rank test (Z-value) on monthly returns over the May 2000 – April 2022 period for the High F-score^{INT} and Low F-score^{INT}, High-Low F-score^{INT}, and the High-Low F-score^{REG} portfolios. Each of the two High-Low F-score portfolios has a long position in the High F-score portfolio and a short position in the corresponding Low F-score portfolio (using the same book-to-market screening). *FS* is used as an abbreviation for *F-score* in the table. Significance levels are indicated by: *p<0.1, **p< 0.05, ***p<0.01.

Null hypothesis 3: *Piotroski's high F-score strategy does not beat the market index when adjusting the book-to-market screening for intangibles*

Table 6 displays the result for both univariate tests for the High F-score^{INT} portfolio against the value-weighted S&P 500 index over the period May 2000 – April 2022. The mean and median monthly return differences were 1.07% and 0.78% respectively and are both statistically significant at the 1% level. Since both tests indicate overperformance in relation to the market index, the third null hypothesis is rejected at the 1% significance level.

Table 6. High F-score^{INT}'s performance compared to the S&P 500 index

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{INT}	0.0172	0.0045	0.0198	264
S&P 500	0.0065	0.0028	0.0120	264
High FS^{INT} – S&P 500	0.0107	0.0028	0.0078	264
T-statistic (Z-value)	3.846***	–	(3.814)***	–
P-value	0.0001	–	0.0001	–

Notes: This table presents the results for the paired sample t-test (T-statistic) and the Wilcoxon signed rank test (Z-value) on monthly returns over the full period for the High F-score^{INT} portfolio and the value-weighted S&P 500 index. *FS* is used as an abbreviation for *F-score* in the table. Significance levels are indicated by: *p<0.1; **p< 0.05; ***p<0.01.

4.3. Further analysis

Does Piotroski's F-score beat a simpler high B^{INT}/M strategy?

Since Piotroski's F-score aims to shift the returns of a simple high book-to-market strategy, a further concern is whether the High F-score^{INT} portfolio can significantly outperform the underlying high B^{INT}/M group, even if winners are better identified in this group. Table 7 displays the return difference between the High F-score^{INT} portfolio and a portfolio formed on the complete underlying high B^{INT}/M group over the period May 2000 to April 2022. The mean and median monthly return differences between the two long-short portfolios corresponded to 0.28% and 0.60%. The return differences are significant at the 10% and 1% levels respectively. As such, we find support for the High F-score^{INT} portfolio having outperformed the complete high B^{INT}/M group at the 10% significance level.

Table 7. High F-score^{INT}'s performance compared to the underlying High B^{INT}/M group

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{INT}	0.0172	0.0045	0.0198	264
High B^{INT}/M	0.0144	0.0048	0.0138	264
High FS^{INT} – High B^{INT}/M	0.0028	0.0019	0.0060	264
T-statistic (Z-value)	1.511*	–	(2.448)***	–
P-value	0.0659	–	0.0071	–

Notes: This table presents the results for the paired sample t-test (T-statistic) and the Wilcoxon signed rank test (Z-value) on monthly returns over the May 2000 – April 2022 period for the High F-score^{INT} and the High B^{INT}/M portfolios. The High B^{INT}/M portfolio consists of firms above the 70th percentile of B^{INT}/M values within each industry. *FS* is used as an abbreviation for *F-score* in the table. Significance levels are indicated by: *p<0.1; **p< 0.05; ***p<0.01.

Is Piotroski's F-score able to separate winners from losers among high B/M stocks?

In our main results, we demonstrate that a portfolio of high F-score (7-9) firms significantly outperform one of low F-score (0-3) firms among high B^{INT}/M stocks in the 21st century. This indicates that Piotroski's F-score can significantly separate between strong and weak prospects in this group. Support is also found for this ability being superior among high B^{INT}/M stocks as compared to high B/M stocks. Subsequently, an area of interest is whether a portfolio of high F-score firms (expected winners) is even able to significantly outperform one of low F-score firms (expected losers) among the unadjusted high B/M group. Panel A of Table 8 displays the results of the univariate tests for the return difference between a high F-score (7-9) and low F-score (0-3) portfolio among high B/M stocks over the period May 2000 to April 2022. The Wilcoxon test indicates a significant return difference at the 10 % level, but the t-test indicates no significant return differences at conventional levels. For completeness, it is also relevant to test whether the High F-score^{REG} portfolio significantly outperforms the complete underlying

B/M group. Panel B of Table 8 displays the results of the univariate tests for the return difference between the High F-score^{REG} portfolio and the returns of a high B/M portfolio over the period May 2000 to April 2022. No significant return difference is found between these portfolios at conventional levels for any of the two tests.

Table 8. F-score's ability to separate winners from losers among high B/M firms

Panel A: Difference in High F-score^{REG} and Low F-score^{REG} returns

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{REG}	0.0127	0.0045	0.0164	264
Low FS^{REG}	0.0129	0.0051	0.0118	264
High FS^{REG} – Low FS^{REG}	-0.0002	0.0024	0.0046	264
T-statistic (Z-value)	-0.081	–	(1.428)*	–
P-value	0.5324	–	0.0769	–

Panel B: Difference in High F-score^{REG} and High B/M returns

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{REG}	0.0127	0.0045	0.0164	264
High B/M	0.0133	0.0047	0.0126	264
High FS^{REG} – High B/M	-0.0006	0.0015	0.0038	264
T-statistic (Z-value)	-0.374	–	(0.884)	–
P-value	0.6456	–	0.1887	–

Notes: This table presents the results for the paired sample t-test (T-statistic) and the Wilcoxon signed rank test (Z-value) on monthly returns over the May 2000 – April 2022 period for the High F-score^{REG}, Low F-score^{REG}, and the High B/M portfolios. The High B/M portfolio consists of firms above the 70th percentile of B/M values. *FS* is used as an abbreviation for *F-score* in the table. Significance levels are indicated by: *p<0.1; **p< 0.05; ***p<0.01.

4.4. Robustness tests

The robustness tests indicate that the rejection of all hypotheses is robust to varying investment rates and depreciation rates of the intangible capital stock (Appendix 3). Furthermore, we confirm that the superior performance of the High F-score^{INT} portfolio relative to the High F-score^{REG} portfolio is not driven purely by within-industry sorting (Appendix 4). We also find that the rejection of the first null hypothesis is robust in both the 2000-2011 and 2011-2022 subperiods (Appendix 5, Panel A). This increases the confidence in our conclusion that the High F-score^{INT} portfolio generates superior returns to the High F-score^{REG} portfolio. It also confirms that this performance is driven by the intangible adjustments to book equity and that it is insensitive to adjustments in these assumptions.

For other tests we find some mixed results. First, the rejection of the third null hypothesis is not robust when solely considering the 2011-2022 period (Appendix 5, Panel D). This suggests that the High F-score^{INT} portfolio's market outperformance is mainly driven by comparatively higher returns in earlier years. Second, expected winners do not significantly outperform expected losers in the 2000-2011 subperiod even when the book-to-market screening is intangible-adjusted (Appendix 5, Panel C). This is true despite that the return difference between high and low F-score firms (winners and losers) is greater when the book-to-market screening is adjusted compared to when it is not (Appendix 5, Panel B). As such, only the 2011-2022 subperiod in isolation fulfills both criteria for rejection of the second null hypothesis.

4.5. Summary of results

To summarize, our results indicate that a high F-score strategy is improved when the book-to-market value screen is adjusted for intangibles. This strategy has also outperformed the S&P 500 market index in the 21st century, which in large part is driven by strong performance in the first decade. Finally, the F-score becomes better at separating winners from losers with adjustments to book-to-market equity, and this ability is improved to such a degree that a strategy that invests in high F-score firms outperforms the complete high B^{INT}/M group. By comparison, the F-score with an unadjusted book-to-market screen is neither able to separate winners from losers nor significantly improve the returns of a high B/M investment strategy.

Table 9. Summary of null hypotheses outcomes

$H_0 1$	*Piotroski's high F-score strategy is not improved when adjusting the book-to-market screening for intangibles*	Rejected
$H_0 2$	*Piotroski's F-score is not better able to separate winners from losers when adjusting the book-to-market screening for intangibles*	Rejected
$H_0 3$	*Piotroski's high F-score strategy does not beat the market index when adjusting the book-to-market screening for intangibles*	Rejected

5. Discussion

In this section we will discuss our results and their possible implications. We will begin by discussing possible explanations for the observed improvement in return performance for Piotroski's F-score when adjusting the book-to-market screening for intangibles. This will be followed by a separate discussion on possible explanations for some inconsistencies in such performance, owing to the inability to separate winners from losers in the early 2000s, as well as the downturn in return performance in the 2010s.

5.1. Improved returns of an intangible-adjusted F-score approach

During our studied period we find evidence that applying Piotroski's F-score to an intangible adjusted book-to-market set of firms delivers significant positive returns. The strategy yielded annualized yearly returns of 18.9 % over the May 2000 - April 2022 period compared to the S&P 500 index (6.7%) and the traditional F-score strategy (12.7%). The results are in line with previous research which demonstrates how the capitalization of intangibles improves the returns of identified value firms (Eisfeldt et al., 2022; Lev and Srivastava, 2019). Furthermore, we find evidence that adjusting the B/M ratio for intangibles improves the F-score's ability to separate winners from losers. Specifically, the average annual return spread between high and low F-score portfolios is 5.8% with B^{INT}/M screening compared to 0.0% for B/M screening over the period May 2000 to April 2022. This may support studies that challenge the semi-strong form of the efficient market hypothesis, since the usage of publicly available information should not be able to predict future price patterns (Fama, 1970). The findings are also consistent with earlier studies on the F-score which demonstrate the importance of context for separating between different quality stocks (Mohanram, 2005; Piotroski, 2000; Piotroski and So, 2012). A few plausible explanations for improved returns are presented below.

The incorporation of stocks with robust business models

The first explanation suggests that the cohort of firms included in the High F-score^{INT} portfolio possess more successful firm characteristics compared to the ones in the High F-score^{REG} portfolio. Lev and Srivastava (2019) argue that there has been a slowdown in mean reversion post-financial crisis which has deteriorated the gains from investing in value stocks. With declined bank financing, it has been more difficult for these firms to perform necessary investments to sustain a competitive position. During this time, the authors identified significant intangible investments, strong financial health, and access to bank financing (to afford investments) as common traits for value firms that have improved their valuations (Lev and Srivastava, 2019).

The main differences between high B^{INT}/M stocks, as compared to high B/M stocks, are that the first will include firms in more intangible-intensive industries (due to within-industry sorting) and that these will have higher intangible-adjusted book values of equity due to SG&A capitalization. Firms with high B/M ratios are less likely to have significant investments in intangibles as these would understate their book equity levels (Francis and Schipper, 1999; Dontoh et al., 2004). This is supported by our data, as we see decreased SG&A intensity for high B/M firms and, unsurprisingly, increased for high B^{INT}/M firms (Appendix 1). Subsequently, the latter subset of firms should contain more firms with sufficient investments in intangibles to improve their valuations, while the opposite is true for high B/M firms. We argue that this could have two implications. First, much of the improved returns in the High F-$score^{INT}$ portfolio may be driven by the higher intangible-intensity among the underlying high B^{INT}/M group. Since the indexed return gap between high B^{INT}/M and high B/M stocks widened at an increasing rate after the financial crisis in 2009 and onward, this may be supported by our data (Figure 1). Second, if greater financing access to afford necessary investments is associated with higher returns (Lev and Srivastava, 2019), some F-score signals may give particularly misleading signals when identifying strong prospects among the high B/M group. Concretely, increased leverage and equity issuance are interpreted as negative signals by the F-score. This could explain why the return spread from high to low F-score firms decreases in the second subperiod when the book-to-market screening is not adjusted for intangibles (Table 2, Panel B).

Superior returns for a more correct proxy of value firms

Modern research suggests that prior high returns from investing in high book-to-market (value) stocks have decreased over time, which in part has been attributed to the misclassification of value and glamor stocks due to the immediate expensing of intangible investments (Eisfeldt et al., 2022; Lev and Srivastava, 2019). One possible explanation for improved returns of a high F-score strategy in our study is thus that value firms are better proxied for when the book-to-market screening is adjusted for intangible investments. Improved returns would then not be attributed to the importance of intangible investments, but an improved capture of distress risk or investor mispricing which drives returns of value stocks (Fama and French, 1993; Lakonishok et al., 1994). These returns would not be found among high B/M firms when book equity is not adjusted, since the B/M ratio would systematically misclassify more intangible-intensive firms as glamor stocks (Lev and Srivastava, 2019). This is supported by the higher SG&A intensity for low B/M high and high B^{INT}/M firms in our data (Appendix 1).

The improved identification of value firms could further explain why Piotroski's F-score is more efficient at separating winners from losers when the book-to-market screening is adjusted. Mohanram (2005) verified that Piotroski's F-score was best applied among value stocks and less suitable among glamor stocks. Since intangible investments have been on a steady rise since the 1980s, the misclassification issue should become more severe with time (Lev and Srivastava, 2019). This could explain why the F-score's ability to separate winning and losing stocks is not found when the book-to-market screening is unadjusted (Table 8) and is improved when the screening is adjusted for intangibles (Table 5, Panel B).

Difference in the capture of size effect between high F-score portfolios

Research has found that small stocks have a tendency to generate higher returns over time than large stocks (Banz, 1981; Reinganum, 1981). Hence, the lower market capitalization of firms in the High F-score^{INT} portfolio compared to the High F-score^{REG} portfolio could explain the higher returns of the former (Table 3). However, since the High F-score^{INT} has larger firms than the Low F-score^{INT} portfolio, size could not explain why the former generates higher returns than the latter. This may decrease the likelihood of size playing a significant role in our results.

5.2. Possible explanations for inconsistencies in return performance

While we find evidence that the ability of Piotroski's F-score to separate winners from losers is improved among high B^{INT}/M stocks, our robustness tests demonstrate that the ability is not significant in the early 2000s in isolation. Similarly, the strategy of investing in identified winners (high F-scores) suffers decreased returns in the 2010s and is subsequently unable to beat the S&P 500 index when solely considering this subperiod. This section will propose three explanations for these momentary declines in return performance. The first is directly related to the aptitude of the F-score, while the other two consider more general issues related to value.

Distortion of F-score inputs and lower accounting relevance

Srivastava (2014) discussed the declining value relevance of accounting information, which he argued was more pronounced for intangible-intensive firms. He argued that the immediate expensing of intangible investments leads to a decreased matching of revenues and costs, as well as increased volatility and depression of earnings figures (Srivastava, 2014). Given that a significant part of intangible expenses is intended to support future profits, these effects could deteriorate the economic interpretation of F-score signals, especially since many of these signals study changes in accounting ratios. Specifically, increased intangible investments between years will yield more negative signals as earnings are understated, and vice versa. Since the high B^{INT}/M group is more intangible-intensive than other book-to-market groups (Appendix 1), these issues would be more pronounced when Piotroski's value screen is adjusted. Moreover, although artificially adjusting accounting figures may seem like a solution to remedy the bias, it could be limited in its effectiveness for more intangible-intensive firms, as their value is largely determined by growth options not recognized in financial statements. Furthermore, such adjustments are likely to introduce other biases to the income statement (Barker et al., 2022; Balachandran and Mohanram, 2011). Hence, identifying value firms that should be more fit for fundamental investing by adjusting for intangibles may be paradoxical, since these adjustments may introduce other limitations to fundamental investing. The F-score, as it currently exists, may thus not be fully optimized to identify winners in this group.

The issue of identifying winners among winners

One potential reason why Piotroski's F-score may have a lower ability to identify winners in the early 2000s could relate to the already strong position of the underlying book-to-market group. First, value stocks as a group performed particularly well before the financial crisis from a historical perspective (Arnott et al., 2021; Lev and Srivastava, 2019). Second, value firms identified by the intangible-adjusted book-to-market metric have been found to have stronger fundamentals compared to value firms identified by the unadjusted book-to-market metric (Eisfeldt et al., 2022). While we demonstrate that comparatively fewer firms qualify for the high F-score portfolio in the high B^{INT}/M group in comparison to the high B/M group (Table 1), which could indicate weaker fundamentals for the first group, it is possible that their strong financial position is reflected in other ratios. This argument is further accentuated since we follow the methodology of Eisfeldt et al. (2022). As such, a lower return spread between expected winners and losers in this period may not be tied to inadequacies of the F-score, but simply the general difficulty of identifying deviant performers for the group in this period.

The slowdown of mean reversion

One apparent reason why the intangible-adjusted high F-score portfolio (High F-score^{INT}) suffered declined returns in the 2010s, and was unable to beat the S&P 500 index in this period, may simply be due to the general decline of value stocks. Researchers argue that adjustments for intangibles may improve returns of the identified value group, and that it is possible to identify financial characteristics that are common among the most successful value firms. Nonetheless, these firms would still belong to a group that has suffered disproportionately relative to other stocks after the financial crisis (Lev and Srivastava, 2019; Eisfeldt et al., 2022; Maloney and Moskowitz, 2020). As such, lower returns of financially superior high F-score firms would not be due to deficiencies in the F-score, but simply due to structural disadvantages which impact all value firms. This interpretation is supported by the fact that decreased returns for high B^{INT}/M firms during the 2011-2022 period coincided with an increase in the return spread between high and low F-score firms (Table 2). This suggests that the F-score's return predictive ability was improved, despite deteriorated high F-score returns.

5.3. Data and method discussion

All firm-level data in this study is gathered from The Center for Research in Security Prices (CRSP) and Compustat. These databases are widely used by finance professionals and academics, whose publications are regularly included in peer-reviewed financial journals. Hence, we argue that the data on which our study is based is reliable. For variable definitions and portfolio creation, we follow the methodology of Piotroski (2000) and Eisfeldt et al. (2022) in detail. Any deviations made are supported by later studies by these researchers (Piotroski and So, 2012). This includes raising the cutoff percentage in book-to-market sorting to the top 30% from 20% of firms and including firms with F-scores of 7-9 in long portfolios and firms with F-scores of 0-3 in short portfolios. Furthermore, sensitivity tests for strong assumptions in the perpetual inventory method increase the overall reliability of our results. We also evaluate the suitability of statistical tests used in this study. First, the normality assumption for the usage of parametric t-tests is reasonable due to our large sample sizes. Second, no issues with kurtosis or skewness have been detected in portfolio data when compared to cutoff values provided by Hair et al. (2010) and Byrne (2010). Finally, it is possible that the factors which are driving the results in our second hypothesis are related to an underlying change in the return distribution when adjusting for intangibles, rather than any improvement of the F-score measure. However, both portfolios are normally distributed and have similar standard deviations, and as such, any difference in return spreads is more likely attributable to a shift than a change in the shape of the distribution. Furthermore, similar methodologies of comparing long-short returns across book-to-market groups to test the F-score's return predictive ability have been used in prior studies, such as Mohanram (2005) and Piotroski and So (2012).

6. Conclusion

Before this paper, one part of investment literature has focused on how adjusting the book-to-market ratio through capitalization of intangible investments to book figures may improve the proxy for and returns of value stocks. Another part has focused on the declining performance of investment strategies, which are based on an imperfect identification of value stocks. However, how these findings may be combined has been largely neglected. This study attempts to bridge that gap by evaluating the performance of Piotroski's F-score after adjusting its book-to-market screening for intangibles, in an effort to improve the strategy's functionality.

We demonstrate that intangible adjustments to Piotroski's methodology can significantly increase the returns of a portfolio investing in high F-score firms. This adjusted investment strategy also beats S&P 500 index during the 21st century, although returns deteriorate in the 2010s. This coincides with prior literature on the recent downturn of the value effect (Lev and Srivastava, 2019; Maloney and Moskowitz, 2020).

We also find evidence that the ability of Piotroski's F-score to identify winning and losing stocks is improved when the book-to-market screening is adjusted for intangibles. Further, a portfolio consisting of identified winners (F-scores 7-9) outperforms a simpler strategy of investing in the complete, underlying, adjusted book-to-market group. When the book-to-market value screen is not adjusted, neither of these abilities is present for the F-score. Specifically, a portfolio of expected winners (F-scores 7-9) generates lower mean returns than one of expected losers (F-scores 0-3), and a portfolio of expected winners is unable to beat a simple high book-to-market strategy. This may support the importance of context for Piotroski's F-score which is brought up in prior studies, as it was originally designed to be applied to value stocks (Mohanram, 2005; Piotroski, 2000). The results also align with previous research which has shown a decline in the effectiveness of an unadjusted F-score strategy in the U.S. (Woodley et al., 2011; Piotroski and So, 2012).

If the results of our study hold true, there are significant implications for both investors and academics. First, the success of the adjusted F-score strategy exemplifies how fundamental analysis of accounting information can generate significant returns, indicating that value relevance of such information is still sufficiently high to be relevant to investors. While the implementation of such strategies is made more cumbersome with amortization schedules for internally generated intangibles, they are still possible to implement for small-stake investors. Second, the functioning of fundamental analysis is contextual as the F-score metric is better able to identify winning stocks among a high book-to-market group when the ratio's book figure is adjusted for intangible investments.

7. Future research

Due to inherent limitations in our study, we acknowledge several interesting possibilities for future research. First, one could use a different method for capitalizing internally generated intangibles to the book-to-market ratio. Other methods capture relevant expenses differently, splitting into organization and knowledge capital, and use industry-specific investment and depreciation rates for the intangible capital stock. Some methods also use pre-IPO growth rates to estimate the initial capital stock for each firm and do not perform within-industry sorting for book-to-market terciles. If such methods find similar results as ours, it would strengthen the case that capitalizing internally generated intangibles increases the applicability of Piotroski's F-score. Second, one could investigate the impact of adjusting inputs to individual F-score signals for internally generated intangibles. If resulting portfolio strategies find significant excess returns, it could generate an interesting discussion on the value relevance of reported accounting figures while also possibly having implications for the aptitude of current accounting standards. Third, a related study could investigate the impact of capitalizing internally generated intangibles to accounting figures for the returns of other value investing strategies. This would broaden the discussion on investor benefits from adjusting for intangibles since our study is naturally delimited to Piotroski's F-score. Finally, since we make no attempt to discuss the risk-adjusted returns of our portfolios, it might be interesting to replicate our study and see whether abnormal returns are achieved. In such a study, one could use an asset pricing model which incorporates the HML factor, and test for risk-adjusted returns when this factor both is and is not adjusted for the expensing of internally generated intangibles.

Bibliography

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. Journal of Financial Economics, 17(2), 223-249.

Arnott, R. D., Harvey, C. R., Kalesnik, V., & Linnainmaa, J. T. (2021). Reports of value's death may be greatly exaggerated. Financial Analysts Journal, 77(1), 44-67.

Balachandran, S., & Mohanram, P. (2011). Is the decline in the value relevance of accounting driven by increased conservatism? Review of Accounting Studies, 16(2), 272-301.

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. Journal of accounting research, 6(2), 159-178.

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

Barker, R., Lennard, A., Penman, S., & Teixeira, A. (2022). Accounting for intangible assets: Suggested solutions. Accounting and Business Research, 52(6), 601-630.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. The journal of Finance, 32(3), 663-682.

Berk, J., & DeMarzo, P. (2019). Corporate finance (5th ed.). Harlow, England: Pearson Education Limited.

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? Journal of Accounting Research, 27(1), 1-36.

Burton, M. (1973). A random walk down wall street (1st ed.). New York, United States: WW Norton.

Byrne, B. M. (2010). Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming (2nd ed.). New York: Taylor and Francis Group.

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, Covariances, and Average Returns: 1929 to 1997. The Journal of Finance, 55(1), 389-406.

De Bondt, W. F., & Thaler, R. H. (1989). Anomalies: A mean-reverting walk down Wall Street. Journal of economic perspectives, 3(1), 189-202.

Dontoh, A., Radhakrishnan, S., & Ronen, J. (2004). The declining value-relevance of accounting information and non-information-based trading: an empirical analysis. Contemporary Accounting Research, 21(4), 795-812.

Eisfeldt, A. L., Kim, E. T., & Papanikolaou, D. (2022). Intangible value. Critical finance review, 11(2), 299-332.

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

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383-417.

French, K. (2023). Kenneth R. French – Data Library. Retrieved April 5, 2023, from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Francis, J., & Schipper, K. (1999). Have financial statements lost their relevance? Journal of Accounting Research, 37(2), 319-352.

Graham, B. (2003). The intelligent investor (4th rev. ed.). Harper Business.

Hair, J. F., Black, W. C., & Babin, B. J. (2010). RE Anderson Multivariate data analysis: A global perspective.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. Journal of financial economics, 6(2/3), 95-101.

Jones, S. T., Reburn, J. P., & Woodley, M. K. (2011). Value stocks and accounting screens: Has a good rule gone bad? Journal of Accounting and Finance, 11(4), 35-42.

Kothari, S. P., & Warner, J. B. (1997). Measuring long-horizon security price performance. Journal of Financial Economics, 43(3), 301-339.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. The journal of finance, 49(5), 1541-1578.

Lev, B., & Srivastava, A. (2019). Explaining the recent failure of value investing. NYU Stern School of Business. Rochester, NY.

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

Maloney, T., & Moskowitz, T. J. (2020). Value and interest rates: Are rates to blame for value's torments? The Journal of Portfolio Management, 47(6), 65-87.

Mohanram, P. (2005). Separating winners from losers among low book-to-market stocks using financial statement analysis. Review of Accounting Studies, 10(2-3), 133-170.

Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. Journal of Accounting and Economics, 11(4), 295-329.

Peters, R. H., & Taylor, L. A. (2017). Intangible capital and the investment-q relation. Journal of Financial Economics, 123(2), 251-272.

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

Piotroski, J. D., & So, E. C. (2012). Identifying expectation errors in value/glamour strategies: A fundamental analysis approach. The Review of Financial Studies, 25(9), 2841-2875.

Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. Journal of financial economics, 22(1), 27-59.

Reinganum, M. R. (1981). Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. Journal of financial Economics, 9(1), 19-46.

Roll, R. (1981). A Possible Explanation of the Small Firm Effect. The Journal of Finance, 36(4), 879-888.

Rosenberg. B, Reid, K., & Lanstein, R. (1998). Persuasive evidence of market inefficiency. Streetwise: The Best of the Journal of Portfolio Management, 48.

Samuelson, P. A. (1973). Proof that properly discounted present values of assets vibrate randomly. The Bell Journal of Economics and Management Science, 4(2), 369–374.

Sciencedirect. (n.d.). Wilcoxon signed-rank test. In Encyclopedia of Mathematics (2nd ed.). Retrieved April 5, 2023, from https://www.sciencedirect.com/topics/mathematics/wilcoxon-signed-rank-test

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? The Accounting Review, 71 (3), 289–315.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), 425-442.

Shumway, T. (1997). The delisting bias in CRSP data. The Journal of Finance, 52(1), 327-340.

Srivastava, A. (2014). Why have measures of earnings quality changed over time? Journal of Accounting and Economics, 57(2-3), 196-217.

S&P Dow Jones Indices. (n.d.). S&P 500. Retrieved April 5, 2023, from https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview

Walkshäusl, C. (2020). Piotroski's FSCORE: international evidence. Journal of Asset Management, 21(2), 106-118

Woodley, M. K., Jones, S. T., & Reburn, J. P. (2011). Value stocks and accounting screens: has a good rule gone bad?. Journal of Accounting and Finance, 11(4), 87-104.

Zhang, X. F. (2006), 'Information uncertainty and stock returns', Journal of Finance 61(1), 105–136.

Appendices

A.1. SG&A intensity for different book-to-market sorts

		SG&A Intensity				
Book-to-market group	N	Mean	s.d.	P25	p50	p75
High B/M	11774	0.29	2.03	0.10	0.20	0.34
Middle B/M	15494	0.33	0.67	0.13	0.25	0.43
Low B/M	10822	0.45	0.32	0.22	0.40	0.65
High B^{INT}/M	12255	0.40	2.10	0.17	0.30	0.50
Middle B^{INT}/M	15623	0.34	0.36	0.14	0.26	0.47
Low B^{INT}/M	10212	0.30	0.32	0.11	0.22	0.41

Notes: This table summarizes the SG&A intensity for different book-to-market groups, both when book equity is and is not intangible-adjusted. SG&A intensity is defined as SG&A expenses scaled by total expenses, in line with Srivastava (2014). We use the following nomenclature for describing book-to-market groups: High = top 30%, Middle = middle 40%, and Low = bottom 30%. N is the total number of firm-year observations for each book-to-market group. Due to overlapping firm-year observations between unadjusted and intangible-adjusted book-to-market groups, the total number of observations exceeds the total number of firm-year observations in our sample. We report statistics using annual data at December$_{t-1}$ 5 months before portfolio formation (May each year) from 1999 to 2021.

A.2. Book-to-market ratios for Fama French 12 industries

		B/M					B^{INT}/M				
Industry	N	Mean	s.d.	p10	p50	p90	Mean	s.d	p10	p50	p90
Consumer Nondurables	2052	0.75	0.88	0.15	0.50	1.45	2.60	4.46	0.41	1.39	5.37
Consumer Durables	1068	0.80	1.09	0.16	0.56	1.40	2.67	4.91	0.39	1.36	5.60
Manufacturing	4843	0.78	0.99	0.20	0.56	1.46	1.90	2.72	0.40	1.13	3.98
Chemicals	1178	0.73	0.98	0.17	0.49	1.39	1.80	2.51	0.44	1.03	4.01
Business Equipment	8390	0.56	0.58	0.12	0.40	1.12	1.96	3.10	0.27	1.07	4.34
Telecommunications	1462	0.85	1.19	0.15	0.59	1.62	1.68	2.71	0.18	0.88	3.87
Oil, Gas, and Coal Extraction	2455	0.89	0.97	0.26	0.65	1.70	1.32	2.14	0.37	0.83	2.56
Wholesale and Retail	3069	0.83	1.04	0.17	0.55	1.60	3.73	8.78	0.43	1.74	8.25
Healthcare	7827	0.44	0.60	0.09	0.29	0.92	1.30	2.79	0.15	0.60	2.78
Other	5749	0.79	1.27	0.16	0.55	1.51	1.89	4.81	0.24	0.95	3.92

Notes: This table summarizes the traditional and intangible-adjusted book-to-market values for Fama and French's 12-industry classification. N is the total number of firm-year observations for each industry. B/M is the traditional book-to-market ratio and B^{INT}/M denotes the intangible-adjusted book-to-market ratio. We report statistics using annual data at December$_{t-1}$ 5 months before portfolio formation (May each year) from 1999 to 2021. The table shows how the impact of incorporating intangibles into book-to-market ratios varies across industries.

A.3. Varying investment and depreciation rates in the computation of B^{INT}/M ratios

PANEL A: Investment rates	0.5	0.6	0.7	0.8	0.9
High FS^{INT} – High FS^{REG}	3.866*** (3.882)***	3.993** (4.125)***	3.876*** (4.105)***	3.915*** (4.191)***	3.543*** (3.812)***
$High^{INT}$ – Low^{INT}	1.724** (3.158)***	1.726** (3.137)***	1.769** (3.203)***	1.752** (3.162)***	1.724** (3.265)***
$(High – Low)^{INT}$ – $(High – Low)^{REG}$	3.670*** (4.138)***	3.872*** (4.313)***	3.716*** (4.393)***	3.801*** (4.427)***	3.673*** (4.069)***
High FS^{INT} – S&P 500	3.855*** (3.956)***	3.852*** (3.890)***	3.869*** (3.860)***	3.839*** (3.818)***	3.731*** (3.686)***
PANEL B: Depreciation rates	**δ=0.1**	**δ=0.2**	**δ=0.3**	**δ=0.4**	**δ=0.5**
High FS^{INT} – High FS^{REG}	3.490*** (3.509)***	3.592*** (3.972)***	4.051*** (4.266)***	3.733*** (4.067)***	3.680*** (3.947)***
High FS^{INT} – Low FS^{INT}	1.741** (3.044)***	1.837** (3.314)***	1.948** (3.341)***	1.855** (3.283)***	1.740** (3.098)***
$(High – Low)^{INT}$ – $(High – Low)^{REG}$	3.390*** (3.810)***	3.669*** (4.285)***	4.211*** (4.882)***	3.949*** (4.698)***	3.646*** (4.396)***
High FS^{INT} – S&P 500	3.806*** (3.673)***	3.846*** (3.814)***	3.853*** (3.877)***	3.773*** (3.778)***	3.704*** (3.766)***

Notes: This table summarizes the significance levels of the 12-month buy-and-hold return differences for our three main hypotheses when varying the investment rate between 0.5-0.9 and the depreciation rate between 0.1-0.5 in the computation of B^{INT}/M ratios. *FS* is used as an abbreviation for *F-score*. Significance levels are indicated by: *p<0.1, **p< 0.05, ***p<0.01.

A.4. Testing against F-score portfolio using within-industry sorting for B/M screening

Monthly returns	Mean	Std. Err.	Median	Obs
High FS^{INT}	0.0172	0.0045	0.0198	264
High FS^{IND}	0.0126	0.0044	0.0133	264
High FS^{INT}– High FS^{IND}	0.0046	0.0012	0.0065	264
T-statistic (Z-value)	3.990***	–	(4.358)***	–
P-value	0.0000	–	0.0000	–

Notes: This table presents the results for the paired sample t-test (T-statistic) and the Wilcoxon signed rank test (Z-value) on monthly returns over the full period for the different F-score portfolios and the S&P 500. The High F-$score^{IND}$ portfolio consists of firms with an aggregate F-score of 7-9 above the 70^{th} percentile of B/M firms within each industry (according to the Fama French 12 industry classification). Significance levels are indicated by: *p<0.1; **p< 0.05; ***p<0.01.

A.5. Univariate analysis for the 2000-2011 and 2011-2022 periods

Panel A: High FS^{INT} tested against High FS^{REG}

	2000-2011				2011-2022			
Monthly returns	**Mean**	**Std. Err.**	**Median**	**Obs**	**Mean**	**Std. Err.**	**Median**	**Obs**
High FS^{INT}	0.0205	0.0061	0.0255	132	0.0140	0.0067	0.0148	132
High FS^{REG}	0.0158	0.0060	0.0231	132	0.0095	0.0067	0.0107	132
High FS^{INT} – High FS^{REG}	0.0046	0.0014	0.0024	132	0.0044	0.0021	0.0041	132
T-statistic (Z-value)	3.364***		(3.337)***		2.089**		(2.328)***	

Panel B: High FS^{INT} – Low FS^{INT} tested against High FS^{REG} – Low FS^{REG}

	2000-2011				2011-2022			
Monthly returns	**Mean**	**Std. Err.**	**Median**	**Obs**	**Mean**	**Std. Err.**	**Median**	**Obs**
High-Low FS^{INT}	0.0041	0.0040	0.0100	132	0.0057	0.0036	0.0091	132
High-Low FS^{REG}	0.0008	0.0036	0.0057	132	-0.0012	0.0030	0.0041	132
High-Low FS^{INT} – High-Low FS^{REG}	0.0033	0.0015	0.0043	132	0.0069	0.0023	0.0051	132
T-statistic (Z-value)	2.163**		(2.351)***		2.974***		(3.629)***	

Panel C: High FS^{INT} tested against Low FS^{INT}

	2000-2011				2011-2022			
Monthly returns	**Mean**	**Std. Err.**	**Median**	**Obs**	**Mean**	**Std. Err.**	**Median**	**Obs**
High FS^{INT}	0.0205	0.0061	0.0255	132	0.0140	0.0067	0.0148	132
Low FS^{INT}	0.0164	0.0081	0.0155	132	0.0083	0.0069	0.0056	132
High FS^{INT} – Low FS^{INT}	0.0041	0.0040	0.0100	132	0.0057	0.0036	0.0091	132
T-statistic (Z-value)	1.030		(2.408)***		1.601*		(2.319)**	

Panel D: High FS^{INT} tested against the S&P 500 index

	2000-2011				2011-2022			
Monthly returns	**Mean**	**Std. Err.**	**Median**	**Obs**	**Mean**	**Std. Err.**	**Median**	**Obs**
High FS^{INT}	0.0205	0.0061	0.0255	132	0.0140	0.0067	0.0148	132
S&P 500	0.0034	0.0043	0.0115	132	0.0095	0.0036	0.0121	132
High FS^{INT} – S&P 500	0.0171	0.0034	0.0140	132	0.0044	0.0043	0.0027	132
T-statistic (Z-value)	4.956***		4.776***		1.017		0.568	

Notes: This table presents the results of paired sample t-test (T-statistic) and the Wilcoxon signed rank test (Z-value) on monthly returns for the 11-year subperiods (2000-2011 and 2011-2022) for the testing of our three main hypotheses. *FS* is used as an abbreviation for *F-score*. Significance levels are indicated by: *p<0.1, **p< 0.05, ***p<0.01.