Intangibles Investment and Asset Quality*

Ravi Jagannathan†
Robert A. Korajczyk‡
Kai Wang§

May 28, 2024

Abstract

A factor using an earnings measure treating intangible and physical investments symmetrically represents "quality." It has smaller left tail risk and co-tail risk with the market than does RMW of Fama and French (2015) and has lower down-market than up-market exposure. Our factor has significant alpha relative to many extant multi-factor asset-pricing models, including the Fama-French model ($\alpha = 2.9\%$). Its performance is due to superior asset selection (market timing) on the long (short) side. When the profitability factor in the Fama-French model is replaced with our factor the resulting model performs better in explaining both the cross section of stock returns and several extant anomalies.

Keywords: Profitability Prediction, Intangible Assets, Return Factors, Asset Pricing

JEL Classification: G11, G12

*We would like to thank Ray Ball, Andrea Eisfeldt, Huseyin Gulen, Matthias Hanauer, Josef Lakonishok, Soohun Kim, Tatjana Xenia Puhan, Anthony Sanford, Michael Weber, and seminar participants at Northwestern University, Stevens Institute of Technology, Vienna Graduate School of Finance, the Advances in Financial Econometrics conference at Copenhagen Business School, Frontiers in Factor Investing Conference, HEC-McGill Winter Finance Conference, UIC Finance Conference, and Alliance Bernstein Quant Conference for helpful comments. Part of this research was done while Korajczyk was an Engelbert Dockner Fellow at the Vienna University of Economics and Business.

†Kellogg School of Management, Northwestern University, and NBER, Email: rjaganna@kellogg.northwestern.edu; Tel: 1-847-491-8338.

‡Kellogg School of Management, Northwestern University, Email: r-korajczyk@kellogg.northwestern.edu; Tel: 1-847-491-8336.

§Kellogg School of Management, Northwestern University, Email: kai.wang2@kellogg.northwestern.edu; Tel: 1-551-226-3715.

1 Introduction

Linear factor models are widely used in the asset pricing literature to represent economy-wide pervasive risk.[1] Connor (1995) categorizes factors models into three types, statistical, macroeconomic, and fundamental, based on the method of choosing the asset pricing factors (as well as hybrid versions that combine statistical, macroeconomic, or fundamental factors (Stroyny (2005))). The logic of the factor model approach is that priced risks should be common, pervasive risks across assets, in the sense that they cannot be diversified away in portfolios with many assets. Statistical factor models (starting with Roll and Ross (1980), Connor and Korajczyk (1986, 1988)) exploit the covariance structure of asset returns to estimate the important common factors. The advantage of statistical models is that they extract factors based on systematic co-movement. The disadvantages are that the factors are not explicitly linked to economic aggregates and that they have a rotational indeterminacy that makes them non-comparable across periods, although they can be rotated to match macroeconomic or fundamental factors (Connor and Korajczyk (1991)).

Fundamental financial factor models specify factor loadings (betas) of stock returns as functions of observable characteristics and estimate latent factors from the relation between asset returns and betas. Examples are, Rosenberg and McKibben (1973), Rosenberg and Guy (1976), Connor et al. (2012), Kelly et al. (2019), Kim et al. (2021), and Ge et al. (2022). These fundamental factor models are economically interpretable to the extent that the characteristics explaining factor loadings are interpretable. This approach has been successful in addressing time variation in the elements of the covariance matrices of stock returns.

Macroeconomic factor models specify, in advance, the nature of the systematic factors. They have the advantage that the factors are interpretable and, ideally, should be correlated with macroeconomic variables that are linked to the marginal utility of investors. Examples are Chen et al. (1986), Breeden et al. (1989), Cochrane (1991), Braun (1993), Jagannathan and Wang (1996), Jagannathan and Wang (2007), Papanikolaou (2011), Kogan and Papanikolaou (2014), among others. While they have had some success in explaining time variation in stock returns at business cycle frequencies, they have several well-recognized disadvantages: macroeconomic time series often have a low correlation

[1] Linear factor models take the market risk premium and the risk-free rate as given when explaining the cross section of returns on securities. In contrast, consumption-based asset pricing models focus on explaining the market risk premium and the risk free rate and their dynamics (see Hansen and Singleton (1982, 1983), Bansal and Yaron (2004), and Parker and Julliard (2005) among others).

with asset returns, possibly due to measurement error, temporal aggregation, aggregation across heterogeneous investors, infrequent decision making due to set up costs (Jagannathan and Wang (2007), Abel et al. (2013)), model misspecification, or asset returns anticipating future changes in macroeconomic factors.

Another approach in the literature is to first identify cross-sectional variation in expected asset returns based on introspection and theory. Then, in a second step, construct zero-investment portfolios that are long stocks with higher expected returns and short stocks with lower expected returns. To the extent the market is informationally efficient, the returns on such long-short portfolios will capture economy-wide pervasive risks. Otherwise, they capture mispricing.

Fama and French (1993) and Fama and French (2015) use this approach to construct three-factor (FF3) and five-factor (FF5) models, respectively. They use the present value relation between asset valuation and future payouts (the dividend discount model) to identify characteristics plausibly related to cross-sectional variation in expected returns. The FF5 model augments their original three-factor model in Fama and French (1993) (FF3) with factors based on profitability (robust-minus-weak (RMW)) and asset growth (aggressive-minus-conservative (CMA)).

The very name of the RMW factor suggests that it is a metric for firm quality: robust firms are of higher quality than weak firms. They perform better in downturns than weak firms.[2] In fact, a wide array of earnings/revenue-based factors have negative market betas and can be thought of as quality metrics (e.g., Novy-Marx (2013), Ball et al. (2015), and Gulen et al. (2023)). There are other quality factors that incorporate profitability as well as other metrics Asness et al. (2019) and other non-factor measures of quality (Piotroski (2000), Monhanram (2005)).

We propose a modification of the FF5 profitability factor which takes into account the fact that profitability based on U.S. generally accepted accounting principles (GAAP) includes currently expensed costs that should rightly be considered investments in intangible assets (such as research and development (R&D) and portions of sales general and Administrative (SG&A) expenses). That is, GAAP profitability reflects investments in physical capital by capitalizing and depreciating those expenditures over time. However, GAAP immediately expenses R&D and SG&A expenditures. Investments in intellectual capital, brand, and organizational capital are, therefore, treated very differently from investments in physical capital, under GAAP.

[2]Jagannathan and Zhang (2020) propose a method for separating higher quality stocks from lower quality stocks based on performance during the month with the worst market return in a calendar year.

We find that the intangible-adjusted profitability factor has positive annualized alpha of 3% (7%) against the both original FF5 profitability factor alone and the FF5 model when our factor portfolios are rebalanced annually (monthly). Substituting the original FF profitability with the intangibles-adjusted profitability factor leads to a model that substantially reduces some of the well-known asset pricing anomalies, such as momentum (Jegadeesh and Titman (1993)), gross profitability (Novy-Marx (2013)), and monthly-rebalanced ROE (Hou et al. (2015)).

The performance of our factor comes from both the long and short sides of the factor portfolio, but the nature of the alphas are quite different. The alpha long side of the strategy comes from positive security alphas while the alpha on the short side comes from market timing, that is from shorting stocks that do particularly poorly in down markets.

2 Characteristics Sorted Portfolios As Factors

Fama and French (2006, 2015) use the dividend discount model to relate a stock's firm characteristics like market-to-book equity ratio, profitability, asset growth on the one hand and its (long-run) expected return - also referred to as the internal rate of return, discount rate, stock yield (Jagannathan et al. (2001)), and implied cost of capital (Gebhardt et al. (2001)) – on the other hand. They then sort stocks based on firm characteristics and construct long-short portfolios of stocks which they use as factors in their linear factor-beta pricing model. In this section, we argue that characteristics sorted portfolios are better proxies for unobserved latent economy-wide pervasive factors than factors estimated using statistical methods when factor betas change over time.

2.1 The Present Value Relation

The present value model equates the market price of a firm's share to the present value of the expected stream of future dividends per share:

$$P_t = \sum_{\tau=1}^{\infty} \frac{\mathbb{E}(D_{t+\tau})}{(1+r)^{\tau}}, \tag{1}$$

where P_t is the share price, $\mathbb{E}(D_{t+\tau})$ is the expected dividend at time $t+\tau$, and r is the ICC. The dividend at any future time $t+\tau$ can be written as the earnings minus retained earnings:

$$D_{t+\tau} = Y_{t+\tau} - (Y_{t+\tau} - D_{t+\tau})$$

where the $Y_{t+\tau}$ is the earnings. When $Y_{t+\tau}$ is after-tax earnings, "clean surplus accounting" implies that the net change of book equity is equal to the retained earnings:

$$dB_{t+\tau} = Y_{t+\tau} - D_{t+\tau} \tag{2}$$

where $dB_{t+\tau} \equiv B_{t+\tau} - B_{t+\tau-1}$ is the net change of book equity in year $t + \tau$.

Replacing $D_{t+\tau}$ with $Y_{t+\tau} - dB_{t+\tau}$ and dividing both sides of (1) by the current book equity B_t, the dividend discount model becomes

$$\frac{P_t}{B_t} = \sum_{\tau=1}^{\infty} \mathbb{E}\left(\frac{Y_{t+\tau}}{B_t} - \frac{dB_{t+\tau}}{B_t}\right) \Big/ (1+r)^{\tau}. \tag{3}$$

There are several variables in the above equation in addition to expected return: the market price, P_t, the book value, B_t, (and the market-to-book ratio P_t/B_t), the stream of expected profitability $\mathbb{E}(Y_{t+\tau})/B_t$, and the stream of expected equity growths $\mathbb{E}(dB_{t+\tau})/B_t$, where $\tau = 1, ..., \infty$. While the present value relation is a mathematical identity, it takes on economic content when the expectations in equation (3) represent the expectations of the investor determining the market price of the share at the margin.

Fama and French (2015) observe that according to the present value relation, ceteris paribus, firms with 1) higher market-to-book ratios should have higher expected returns; 2) higher expected profitability should have higher expected returns; 3) higher expected growth in book value should have lower expected returns; and 4) higher market value (holding book value constant) should have lower expected returns. Fama and French (2015) use this observation to motivate their use of the returns on characteristics sorted portfolios as proxies for economy-wide pervasive factors.

We provide additional support for the use of characteristics sorted long-short portfolio returns as proxies for economy-wide pervasive factors by considering an example economy where the expected growth rate and one period expected returns remain constant over time, but differ across firms — i.e., the Gordon Growth Model holds. In addition, returns are generated by four latent factors, and the expected excess return on a stock is a linear function of its return factor betas, i.e., a linear four beta model explains the cross section of expected excess returns. In such an economy, we show that factors constructed by sorting stocks based on firm characteristics -– book/market, ROE, and asset growth – along with the equally weighted market index factor proxy well for the four unobserved factors. Factors constructed using the Connor and Korajczyk (1986) asymptotic principal component method are better proxies when betas do not vary over time. However, long-short characteristics sorted portfolios are better proxies for the latent factors when factor betas vary over time as in Kim et al. (2021). The internet appendix to the paper provides the details.

Any set of variables that help explain the cross section of expected profitability and expected asset growth would be candidates for constructing characteristics sorted portfolios.

These include lagged values of profitability and asset growth, which form the basis of some FF5 factors. The factors in FF5 model are the excess return on the market portfolio (clearly motivated by the Capital Asset Pricing Model (CAPM)) and four long-short portfolios, all of which can be motivated by the present value relation, (3). SMB (Small Minus Big) is a portfolio that is long small capitalization stocks and short large capitalization stocks. HML (High Minus Low) is a portfolio that is long high book-to-market stocks and short low book-to-market stocks. RMW (Robust Minus Weak) is a portfolio that is long high expected profitability stocks and short low expected profitability stocks — often referred to as the profitability factor. Finally, CMA (Conservative Minus Aggressive) is a portfolio that is long low expected asset growth stocks and short high expected asset growth stocks. Holding, B_t, $\mathbb{E}(Y_{t+\tau})/B_t$, and $\mathbb{E}(dB_{t+\tau})/B_t$, the present value relation suggests that SMB should have a positive expected return, as argued by Berk (1995). By similar logic, HML, RMW, and CMA should have positive average returns.

2.2 Profitability and Future Cash Flows

Our focus is on the profitability factor. Fama and French (2006) use net income divided by the book equity (which we call bottom-line earnings hereafter) as a proxy for expected profitability. They find a significant relation between expected returns and current profitability. However, they find no significant relation between expected returns and expected profitability measured by fitted values from cross-sectional regression of future profitability on current profitability and additional variables. They offer two possible explanations for the puzzling results: (1) the measurement error in the first-stage cross-section profitability regressions and (2) the collinearity problem since the book-to-market ratio enters both the first-stage and the second-stage regressions.

Measurement error in profitability has been discussed extensively in the literature. Several alternative measures of profitability have been suggested as alternatives to bottom-line earnings since they may be less susceptible to manipulation by management. Using Compustat definitions, we move from the top-line of the income statement (revenue) to bottom-line earnings, as follows:[3]

$$
\begin{aligned}
&\text{Revenue (REVT)}\\
-&\text{Cost of goods sold (COGS)} = \text{Gross profit (GP)}\\
-&\text{Selling, general and administrative expenses (XSGA)} = \text{Operating income before depreciation (OIBDP)}\\
-&\text{Depreciation and amortization (DP)} = \text{Operating income after depreciation (OIADP)}\\
-&\text{Interest and related expense (XINT)}\\
+&\text{Special items (SPI)}\\
+&\text{Non-operating income (NOPI)} = \text{Pretax income (PI)}\\
-&\text{Income taxes (TXT)}\\
-&\text{Minority interest (MII)} = \text{Income before extraordinary items (IB)}
\end{aligned}
$$

To avoid the largely non-recurring items, Li and Mohanram (2014) measure earnings as income before extraordinary items minus special items, IB-SPI. Novy-Marx (2013) uses gross profit (GP) as the proxy for profitability and argues that *"The farther down the income statement one goes, the more polluted profitability measures become, and the less related they are to true economic profitability."* To measure earnings in their profitability factor, Fama and French (2015) subtract the selling, general, and administrative expenses and interest expenses from the gross profit (EBTDA = GP-XSGA-XINT).

We assess the ability of various earnings measures to predict future cash flows (years $t+1$ through $t+10$) by regressing future cash flows (measured by the Compustat Cash Flow from Operations, OANCF) on current earnings metrics (Table 1). The cross-sectional regressions are for 50 portfolios sorted by 5 industry groups and the Fama-French measure of profitability. Of the measures that do not adjust for intangibles, we use (moving from bottom-line to top-line measures) net income (Fama and French (2006)), earnings before extraordinary items (Li and Mohanram (2014)), earnings before taxes (EBT), earnings before taxes, depreciation, and amortization (EBTDA, Fama and French (2015)), gross profit, (Novy-Marx (2013)), and revenue. Since the Compustat variable for cash flows from operations (OANCF) is available starting in 1988, our predictive regressions are for the period 1988-2022. For all horizons, the average R^2 values of the regression increase from net income through the Fama and French (2015) EBTDA measure and then decline as we move to more top-line measures. The results for a subset of these measures, net income; earnings before taxes, depreciation, and amortization (EBTDA); and gross profit, are reported in Table 1. Across all horizons, the gross profit measure of profitability has higher average values of R^2 than net income. Additionally, the FF5

[3] All the items have quarterly counterparts.

measure of profitability has higher values of R^2 than gross profitability Because of this result, we follow Fama and French (2015) and measure unadjusted profitability using EBTDA instead of the more bottom-line measure used in Fama and French (2006) or the pore top-line measure of gross profitability in Novy-Marx (2013).

2.3 Earnings, Value, and Growth Factors with Investments in Intangibles

Analysts have long recognized that the accounting treatment of investments in physical assets and intangible assets has been asymmetric (for example, Bloomberg (1934), Graham et al. (1962), Graham (1973), Greenwald et al. (2001)).[4] Investments in physical capital are capitalized and depreciated over time while investments in intangible capital, which build intellectual property, brands, and organizational capital (reported in R&D and SG&A), are expensed immediately, unless they are purchased intangibles. Treating investments in intangible capital similarly to physical capital has implications for the measurement of earnings and book value and influences three of the variables in the present value relation (3), P_t/B_t, $dB_{t+\tau}/B_t$, and $Y_{t+\tau}/B_t$.

Similarly, incorporating investment in intangible capital might affect the way we define three of the factors in the FF5 model, RMW, HML, and CMA. Arnott et al. (2021) and Eisfeldt et al. (2022) adjust the HML for the intangible capital and argue that the intangible-adjusted HML has significant advantages, in terms of pricing the cross-section, relative to the original HML. Wang (2023) proposes an intangibles-adjusted asset growth factor, similar to CMA, and argues it outperforms the original CMA factor. Gulen et al. (2023) construct variants of the FF5 factors that incorporate the effects of off-balance-sheet intangible assets and find that some outperform the FF5 factors.

Fundamental analysts often adjust earnings for investment in intangibles, e.g., Greenwald et al. (2001) add 25% of both the R&D and SG&A expenses back to Intel's earnings to project its future earnings (Table 7.7 in the chapter "Inside Intel" in Greenwald et al.

[4]In the 2010 CEO's letter to Amazon shareholders, Jeffrey Bezos emphases that Amazon's continuing investment in intangible assets has played a determinant role in advancing the company and the ultimate payoff to shareholders: *"The advances in data management developed by Amazon engineers have been the starting point for the architectures underneath the cloud storage and data management services offered by Amazon Web Services (AWS). ... in my opinion, these techniques are not idly pursued – they lead directly to free cash flow."*

In the 1997 annual report Bezos illustrates his principle of investing: *"We will make bold rather than timid investment decisions where we see a sufficient probability of gaining market leadership advantages. ... When forced to choose between optimizing the appearance of our GAAP accounting and maximizing the present value of future cash flows, we'll take the cash flows."*

(2001)). Even earlier, Graham (1973) highlights the danger of earnings "smoothing" and the choice between fully charging off R&D costs or amortizing them.[5]

The academic literature has advocated several approaches to account for investment in intangibles when explaining equity valuation. For example, Lev and Sougiannis (1996) capitalize R&D expense and adjust bottom-line earnings for R&D expense. Peters and Taylor (2017) and Crouzet and Eberly (2023) consider all R&D expense as investment in knowledge capital and 30% of the SG&A expense, net of R&D expense, as the investment in organizational capital.[6] Belo et al. (2022) treat R&D and advertising expenses and the change in the number of employees as separate components of investment in intangibles. Eisfeldt and Papanikolaou (2013) capitalize the sum of R&D and SG&A expenses, again SG&A is net of R&D expense, as measure of firms' organizational capital. Eisfeldt et al. (2022) suggest: *"Without better estimates of the fraction of SG&A spending that is investment in intangible assets, ... it is best to use 100% of SG&A ... to avoid introducing noise."*

Because of its superior ability to forecast future cash flows, we wish to stay as close to the FF5 definition of earnings as possible. Since the FF5 RMW factor corresponds to earnings before taxes, depreciation, and amortization (EBTDA), our corresponding adjustment to earnings is the equivalent of EBTDA with intangibles investment. That is, we add back the expense of intangibles investment to the FF5 version of earnings. We focus on three alternative measures of intangibles investment a) one that adds to earnings R&D expenses only as in Lev and Sougiannis (1996) (noted with a superscript RD below), b) one that adds to earnings R&D expenses and 30% of SG&A expenses as in Peters and Taylor (2017) and Crouzet and Eberly (2023) (noted with a superscript RD+.3SG below), and c) one that adds to earnings R&D expenses and 100% of SG&A expenses as in Eisfeldt et al. (2022) (noted with a superscript RD+SG below). The first of these measures was included in early versions of Fama and French (2018) but not in the published version.

We first study the ability of intangibles-adjusted profitability measures to predict future cash flows, a requirement for them to be proxies for expected cash flows in (3). There are fewer individual firms available at longer horizons due to firms entering and dropping

[5]In chapter 12 "Things to consider about per-share earnings": *"Still another factor, important at times, is the choice between charging off research and development costs in the year they are incurred or amortizing them over a period of years"*.

[6]The SG&A expense has to be net of R&D because it is the practice of Compustat to include the R&D expenses in the SG&A expenses and record it as the data item XSGA unless it is included in the cost of goods sold by the company.

out of the sample (the number of firms with cash flow data at a horizon of ten years is 60% of those with data at a horizon of one year). Because of this, we run regressions using 50 portfolios sorted by the FF5 profitability measure ($Y(\text{FF})/B_t$) and Fama-French industry classification (details on portfolio formation are given below). Portfolios are rebalanced as firms exit the sample. In the profitability regressions (Table 1), $Y(\text{FF})/B_t$ is the single best predictor of profitability among the unadjusted profitability measures. A subset of these is reported in the table. The gross profitability measure of Novy-Marx (2013) has a higher average R^2 than net income (IB), for all horizons (Panels A and B of the table). The difference in average R^2 values is significant for horizons 3 and 10 (t-statistics for the difference are reported in Panel B). The EBTDA measure of Fama and French (2015) has a higher average R^2, relative to gross profitability, for all horizons (Panels A and B of the table) and the differences are significant at horizons of 1 through 7 years.

When we include the intangibles adjustments (Panels C and D) $Y(\text{RD})/B_t$ has less predictive power (measured by R^2) than $Y(\text{FF})/B_t$ at horizons of 1 to 2 years and greater at horizons of 3 to 10 years, with horizons 4 to 10 being significantly higher. This makes intuitive sense since R&D expenditures may take some time to impact cash flows. $Y(\text{RD}+.3\text{SG})/B_t$ has lower predictive power (measured by R_2) than $Y(\text{FF})/B_t$ at horizons of one year through three years and higher average R^2s for longer horizons (significant for years six through 10). $Y(\text{RD}+\text{SG})/B_t$ has lower predictive power (measured by average R^2) than $Y(\text{FF})/B_t$ from year one to year eight.

In Figure 1 of the internet appendix, we plot the time series of the difference in R^2 between the profitability measures $Y(\text{NM})/B$ versus $Y(\text{IB})/B$, $Y(\text{FF})/B$ versus $Y(\text{NM})/B$, and $Y(\text{RD})/B$ versus $Y(\text{FF})/B$ across three different forecasting horizons $\tau = 1, 5, 10$. Though $Y(\text{NM})/B$ does not always outperform $Y(\text{IB})/B$ in forecasting future profitability, $Y(\text{FF})/B$ clearly performs better than $Y(\text{NM})/B$. $Y(\text{RD})/B$ forecasts future profitability better than $Y(\text{FF})/B$ in the longer-horizon (5- and 10-year ahead) but worse in the short-horizon (1-year ahead).

Among measures that do not adjust for intangibles, the results in Table 1 provide empirical support for the ability of the earnings measure used in Fama and French (2015) to predict future cash flows (Panels A and B). The results also provide support for increased predictive power at longer horizons when using intangibles-adjusted profitability (Panels C and D).

Table 2 performs a similar portfolio-level cross-sectional regression in which the FF5 profitability measure, RD_t/B_t, SGA_t/B_t, asset growth, and book-to-market are included as explanatory variables together. Both the Fama-French profitability measure and R&D

adjustment are significant at all horizons. The SG&A adjustment is also positive at all horizons except for year one. The slope coefficients of R&D and SG&A adjustments generally increase with horizon. The results in Table 2 confirm the intuition from Table 1 that R&D takes some time to influence future cash flows. SG&A kicks in by year 1 and also grows in importance at longer horizons, consistent with portions of SG&A being investments in intangible capital that payoff in future cash flows. Thus the Fama and French (2015) EBTDA profitability measure as well as measures of EBTDA that adjust for investment in intangibles seem to pass the test of being able to predict future cash flows.

We report the predictive regression results for the cash dividends in Tables 3 and 4. In these tables, the dependent variables are future cash dividends in year $t+\tau$ divided by the book equity in year t and the independent variables remain the same as in the operating cash flows regressions. The portfolio-level regressions largely overcome the problem that many firms do not pay dividends. Differences between the results of forecasting cash dividends and the results of forecasting operating cash flows are that dividends are generally less forecastable than operating cash flows and $Y(\text{NM})/B$ does not appear better than $Y(\text{IB})/B$ when forecasting cash dividends while $Y(\text{RD})/B$ outperforms $Y(\text{FF})/B$ even in the short horizon (1-year ahead). The importance of the intangibles adjustment increases (both in magnitude and t-statistics of the slope coefficients) as the forecasting horizon grows.

Now that the profitability measures seem to pass the minimum requirement for being useful for being tied to future expected returns (or the ICC), that is having the ability to forecast future cash flows, we can move on to asking whether factors derived from them are useful in pricing the cross-section of asset returns. Before we do that, we provide a more detailed discussion of the underlying data.

3 R&D Disclosure and Compustat Data

The Financial Accounting Standards Board Statement number 2 (FASBS2, FASB (1974)) mandated all material R&D costs encompassed by the statement to be charged to expense when incurred and the statement became effective for fiscal years beginning on or after January 1, 1975. The accounting treatment of R&D costs varies across firms before 1975. Moreover, FASB (1974) required *"... research and development costs be charged to expense when incurred shall be applied retroactively by prior period adjustment, ... The prior period adjustment shall recognize any related income tax effect."* Therefore, the recorded R&D expense in restated financial statements prior to the fiscal year of 1975 may not have been available to the public at the time of original publication. Therefore, the performance of portfolios constructed using those restated data may contain look-ahead biases. For this reason, we restrict our analysis to the sample period from 1975 to 2021 when using any accounting information.

Moreover, even post-FASB2, not all firms report (positive) R&D expenses. In the calendar year 1975 40% of firms report R&D expenses, experiencing a slight decline (but always above 35%) before 2000, and then gradually increasing to 52% in 2021. Koh and Reeb (2015) state that the materiality threshold is typically defined as amounts exceeding one percent of sales. Since R&D expense is an item of mandatory disclosure, blank or empty R&D fields should imply zero or immaterial R&D expense. However, Koh and Reeb (2015) find that 10.5% of the missing R&D firms file and receive patents, which is 14 times greater than zero R&D firms. Koh and Reeb (2015), therefore, argue that a blank R&D field could also represent a firm's conscious decision not to separate R&D expense from other reported expenses, rather than implying zero R&D expense. Without a clean approach to impute missing R&D expense, we set it as zero if missing in this paper. The consequence is that our R&D measure is subject to an errors in variables (EIV) problem. The EIV issue should weaken our results and lead to a bias toward failing to find a significant effect of the intangibles adjustment. Goyal and Wahal (2023) hand-collect data on R&D for firms with missing data to either supplement or impute missing R&D data. They find that, while the number of firms with non-zero R&D increases significantly, aggregate industry R&D changes imperceptibly for most industries at the two-digit SIC code level (see their Table 1).

For the firms reporting R&D expense, the expense can be shown either as (1) a separate line item on the income statement or (2) separately stated in notes to their financial statements.[7] It is Compustat's standard to collect R&D expense (Compustat data item XRD) from either line items or notes to financial statements.

Since we construct both annual- and monthly-rebalanced profitability factors using annual and quarterly filings data respectively, we also examine the R&D data recorded in quarterly Computat. First, we find that no firms report R&D in the quarterly filings (quarterly Compustat data item XRDQ) before 1989 and only on average 41% of the firms have XRDQ record in the quarterly Compustat data between 1989 and 2021.[8] Moreover, among all firms that have XRDQ records, there are about 25% of the firms

[7] `https://www.irs.gov/businesses/corporations/faqs-irc-41-qres-and-asc-730-lbi-directive`.
[8] 98% of the firms reporting annual R&D expense report quarterly R&D expense.

report only once a year and the amount is the same as in the XRD recorded in annual Compustat data. We verify that firms reporting their R&D expense as a separate line item in the income statement typically have multiple XRDQ records in a year while the firms integrating R&D expense into other line items and only reporting in the notes to financial statement typically have only one XRDQ record each year.[9] From 1989 through 2021, the aggregate R&D expense for all firms with more than one XRDQ record is 8.8% of their total sales which is a much larger fraction relative to the 1.8% for firms with only one XRDQ record.

As an example, consider General Electric (GE). GE reports R&D expenses only in the notes to financial statements in the annual filings before the fiscal year ending in December 31, 2019. In their notes to financial statements in the 2019 10-K, GE explains "R&D expenses are classified in cost of goods and services sold in our consolidated Statement of Earnings (Loss)." After 2019, GE changes this practice and reports R&D expense as a line item in both quarterly and annual filings. This change is also reflected in the quarterly Compustat data.

GE integrates R&D expense in the cost of goods sold (Compustat data item COGS), though most firms include R&D expense in the selling, general and administrative expense (Compustat data item XSGA). Compustat does not provide any information to identify with which item the firm integrates R&D expense. This practice has implications for our analysis as we need to separate SG&A and R&D expenses.

We follow the procedure in Peters and Taylor (2017) and that R&D is included in reported SG&A. We, therefore, set our SG&A measure equal to reported SG&A minues reported R&D (XSGA minus XRD minus RDIP[10]) We add an additional screen: when XRD exceeds XSGA but is less than COGS, or when XSGA is missing, SG&A is set equal to XSGA with no further adjustments or zero if XSGA is missing. We recognize that this approach is not flawless as it does not work for firms with higher SG&A than R&D but integrating R&D in the cost of goods sold. We denote firms' ex-R&D values of SG&A as SGA and use XSGA to refer to the cum-R&D Compustat data item.

Panel A and B in Figure 1 report R&D expense as fractions of the total sales and book equity, respectively, for all firms for five industries (using the Fama and French industry definitions). R&D expense relative to sales and book equity exhibit heterogeneity across industries with HighTech and Health industries being the most R&D intensive industries

[9] 64% of the firms reporting quarterly R&D expense report four times a year.

[10] RDIP represents the portion of R&D considered to be "purchased" and written off immediately upon acquisition if the R&D items are deemed not to have an alternative use.

and experiencing the most rapid growth over the sample period.

4 Data

Monthly stock returns, dividends, stock prices, and numbers of common shares outstanding are from the Center for Research in Security Prices (CRSP), and annual and quarterly accounting information is from Compustat.[11] To avoid look-ahead bias in using R&D data, the accounting information is from 1975 to 2022. We include all firms with share codes 10 and 11 and traded on NYSE, Amex, and Nasdaq. Table 16 contains a comprehensive list of variable definitions.

The profitability of firms is measured as the earnings divided by total book equity. We define annual total book equity as Compustat data item CEQ[12] and quarterly book equity following Hou et al. (2015).[13] As mentioned above, we perform analyses using several measures of earnings but only report a subset of results to conserve space. Our earnings measures include (moving from top-line to bottom-line measures) (1) total revenue, $Y(\text{REVT})$, (2) gross profit, $Y(\text{NM})$ following Novy-Marx (2013), (3) gross profit minus interest expense, $Y(\text{RD}+\text{SG})$ as in Eisfeldt et al. (2022) for investment in intangibles, (4) gross profit minus interest expense minus 70% of SG&A expense, $Y(\text{RD}+.3\,\text{SG})$ following the definition in Peters and Taylor (2017) for investment in intangibles, (5) gross profit minus interest expense minus 100% of SG&A expense, $Y(\text{RD})$, (6) gross profit minus interest expense minus 100% of SG&A expense minus 100% of R&D expense, $Y(\text{FF})$ following definition in Fama and French (2015) for profitability, (7) gross profit minus interest expense minus 100% of SG&A expense minus 100% of R&D expense minus 100% depreciation, $Y(\text{EBT})$, (8) income before extraordinary items minus special items, $Y(\text{SPI})$, and (9) income before extraordinary items, $Y(\text{IB})$.

Due to the quarterly R&D data issues described in the earlier section, we focus on the annual-rebalanced profitability factors using annual Compustat data in most of our

[11]The results presented in this paper are based on the CRSP and Compustat data downloaded on December 2, 2023.

[12]The book equity computation in Fama and French (2018) has broader coverage than the Compustat data item CEQ before 1968. Since our sample period starts in 1975, using CEQ or computed book equity does not alter our results.

[13]Quarterly book equity is shareholders' equity, plus balance-sheet deferred taxes and investment tax credit (item TXDITCQ) if available, minus the book value of the preferred stock. Depending on availability, we use stockholders' equity (item SEQQ), or common equity (item CEQQ) plus the carrying value of the preferred stock (item PSTKQ), or total assets (item ATQ) minus total liabilities (item LTQ) in that order as shareholders' equity. We use redemption value (item PSTKRQ) if available, or carrying value for the book value of the preferred stock.

analyses. To be included in the profitability forecasting regression for $t + \tau$ where $\tau = 1, 2, ..., 10$ in Tables 1,2, 3, and 4, a firm must have Compustat data for calendar year t on book equity and revenue or income before extraordinary items. For the firms reporting accounting information more than once in a given calendar year, potentially due to the change of the fiscal year-end month, we use the most recently released information. In addition, to ensure that the accounting variables are known at the month of portfolio formation and return regressions, we follow Fama and French (2006) and include a firm in the portfolio and return regression for July of t+1 to June of t+2 if it has available book equity, revenue or income before extraordinary items in the calendar year t. This procedure guarantees an at least six-month period for the information to be released.[14] Figure 2 in the internet appendix depicts the timeline for the empirical implementation.

We exclude firms with negative book equity, missing share price, shares outstanding when they are used in the explanatory variables (that is, at time t). Following Fama and French (2006), we remove firms with total assets less than \$25 million or book equity less than \$12.5 million in year t to avoid the influence of small firms.

Other data, including various competing asset pricing factors, are taken from the authors' websites. In particular, the factors in the Fama-French five factor model (FF5) and the momentum (UMD and UMD(d)) are taken from Ken French's data library (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`). The same data library is also the source of the five industry portfolios that we use in the ROE prediction regressions. The factors in the Hou-Xue-Zhang four factor model (HXZ4) and their expected growth factor (R_EG) are from `https://global-q.org/factors.html`. The factors in the Daniel et al. (2020) behavioral three factor model (DHS3) are from `https://sites.google.com/view/linsunhom`. The factors in the Stambaugh and Yuan (2017) four factor model (SY4) are from `https://finance.wharton.upenn.edu/~stambaug/`. The intangibles-adjusted HML factor, HML(EKP) is from Eisfeldt et al. (2022) and available from `https://github.com/edwardtkim/intangiblevalue,`. The monthly rebalanced version of HML, HML(Devil) of Asness and Frazzini (2013) is from AQR.[15] All factor data were downloaded on December 2, 2023. SY4 is available until December 2016, HML(EKP) are available until March 2022, HXZ4, R_EG, and DHS3 data are available until December 2022, FF5, UMD, and HML(Devil) are available until

[14]This conservative implementation could include some stale accounting information. For example, for a firm with a fiscal year-end in February 2021 with announced earnings in April 2021, the accounting information is already fourteen months ago when we first use it in the portfolio and return regression analysis in July 2022. This will tend to bias against finding significant results.

[15]`https://www.aqr.com/Insights/Datasets/The-Devil-in-HMLs-Details-Factors-Monthly`.

June 2023.

Since firm-level cross-section predictive regressions with horizons of one to ten years are subject to severe survivorship biases, we forecast the future operating cash flows of fifty managed portfolios. In the internet appendix, we report the number of firms that remain available after τ years. 17%, 28%, and 47% of firms drop out of the sample after 3, 5, and 10 years respectively. To deal with the severe survivorship bias in the firm-level profitability regressions, we assign firms into 5-by-10 industry and $Y(\text{FF})/B$ groups. The portfolios are rebalanced at the end of each month in the next ten years to ensure that the portfolio weights on each firm are proportional to their respective market size. The portfolio-level profitability regressions take the portfolio-level profitability (earnings in year $t+\tau$ divided by book equity in year t) in year $t+\tau$ as the dependent variables and the current profitability, and other variables, in year t as the independent variables.

In the regressions reported in Tables 1, 2, 3, and 4), the earnings of portfolio k in the year τ after its formation as its net worth times the value-weighted earnings-to-market ratio, i.e.,

$$\mathrm{Y}_{k,t+\tau} = \underbrace{w_{k,t+\tau}}_{\text{portfolio net worth}} \sum_{i\in k} \underbrace{\frac{M_{i,t+\tau}}{\sum_{j\in k} M_{j,t+\tau}}}_{\text{weight on firm i}} \frac{\mathrm{Y}_{i,t+\tau}}{M_{i,t+\tau}} \qquad \text{where } \tau=\{1,2,\cdots,10\} \tag{4}$$

where $w_{k,t+\tau} = \prod_{s=\{1,2,\cdots,\tau\times 12\}} \left(1+\sum_{i\in k} \frac{M_{i,s-1}}{\sum_{j\in k} M_{j,s-1}} r_{i,s}\right)$ is one plus the cumulative ex-dividend return (CRSP data item RETX) of the portfolio from its formation to τ years later, where $r_{i,s}$ is the monthly return of firm i in month s after the portfolio formation.

Following Fama and French (2015), we leave at least six months for the firms to announce earnings. The portfolios are formed by the June-end in year t based on the accounting information in calendar year t-1 from Compustat. All the firms are sorted independently into 5-by-10 industry and $Y(\text{FF})/B$ groups and we keep tracking the portfolios' returns at the end of each month, earnings and cash flows at end of each June end for the next ten years after their formations based on the rebalanced scheme described earlier.

5 Profitability Factors with an R&D Adjustment

Though FF5 contains only annual-rebalanced factors, many other factor models are based on monthly-rebalanced portfolios. Therefore, we also construct monthly-rebalanced profitability factors using quarterly accounting data. While our focus is on annual-rebalanced, we also report results with monthly-rebalanced.

5.1 The performance of Profitability in Pricing Assets

Table 5 reports mean returns and (t-statistics) of portfolios ranked independently on market capitalization (based on the median market equity for NYSE firms) and various measures of profitability (three groups based on the 30^{th} and 70^{th} NYSE percentiles of profitability). The table also provides the difference in average returns in high- versus low-profitability firms for each size category. The last two columns show the mean profitability factor return spread and its alpha, relative to a one-factor model using the FF RMW factor. The return spreads are generally significant for all unadjusted profitability measures with t-statistics increasing from bottom-line profitability to top-line, except for revenue, which is less significant than gross profitability.

However, the largest average return spreads and t-statistics are for the three intangibles-adjusted measures, followed by the gross profitability measure of Novy-Marx (2013). Similarly, the only measures with significantly positive alpha versus RMW are the intangibles-adjusted measures and gross profitability.

Panel A and B in Figure 2 present the cumulative return of four selected annual- and monthly-rebalanced profitability factors in three equally divided subsamples. For annual-rebalanced profitability factors, Y(IB)/B underperforms in all three subsamples, Y(RD)/B performs the best in the later two subsamples, and Y(FF)/B leads the performance in the first subsample. For the monthly-rebalanced factors, the R&D-adjusted profitability factor has the best performance in all subsamples. Interestingly, Y(FF)/B and Y(IB)/B consistently deliver very similar results. Y(NM)/B does not do as well before 2008 but has strong performance after 2008 for both annual- and monthly-rebalanced factors.

It is common to test alternative pricing models by choosing some test assets and asking whether one model provides lower estimated abnormal returns (alphas) than a competing model. The alphas are typically measured as the intercept in a time series regression of the excess return on each test asset on the factor returns (or factor-mimicking portfolios). Barillas and Shanken (2017) and Barillas et al. (2020) show that, when factors are traded portfolios, alternative models can be compared (in terms of which one leads to higher squared Sharpe ratios) by asking if each model can price the factors in the alternative model. The left-hand side "test assets" drop out in this case (they also show that looking at test asset alphas may be misleading).

We study the mispricing (or alphas) of the profitability factors studied in Table 5 by regressing their excess returns on eight important multi-factor asset pricing models in the literature. The profitability alphas are reported in Table 6. The models that we use are (1) the FF5 model (Fama and French (2015)); (2) the FF5(C) model with RMW replaced by the cash-based operating profitability factor (Fama and French (2018)); (3) the FF6 model which adds the momentum factor, UMD, from Carhart (1997) to FF5; (4) FF6/Devil which is a version of the FF6 model in which the value factor, HML is replaced with a version that is updated on a monthly basis (Asness and Frazzini (2013)); (5) FF6/EKP which is a version of the FF6 model in which the value factor, HML is replaced with the intangibles-adjusted value factor from Eisfeldt et al. (2022); (6) the four-factor model of Hou et al. (2015); (7) the five-factor model from Hou et al. (2021); (8) the long- and short-horizon-mispricing thee-factor model of Daniel et al. (2020); and (9) the mispricing four-factor model of Stambaugh and Yuan (2017).

Our main results are the estimated values of α, $\hat{\alpha}$, for the array of profitability factors, and their associated t-statistics, relative to the eight competing models, which are reported in Table 6. For each measure of profitability, we do 2-by-3 sorts on size and profitability. For the annual rebalanced factors, the size sort breakpoint is the median size for NYSE firms. The profitability-sort breakpoints ate the 30^{th} and 70^{th} NYSE percentiles of profitability. Using the Fama-French terminology "Small Robust" is the value-weighted portfolio of small-cap firms with high profitability and "Small Weak" is the portfolio of small-cap firms with low profitability (with equivalents of "Big Robust" and "Big Weak" for large-cap firms). Our factors are constructed as 1/2 (Small Robust + Big Robust) - 1/2 (Small Weak + Big Weak).

The first three columns of results show the profitability factors moving from bottom-line to top-line measures: net income (IB), income before special items (SPI), earnings before taxes (EBT), and earnings before taxes depreciation and amortization (EBTDA), used in Fama and French (2015) (FF), have insignificant values of $\hat{\alpha}$. This is not surprising since our predictive regressions for operating cash show that IB, SPI, and EBT all have less predictive power than (EBTDA). The results for FF in column 4 are also not surprising since the RMW is one of the factors in the first 4 models. The fact that $\hat{\alpha}$ for our version of the FF5 profitability factor is not identically zero is just an indication that our construction of RMW differs slightly from the version of RMW on Ken French's data library, which is the source of RMW on the LHS of the regression.

The next three columns include versions of our intangibles-adjusted profitability factors. The factor $r_{\text{ROE}}^{\text{RD}}$ uses the FF5 measure of profitability plus R&D expenditures. The factor $r_{\text{ROE}}^{\text{RD+.3SG}}$ also adds 30% of SG&A and $r_{\text{ROE}}^{\text{RD+SG}}$ also adds 100% of SG&A. For $r_{\text{ROE}}^{\text{RD}}$ and $r_{\text{ROE}}^{\text{RD+.3SG}}$ the values of $\hat{\alpha}$ are positive and the largest across all profitability measures and statistically are significant across all eight benchmark models. That is, adding these factors to the benchmark factors would significantly increase the squared Sharpe ratio of the portfolio space spanned by the factors. The t-statistics range from 2.24 to 7.05 and exceed 3.63 for all models except the HMXZ5 model. The factor which adds R&D and all of SG&A, $r_{\mathrm{ROE}}^{\mathrm{RD+SG}}$, and the gross profitability measure, $r_{\mathrm{ROE}}^{\mathrm{NM}}$, have slightly smaller alphas and have significantly positive $\hat{\alpha}$ for all models except HMXZ5 and SY4. The seven significant alphas for $r_{\mathrm{ROE}}^{\mathrm{RD+SG}}$ have t-statistics ranging from 2.12 to 4.42 while the significant alphas for $r_{\mathrm{ROE}}^{\mathrm{NM}}$ have t-statistics ranging from 1.94 to 3.58. The results for $r_{\mathrm{ROE}}^{\mathrm{RD+SG}}$ and $r_{\mathrm{ROE}}^{\mathrm{NM}}$ are close since profitability, when adding back both R&D and SG&A, differs from gross profitability only by interest expense.

Our main focus is on factor risk premia and the results of Table 6 are consistent with a significant premium to the intangibles-adjusted profitability factors. For some factors, it can be the case that abnormal returns come mainly from the short side of the factor portfolio. A practitioner attempting to implement a smart beta portfolio based on a given factor might have concerns that there are large asset borrowing costs on the short side. We study which side is the source of abnormal returns by looking at the values of $\hat{\alpha}$ for the long and short sides separately.

Panel B-1 and C-1 of Table 6 show our results. Panel B-1 shows the long side for annual-rebalanced factors. $r_{\mathrm{ROE}}^{\mathrm{RD}}$ and $r_{\mathrm{ROE}}^{\mathrm{RD+.3SG}}$ have significant long-side alphas for each benchmark model, with one exception. $r_{\mathrm{ROE}}^{\mathrm{RD+.3SG}}$ has an insignificant long-side alpha relative to SY4. Panel C-1 shows the short side for annual-rebalanced factors. The values of $\hat{\alpha}$ for $r_{\mathrm{ROE}}^{\mathrm{RD}}$ and $r_{\mathrm{ROE}}^{\mathrm{RD+.3SG}}$ are statistically significant for all models, but two (HMXZ5 and DHS3). In particular, while the long-side alphas were not very strong against SY4, the short-side alphas are. Similarly, while the short-side alphas were not very strong against HMXZ5 and DHS3, the long-side alphas are.

The takeaway here is that the intangibles-adjusted profitability measures add explanatory power for the cross-section of assets beyond that captured by the models studied here. Against the nine benchmark models $r_{\mathrm{ROE}}^{\mathrm{RD}}$ and $r_{\mathrm{ROE}}^{\mathrm{RD+.3SG}}$ have abnormal returns coming from both sides of the position for five models, from the long side for two models, and from the short side for one model. Therefore, the significant intangibles adjusted factor alphas are not solely coming from the short side.

5.2 What Fraction of SG&A?

The evidence so far suggests that adjusting earnings for R&D expense is important, however, it remains unclear how much of SG&A should be added back to earnings. The results in Table 5 suggest that adding 30% of SG&A can improve the explanatory power of the intangibles-adjusted factors. Alpha increases in size, versus only adjusting for R&D, for seven of the eight models, although the t-statistics only increase for two of the eight models. Adding 100% of SG&A always reduces both the magnitude of the estimated αs and the t-statistics. Table 7 repeats the type of analysis of Table 5, for profitability factors that adjust the fraction of SG&A added back to earnings from 0% to 100% in increments of 10%. The highest levels of α are always for adding a fraction of 20% (with two ties to 30% and one tie to 0%-10%-30%). The t-statistics for the estimated α are the largest at 20% for seven models and at 10% for one model. This suggests that R&D expense plays a crucial role in strengthening the relationship between profitability measure and expected returns while a fraction of SG&A expense is important. This is consistent with our predictive regressions in Tables 1 and 3.

5.3 Subsample Performance

The cumulative returns shown in Figure 2, Panel A, suggest a potential regime shift in the returns to both the FF5 factor, RMW, and the intangibles-adjusted profitability factors, around the year 2000. We wish to determine if the significance of these factors is a post-2000 phenomenon. We report the time-series average monthly percent returns and t-statistics of several related factors/anomalies for the full sample and two subperiods (using December 1999 as the breakpoint) in Table 8. All the factors considered here are significant in the full sample. Most of the factor t-statistics fall in the second subsample, except for three factors, $r^{\mathrm{NM}}_{\mathrm{ROA}}$, $r^{\mathrm{BGLN}}_{\mathrm{ROA}}$, and R&D-adjusted profitability. The monthly-rebalanced ROE factor from Hou et al. (2015) (R_ROE) and the momentum factor become insignificant in the second subsample. The two R&D-adjusted profitability factors, $r^{\mathrm{RD}}_{\mathrm{ROE}}$ and $r^{\mathrm{RD}}_{\mathrm{ROEQ}}$, have the highest t-statistics in the second subsample. Importantly, the two R&D-adjusted profitability factors are not solely post-2000 phenomena, with both being statistically significant in each subperiod.

5.4 Relations with Profitability/Intangibles-Related Factors

Many existing factor models include a profitability or earnings factor, e.g., RMW in FF5, R_ROE in HXZ4, PEAD in DHS3, and PERF in SY4. The (price) momentum factor is often related to the earnings momentum phenomenon. For example, Chan et al. (1996) argue that price momentum is largely due to the market's underreaction to earnings-related information. Even though these factors are all connected to the earnings or post-earnings market reactions, they manifest very different properties of some asset price anomalies. For example, the monthly-rebalanced PEAD and R_ROE factors can subsume the (price) momentum factor while Fama and French (2016) show that a momentum factor is required to take away the significance of the momentum phenomenon though they have an annual-rebalanced profitability factor, RMW, in FF5.

We examine the relation of the R&D-adjusted profitability factor with profitability factors in existing models and five other potentially related factors, gross profitability ($r^{\text{NM}}_{\text{ROA}}$), gross profit divided by total assets, by Novy-Marx (2013), profitability factors by Ball et al. (2015) ($r^{\text{BGLN}}_{\text{ROA}}$) and Ball et al. (2016) ($r^{\text{BGLN}}_{\text{CROA}}$), quality-minus-junk factor by Asness et al. (2019) (QMJ), and an intangibles-adjusted value factor (HML(EKP)) by Eisfeldt et al. (2022). Panel A-1 and A-2 in Table 9 report the regression results of the related factors on RMW and the full Fama-French five-factor model respectively. All included factors are significant against RMW and FF5. Panel B-1 and B-2 report results of similar regressions but with RMW replaced by $r^{\text{RD}}_{\text{ROE}}$. In the univariate regressions, $r^{\text{RD}}_{\text{ROE}}$ takes away the significance of most of the factors except for PEAD and HML(EKP). In the multiple regressions, however, some of the related factors turn significant again due to their negative relation with the value factor (HML) in FF5. Comparing the results on panels A and B, it is clear that the R&D adjustment is important for the profitability factor.

Panel C-1 and C-2 regress $r^{\text{RD}}_{\text{ROE}}$ on the related factors. The alphas of $r^{\text{RD}}_{\text{ROE}}$ are significant against all included factors in both univariate and multiple regressions. As independent variables, RMW, $r^{\text{FF}}_{\text{CROE}}$, $r^{\text{BGLN}}_{\text{ROA}}$, $r^{\text{BGLN}}_{\text{CROA}}$, PERF, and QMJ have the strongest explanatory power for $r^{\text{RD}}_{\text{ROE}}$. Two factors, PEAD and HML(EKP) are unrelated with $r^{\text{RD}}_{\text{ROE}}$ with insignificant (or negative) slope coefficients and very low R^2s.

5.5 Within Industry Performance

Figure 1 shows that firms within HighTech and Health industries are highly R&D intensive and their R&D expenses as a fraction of total sales have been growing through the entire sample period. The R&D adjustment might, therefore, have a stronger impact on the firms within these two industries. We examine the performance of the profitability factors within each of the Fama-French five industries and report them in Table 10. Specifically, we assign firms into two size groups based on median within-industry NYSE market equity and three profitability groups based on 30^{th} and 70^{th} percentile within-industry NYSE profitability at the June end of year t. The firms are held until the June end of the year $t+1$. The within-industry profitability factors are 1/2(Small Robust+Big Robust) - 1/2(Small Weak+Big Weak) within each industry. As before, we select the earnings measures from the bottom-line earnings to the top-line, including IB, SPI, EBT, FF, RD, RD+.3SG, NM, and REVT.

Panel A shows that the HighTech and Health industries have the highest fraction of firms reporting positive R&D expenses (column RD>0) and the highest R&D expenses as factions of both book equity (column RD/B) and total sales (column RD/REVT). Panel B reports the mean excess return of the profitability factors and the associated t-statistics. The only factor portfolios that have a statistically significant mean return across all industries are our three factors RD RD+.3SG, and RD+SG; and the Novy-Marx factor, NM.

5.6 Alternative Scaling Factors

The valuation equation (3) uses the book value of equity only as a scaling factor. The use of book equity makes our results directly comparable to those of Fama and French (2006). However, there may be better scaling factors, which is an empirical question. We study the alternative profitability factors withan array of alternative scalings: revenue (REVT), book value of equity (B), total assets (AT), gross PP&E plus intangible capital (as in Peters and Taylor (2017), B^{PT}), and market equity. Table 11 reports the time-series average of the monthly value-weighted strategy returns and the alpha against FF5 and HXZ4 (along with t-statistics). For all scaling factors either $Y(\mathrm{RD})$ or $Y(\mathrm{RD}+.3\mathrm{SG})$ deliver the highest mean return and either $Y(\mathrm{RD})$, $Y(\mathrm{RD}+.3\mathrm{SG})$, or $Y(\mathrm{RD}+\mathrm{SG})$ deliver the highest t-statistics for the mean return. Similarly, either $Y(\mathrm{RD})$, $Y(\mathrm{RD}+.3\mathrm{SG})$, or $Y(\mathrm{RD}+\mathrm{SG})$ deliver the highest alphas (relative to FF5 and HXZ4) and t-statistics for alpha.

Across scaling factors using intangibles-adjusted book value (B^{PT}) almost universally yields the smallest t-statistics for the alphas against FF5 and HXZ4. While scaling by intangibles-adjusted book value has some intuitive appeal, the adjustment for intangibles might by yielding a noisy estimate of adjusted book value. Gulen et al. (2023) propose a intangibles-adjusted profitability factor that adjusts both the numerator and denominator, with industry-specific capitalization rates for SG&A and depreciation rates for R&D. Adding industry specific parameters should reduce bias in estimating intangibles-adjusted book values. However, there is loss in degrees of freedom from estimating a larger number parameters, so the net effect on precision is an open question.

6 Model Comparison Tests

In this section, perform additional tests of the ability of the intangibles-adjusted factor to expand the portfolio space, relative to other asset pricing models. We rely on the test of Barillas et al. (2020) which tests whether there is a statistically significant improvement in the squared Sharpe ratio of one model versus another. Our model comparisons take two forms.

The first, which is only applicable to models which include profitability-related factors (all except DHS3), we replace the original profitability factor in the benchmark model with $r^{\mathrm{RD}}_{\mathrm{ROE}}$. Thus the comparison is the model with its profitability factor replaced with $r^{\mathrm{RD}}_{\mathrm{ROE}}$. We denote this as MODEL/$r^{\mathrm{RD}}_{\mathrm{ROE}}$ in Table 12.

The second type of test asks whether adding the R&D-adjusted profitability factor to the benchmark model improves the squared Sharpe ratio. We denote this as MODEL+$r^{\mathrm{RD}}_{\mathrm{ROE}}$ in the table. For example, this tests, when applied to the FF5 model, asks whether a six-factor model consisting of the FF5 factors, plus $r^{\mathrm{RD}}_{\mathrm{ROE}}$, has a significantly higher R^2 than the FF5 model.

In Table 12 we report the results of the Barillas et al. (2020) test for the full period and two subperiods. The tests are against the same nine benchmark models used above. The first row for each model is the change in the squared Sharpe ratio from either swapping profitability factors or adding the R&D-adjusted profitability factor as an extra factor (that is the improvement of the model listed in the column over the model listed in the row). The second row for each comparison is the p-value for the null hypothesis that the change in the Sharpe ratio is zero.

For the augmented models, where we add $r^{\mathrm{RD}}_{\mathrm{ROE}}$ as a factor, all of the tests reject the null of no improvement. The largest p-value across the full sample and two subsamples is 0.012. Thus the R&D-adjusted factor improves the portfolio opportunity set.

For the models that replace an existing profitability factor with our R&D-adjusted profitability, the improvement in squared Sharpe ratio is not significant for the FF5(C) (for the full sample and the first subsample) nor for HMXZ5 (for the first subsample only).

The overall picture from the Barillas et al. (2020) tests is that the R&D-adjusted profitability factor is important for explaining the cross-section of asset returns.

7 Factor or Anomaly

Whenever a new pricing factor is proposed there is always the concern about whether the identified factor is a true proxy for systematic risk or an anomaly. If the latter, is it true mispricing, which is likely to be arbitraged away over time, or is it due to data mining? The data realities of R&D reporting do not allow us the type of out-of-sample analysis in Mclean and Pontiff (2016) or Linnainmaa and Roberts (2018), so we report some suggestive statistics.

The profitability factors across various models are linked to expected returns through the present value relation in Equation 3. However, as noted by Fama and French (2006, 2015), Equation 3 is unable to distinguish between risk factors and anomalies. The fact that the pricing results are stable across time is of some comfort, but certainly not definitive.

A true factor should be risky. We look at the risk of profitability factors in two ways. We first ask what is the maximal drawdown for each factor. The results are reported in Table 13. All of the profitability factors have non-trivial downside risk although it is clear that the intangibles-adjusted profitability factors have the smallest realized downside.

Table 14 looks at the behavior of the profitability factors during market downturns. One might expect from their construction (and even the name "Robust Minus Weak") that these factors might provide some downside protection since weak firms might be expected to suffer more from economic downturns. That is born out in the table. The factors have higher returns when the excess return on the market is negative than when it is positive. This is true for all of the profitability factors, not just those adjusted for intangibles.

The resilient performance during the down markets could be attributed to a "protective" put option feature embedded in the profitability factors (see, e.g., Henriksson and Merton (1981) and Jagannathan and Korajczyk (1986)). We can test this conjecture by having an additional term to capture the put option features in the a CAPM-style time-series regression.

$$r^{i,t}_{\text{ROE}} = \alpha^s_i + \beta^U_i \text{RMRF} + \beta^{U-D}_i \max[-\text{RMRF}, 0] + \epsilon_{i,t}. \tag{5}$$

The intercept term, α^s_i, measures the abnormal return from pure security selection, that is, returns independent of any put like feature. The coefficient β^U_i measures the portfolio's up-market beta and the coefficient β^{U-D}_i measure the difference between the portfolio's up-market and down-market betas. If the factor portfolio has a put-like feature, then we should see that the estimated value of β^{U-D}_i is significantly positive. That is, a strategy with an embedded put will have a higher (lower) beta when the excess return on the market is positive (negative). β^{U-D}_i measures the number of put options needed to replicate the strategy's put feature. In a managed portfolio, this put feature could be ascribed to market timing ability. For the factor portfolios the put-like feature could be from the underlying put-like properties of the assets held or dynamic rebalanced inherent in the portfolio strategy. Jagannathan and Korajczyk (1986) argue that a measure which combines both sources of performance is

$$\frac{\alpha^s_i}{1 + r_f} + \beta^{U-D}_i \times p. \tag{6}$$

Here p is the value of a put option on the market index normalized to unity, with one period to expiration and an exercise price of $1 + r_f$.

Table 15 reports the regression results from 5 for three profitability factors, RMW from FF5 (FF); our measure that adds R&D to earnings (RD); and gross profitabilty (NM).

For our factor portfolio, r^{RD}_{ROE}, the intercept, 0.24, and the coefficient of the put, 0.14, are both significant. We estimate the combined measure of performance assuming the monthly interest rate is zero, and the monthly standard deviation of the market is 5%. The Black-Scholes value of the p is \$0.02 per month, or 2% of the amount invested in the index. Hence the option adjusted total alpha of $r^{\text{RD}}_{\text{ROE}}$ is $0.24\% + 0.14 \times 2\% = 0.24 + 0.28 = 0.52\%$. It is interesting to have both α^s_i and β^{U-D}_i positive and significant since it suggests that investors get paid a premium for holding a "protective" put.

The three other profitability factors, IB, FF and NM, also have positive estimates of α^s_i and β^{U-D}_i, but only the value of α^s_i for FF is statistically significant.

We report the regression results of the long side and short side of the profitability factors separately in the middle and right columns of Table 15. All four factor portfolios have statistically significant estimates of α^s_i for the long side of the portfolio. None of the four factor portfolios have significant estimated values of β_i^{U-D} on the long side. In fact, they are all negative. On the short side, the numbers in the table are for a long position on the short side of the portfolio. Shorting the short side would yield coefficients with the opposite sign. There is no indication of the short side of the factor portfolios adding to the factor portfolios level of α_i^s. In fact, the short side leads to smaller α_i^s for the long-short portfolio. Three out of four factors have significantly negative values for β_i^{U-D}. This implies that the assets in the short side decline much more in a down market than they rebound in an up market. Taking a short position in these assets yields positive values of β_i^{U-D} of the long-short portfolio. However the net effect is only statistically significant for our R&D-adjusted earnings factor.

8 Conclusions

Fama and French (2015) develop a clever method for constructing an economy-wide pervasive risk factor that captures shocks to firms' profitability, which they denote RMW (robust-minus-weak). Constructing this factor requires measuring firms' profitability. Fama and French (2015) use earnings before taxes and depreciation (EBTD). Even though R&D and some components of SG&A expenditures should rightly be be treated as investments in intangible assets, GAAP requires full expensing of such expenditures over our sample period. We use the Fama-French definition of earnings (EBTD) but adjust it to reflect investments in intangibles by adding the R&D expenditures (or R&D and a fraction of SG&A). We call the resulting earnings intangibles-adjusted earnings.

Intangibles-adjusted earnings forecast the cross section of future operating cash flows and cash dividends of firms better than unadjusted earnings at longer horizons. The profitability factors that we construct based on this adjustment have a statistically and economically significant alphas against an array of existing factor models. When the profitability factor in the FF5 model is replaced by this R&D-adjusted profitability factor, it is able to subsume various anomalies like momentum and operating leverage.

The long side of our intangibles-adjusted factor portfolios picks stocks which outperform the market, on a risk adjusted basis, independently of whether the market has positive or negative excess returns. The short side of our intangibles-adjusted factor portfolios does well by shorting stocks that do particularly poorly in down markets, thus giving the strategy a put-like feature. This the nature of the superior performance of the intangibles-adjusted factors comes from two very distinct sources.

References

Abel, Andrew B., Janice C. Eberly, Stavros Panageas. 2013. Optimal inattention to the stock market with information costs and transactions costs. *Econometrica* **81**(4) 1455–1481. URL http://www.jstor.org/stable/23524182.

Arnott, Robert D, Campbell R Harvey, Vitali Kalesnik, Juhani T Linnainmaa. 2021. Reports of value's death may be greatly exaggerated. *Financial Analysts Journal* **77**(1) 44–67.

Asness, Clifford, Andrea Frazzini. 2013. The devil in hml's details. *The Journal of Portfolio Management* **39**(4) 49–68. URL https://doi.org/10.3905/jpm.2013.39.4.049.

Asness, Clifford S, Andrea Frazzini, Lasse Heje Pedersen. 2019. Quality minus junk. *Review of Accounting Studies* **57** 34–112. URL https://doi.org/10.1007/s11142-018-9470-2.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, Valeri Nikolaev. 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics* **121**(1) 28–45.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, Valeri V Nikolaev. 2015. Deflating profitability. *Journal of Financial Economics* **117**(2) 225–248.

Bansal, Ravi, Amir Yaron. 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance* **59**(4) 1481–1509.

Barillas, Francisco, Raymond Kan, Cesare Robotti, Jay Shanken. 2020. Model comparison with sharpe ratios. *Journal of Financial and Quantitative Analysis* **55**(6) 1840–1874.

Barillas, Francisco, Jay Shanken. 2017. Which alpha? *The Review of Financial Studies* **30**(4) 1316–1338.

Belo, Frederico, Vito D Gala, Juliana Salomao, Maria Ana Vitorino. 2022. Decomposing firm value. *Journal of Financial Economics* **143**(2) 619–639.

Berk, Jonathan B. 1995. A critique of size-related anomalies. *The Review of Financial Studies* **8**(2) 275–286. URL http://www.jstor.org/stable/2962273.

Bloomberg, Lawrence N. 1934. The investment value of goodwill. Ph.D. thesis, Johns Hopkins University, Baltimore, MD.

Braun, Phillip A. 1993. Asset pricing and capital investment. Ph.D. thesis, University of Chicago, Chicago, IL.

Breeden, Douglas T., Michael R. Gibbons, Robert H. Litzenberger. 1989. Empirical test of the consumption-oriented capm. *The Journal of Finance* **44**(2) 231–262. URL http://www.jstor.org/stable/2328589.

Carhart, Mark M. 1997. On persistence in mutual fund performance. *The Journal of Finance* **52**(1) 57–82. URL http://www.jstor.org/stable/2329556.

Chan, Louis KC, Narasimhan Jegadeesh, Josef Lakonishok. 1996. Momentum strategies. *The Journal of Finance* **51**(5) 1681–1713.

Chen, Nai-Fu, Richard Roll, Stephen A. Ross. 1986. Economic forces and the stock market. *The Journal of Business* **59**(3) 383–403. URL http://www.jstor.org/stable/2352710.

Cochrane, John H. 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. *The Journal of Finance* **46**(1) 209–237. URL http://www.jstor.org/stable/2328694.

Connor, Gregory. 1995. The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal* **51**(3) 42–46. URL http://www.jstor.org/stable/4479845.

Connor, Gregory, Matthias Hagmann, Oliver Linton. 2012. Efficient semiparametric estimation of the Fama–French model and extensions. *Econometrica* **80**(2) 713–754.

Connor, Gregory, Robert A. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* **15**(3) 373–394. doi:https://doi.org/10.1016/0304-405X(86)90027-9. URL https://www.sciencedirect.com/science/article/pii/0304405X86900279.

Connor, Gregory, Robert A. Korajczyk. 1988. Risk and return in an equilibrium apt: Application of a new test methodology. *Journal of Financial Economics* **21**(2) 255–289. doi:https://doi.org/10.1016/0304-405X(88)90062-1. URL https://www.sciencedirect.com/science/article/pii/0304405X88900621.

Connor, Gregory, Robert A. Korajczyk. 1991. The attributes, behavior, and performance of u.s. mutual funds. *Review of Quantitative Finance and Accounting* **1**(1) 5–26. URL `https://doi.org/10.1007/BF02408404`.

Crouzet, Nicolas, Janice Eberly. 2023. Rents and intangible capital: A q+ framework. *The Journal of Finance* **78**(4) 1873–1916.

Daniel, Kent, David Hirshleifer, Lin Sun. 2020. Short-and long-horizon behavioral factors. *The Review of Financial Studies* **33**(4) 1673–1736.

Eisfeldt, Andrea L., Edward T. Kim, Dimitris Papanikolaou. 2022. Intangible value. *Critical Finance Review* **11**(2) 299–332.

Eisfeldt, Andrea L, Dimitris Papanikolaou. 2013. Organization capital and the cross-section of expected returns. *The Journal of Finance* **68**(4) 1365–1406.

Fama, Eugene F, Kenneth R French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**(1) 3–56.

Fama, Eugene F, Kenneth R French. 2006. Profitability, investment and average returns. *Journal of Financial Economics* **82**(3) 491–518.

Fama, Eugene F, Kenneth R French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* **116**(1) 1–22.

Fama, Eugene F, Kenneth R French. 2016. Dissecting anomalies with a five-factor model. *The Review of Financial Studies* **29**(1) 69–103.

Fama, Eugene F, Kenneth R French. 2018. Choosing factors. *Journal of Financial Economics* **128**(2) 234–252.

FASB. 1974. Statement of financial accounting concepts no. 2, accounting for research and development costs. *Financial Accounting Standards Board, Norwalk, CT* .

Ge, Shuyi, Shaoran Li, Oliver Linton. 2022. Dynamic peer groups of arbitrage characteristics. *Journal of Business & Economic Statistics* 1–24.

Gebhardt, William R, Charles MC Lee, Bhaskaran Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Research* **39**(1) 135–176.

Goyal, Amit, Sunil Wahal. 2023. R&d, innovation, and the stock market. *Available at SSRN 4568392* .

Graham, Benjamin. 1973. The intelligent investor, a book of practical counsel, 4th rev. ed. *New York: Harper & Brothers Publishers* .

Graham, Benjamin, David L. Dodd, Sidney Cottle. 1962. *Security Analysis: Principles and Technique*. McGraw-Hill Book Company, New York, NY.

Greenwald, Bruce C, Judd Kahn, Paul Sonkin, Michael van Biema. 2001. Value investing: from graham to buffett and beyond. *John Wiley & Sons* 107–137.

Gulen, Huseyin, Dongmei Li, Ryan H. Peters, Morad Zekhnini. 2023. Intangible capital in factor models. *Management Science* **Forthcoming**. URL `http://dx.doi.org/10.2139/ssrn.3725002`.

Hansen, Lars Peter, Kenneth J Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society* 1269–1286.

Hansen, Lars Peter, Kenneth J Singleton. 1983. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of political economy* **91**(2) 249–265.

Henriksson, Roy D., Robert C. Merton. 1981. On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills. *The Journal of Business* **54**(4) 513–533. URL `http://www.jstor.org/stable/2352722`.

Hou, Kewei, Haitao Mo, Chen Xue, Lu Zhang. 2021. An augmented q-factor model with expected growth. *Review of Finance* **25**(1) 1–41.

Hou, Kewei, Chen Xue, Lu Zhang. 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies* **28**(3) 650–705.

Jagannathan, Ravi, Robert A. Korajczyk. 1986. Assessing the market timing performance of managed portfolios. *The Journal of Business* **59**(2) 217–235. URL `http://www.jstor.org/stable/2353018`.

Jagannathan, Ravi, Ellen McGrattan, Anna Scherbina. 2001. The declining us equity premium.

Jagannathan, Ravi, Yong Wang. 2007. Lazy investors, discretionary consumption, and the cross-section of stock returns. *The Journal of Finance* **62**(4) 1623–1661. URL http://www.jstor.org/stable/4622313.

Jagannathan, Ravi, Zhenyu Wang. 1996. The conditional capm and the cross-section of expected returns. *The Journal of Finance* **51**(1) 3–53.

Jagannathan, Ravi, Yang Zhang. 2020. A return based measure of firm quality. Tech. rep., National Bureau of Economic Research.

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

Kelly, Bryan, Seth Pruitt, Yinan Su. 2019. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics* **134** 501–534.

Kim, Soohun, Robert A. Korajczyk, Andreas Neuhierl. 2021. Arbitrage Portfolios. *The Review of Financial Studies* **34**(6) 2813–2856. doi:10.1093/rfs/hhaa102. URL https://doi.org/10.1093/rfs/hhaa102.

Kogan, Leonid, Dimitris Papanikolaou. 2014. Growth opportunities, technology shocks, and asset prices. *The Journal of Finance* **69**(2) 675–718.

Koh, Ping-Sheng, David M Reeb. 2015. Missing r&d. *Journal of Accounting and Economics* **60**(1) 73–94.

Lev, Baruch, Theodore Sougiannis. 1996. The capitalization, amortization, and value-relevance of r&d. *Journal of Accounting and Economics* **21**(1) 107–138.

Li, Kevin K, Partha Mohanram. 2014. Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies* **19**(3) 1152–1185.

Linnainmaa, Juhani T, Michael R Roberts. 2018. The history of the cross-section of stock returns. *The Review of Financial Studies* **31**(7) 2606–2649.

Mclean, R. David, Jeffrey Pontiff. 2016. Does academic research destroy stock return predictability? *The Journal of Finance* **71**(1) 5–32. doi:https://doi.org/10.1111/jofi.12365. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12365.

Monhanram, Partha S. 2005. Separating winners from losers among lowbook-to-market stocks using financial statement analysis. *Review of Accounting Studies* **10** 133–170.

Novy-Marx, Robert. 2013. The other side of value: The gross profitability premium. *Journal of Financial Economics* **108**(1) 1–28.

Papanikolaou, Dimitris. 2011. Investment shocks and asset prices. *Journal of Political Economy* **119**(4) 639–685.

Parker, Jonathan A, Christian Julliard. 2005. Consumption risk and the cross section of expected returns. *Journal of Political Economy* **113**(1) 185–222.

Peters, Ryan H, Lucian A Taylor. 2017. Intangible capital and the investment-q relation. *Journal of Financial Economics* **123**(2) 251–272.

Piotroski, Joseph D. 2000. Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research* **38** 1–41. URL `http://www.jstor.org/stable/2672906`.

Roll, Richard, Stephen A. Ross. 1980. An empirical investigation of the arbitrage pricing theory. *The Journal of Finance* **35**(5) 1073–1103. URL `http://www.jstor.org/stable/2327087`.

Rosenberg, Barr, James Guy. 1976. Prediction of beta from investment fundamentals: part one. *Financial Analysts Journal* **32**(3) 60–72.

Rosenberg, Barr, Walt McKibben. 1973. The prediction of systematic and specific risk in common stocks. *Journal of Financial and Quantitative Analysis* **8**(2) 317–333. URL `http://www.jstor.org/stable/2330027`.

Stambaugh, Robert F, Yu Yuan. 2017. Mispricing factors. *The Review of Financial Studies* **30**(4) 1270–1315.

Stroyny, Alvin L. 2005. Estimating a combined linear factor model. *Linear Factor Models in Finance*. Elsevier, 210–225.

Wang, Kai. 2023. Organizational capital, tobin's q, and the cross section of stock returns. *Tobin's Q, and the Cross Section of Stock Returns (February 19, 2023)* URL `http://dx.doi.org/10.2139/ssrn.4364379`.

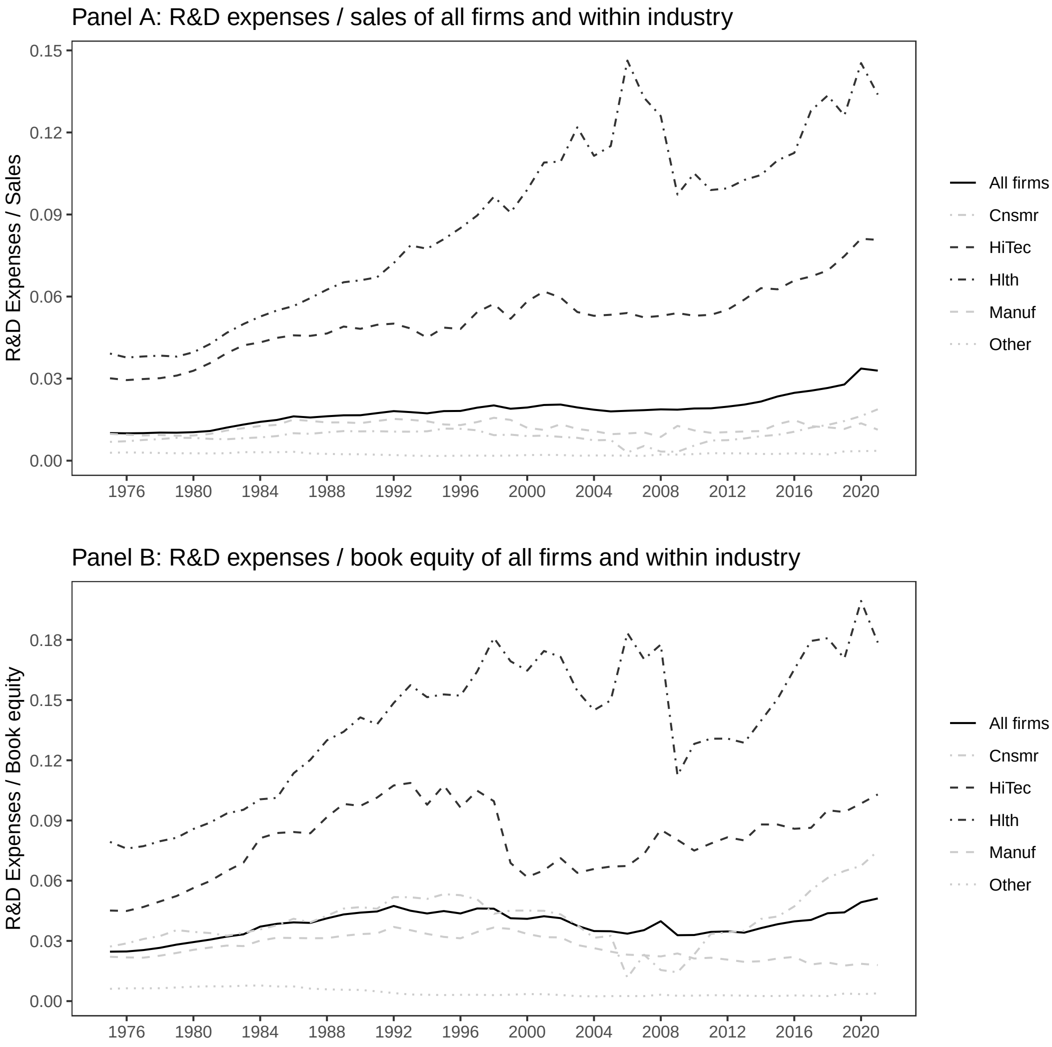

Figure 1: R&D expenses as fractions of sales and book equity, 1975 to 2022.

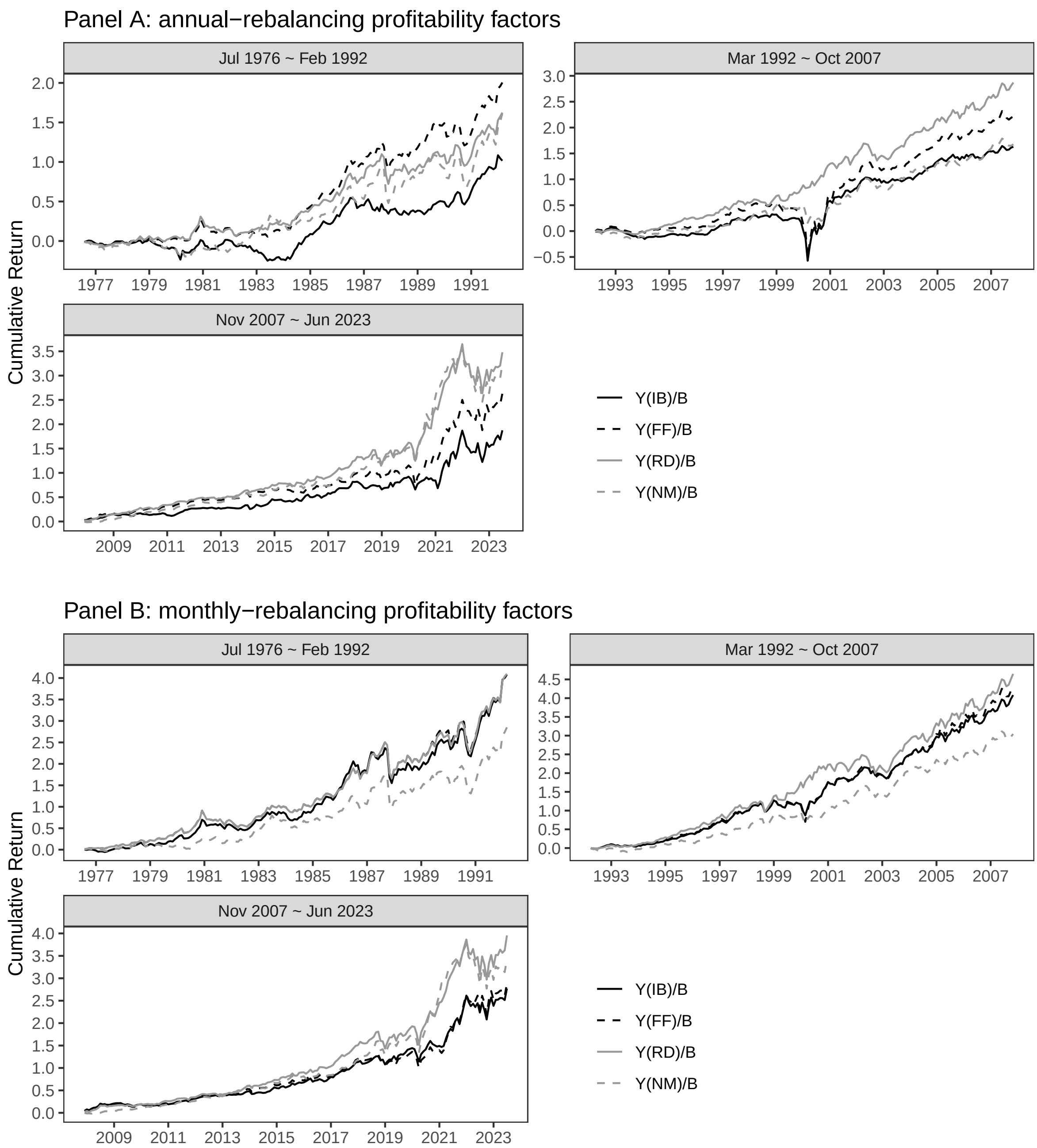

Figure 2: Cumulative return of annual- and monthly-rebalanced profitability factors, July 1976 to June 2023. Cumulative returns for each subsample are compounded long-side of the factor minus the compounded short-side of the factor.

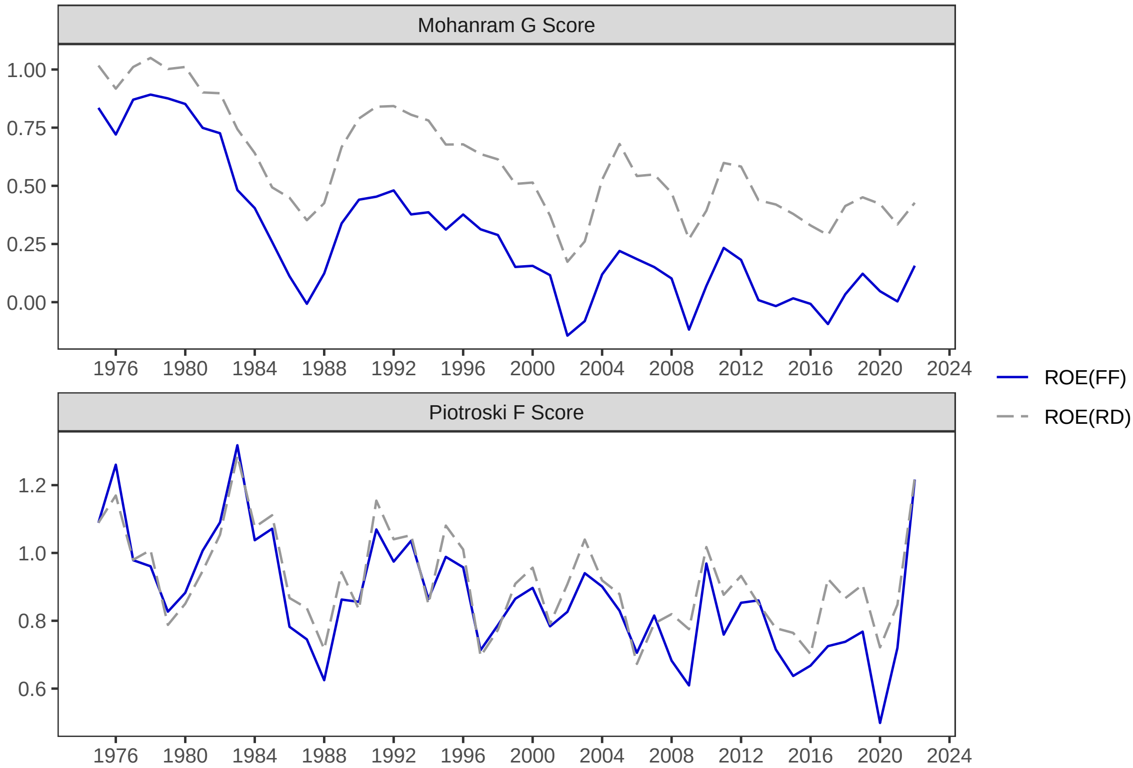

Figure 3: Difference in the equal-weighted average of Mohanram G Score and Piotroski F Score between high Y(FF)/B or Y(RD)/B group and low Y(FF)/B or Y(RD) group. High and low Y(FF)/B or Y(RD)/B groups are based on the 30th and 70th NYSE Y(FF)/B or Y(RD)/B. The time-series average of the difference in Mohanram G Score for Y(FF)/B is 0.28, and it is 0.60 for Y(RD)/B. The time-series average of the difference in Piotroski F Score for Y(FF)/B is 0.87, and it is 0.92 for Y(RD)/B.

Table 1: Cross-section univariate regression of future profitability (operating cash flows) at the portfolio level (1978 - 2022).

This table reports the time-series averages of the slope coefficients, t-statistics, and r-squared for the cross-section regressions of future operating cash flows (divided by current book value of equity) at the portfolio level. The operating cash flow is the Compustat data item OANCF which becomes available on Compustat after 1988. Firms are assigned into 5-by-10 industry and ROE(FF) groups in year t. Industry groups are based on Fama-French five industry classification and ROE(FF) cut points are NYSE ROE(FF) decile within each industry group. For notation convenience, we denote $Z^{X}_{t+\tau}$ as $Y^{X}_{t+\tau}/B_t$ where X includes IB, NM, FF, RD, RD+.3 SG, and RD+SG. The portfolio-level earnings and book equity are the value-weighted earnings and book equity for a given portfolio in year t. The portfolio-level future cash flows is computed following equation (4). We also test the null hypothesis that the difference of the r-squared time series between $Z^{\mathrm{NM}}_{t+\tau}$, $Z^{\mathrm{FF}}_{t+\tau}$, $Z^{\mathrm{RD}}_{t+\tau}$, $Z^{\mathrm{RD+.3\,SG}}_{t+\tau}$, $Z^{\mathrm{RD+SG}}_{t+\tau}$ and $Z^{\mathrm{IB}}_{t+\tau}$, $Z^{\mathrm{NM}}_{t+\tau}$, $Z^{\mathrm{FF}}_{t+\tau}$, $Z^{\mathrm{FF}}_{t+\tau}$, $Z^{\mathrm{FF}}_{t+\tau}$ is zero and report the t-statistics in the panel B and D. We remove firms with total assets less than \$12.5 million or book equity less than \$25 million in year t.

Panel A: Average slopes for Z^{IB}_t, Z^{NM}_t, and Z^{FF}_t

	R^2	Intc	Z^{IB}_t		R^2	Intc	Z^{NM}_t		R^2	Intc	Z^{FF}_t
Z^{OCF}_{t+1}	0.56	0.15	1.03	Z^{OCF}_{t+1}	0.58	0.04	0.31	Z^{OCF}_{t+1}	0.74	0.06	0.71
Z^{OCF}_{t+2}	0.50	0.17	1.05	Z^{OCF}_{t+2}	0.55	0.04	0.32	Z^{OCF}_{t+2}	0.67	0.08	0.72
Z^{OCF}_{t+3}	0.47	0.19	1.07	Z^{OCF}_{t+3}	0.53	0.05	0.34	Z^{OCF}_{t+3}	0.62	0.10	0.73
Z^{OCF}_{t+4}	0.45	0.21	1.15	Z^{OCF}_{t+4}	0.51	0.06	0.36	Z^{OCF}_{t+4}	0.58	0.11	0.77
Z^{OCF}_{t+5}	0.41	0.23	1.18	Z^{OCF}_{t+5}	0.47	0.08	0.36	Z^{OCF}_{t+5}	0.53	0.13	0.78
Z^{OCF}_{t+6}	0.38	0.26	1.21	Z^{OCF}_{t+6}	0.44	0.10	0.38	Z^{OCF}_{t+6}	0.49	0.16	0.79
Z^{OCF}_{t+7}	0.35	0.28	1.26	Z^{OCF}_{t+7}	0.40	0.12	0.39	Z^{OCF}_{t+7}	0.44	0.18	0.83
Z^{OCF}_{t+8}	0.30	0.30	1.32	Z^{OCF}_{t+8}	0.35	0.14	0.40	Z^{OCF}_{t+8}	0.37	0.21	0.84
Z^{OCF}_{t+9}	0.28	0.33	1.41	Z^{OCF}_{t+9}	0.34	0.16	0.43	Z^{OCF}_{t+9}	0.35	0.24	0.87
Z^{OCF}_{t+10}	0.25	0.36	1.51	Z^{OCF}_{t+10}	0.31	0.18	0.46	Z^{OCF}_{t+10}	0.31	0.27	0.90

Panel B: t-statistics for Panel A

		Intc	Z^{IB}_t		ΔR^2 (NM-IB)	Intc	Z^{NM}_t		ΔR^2 (FF-NM)	Intc	Z^{FF}_t
1		11.09	13.67	1	0.55	3.87	19.25	1	9.33	4.66	24.66
2		10.74	15.41	2	1.64	4.34	17.99	2	6.83	5.96	25.76
3		12.55	12.62	3	2.64	5.27	20.39	3	5.36	7.16	22.07
4		14.64	13.32	4	2.49	5.89	21.22	4	4.19	7.60	20.80
5		21.14	13.84	5	2.46	7.02	17.58	5	3.65	10.36	18.72
6		25.98	16.06	6	2.60	7.62	17.82	6	2.64	12.13	18.55
7		30.84	24.00	7	1.85	5.33	12.30	7	2.21	13.11	17.48
8		32.62	23.43	8	2.19	4.41	8.76	8	1.66	12.52	13.78
9		37.63	28.63	9	2.61	5.83	11.01	9	0.72	13.64	16.65
10		48.36	28.94	10	3.08	6.54	13.84	10	-0.29	10.73	13.24

Panel C: Average slopes for Z^{RD}_t, $Z^{\mathrm{RD+.3\,SG}}_t$, and $Z^{\mathrm{RD+SG}}_t$

	R^2	Intc	Z^{RD}_t		R^2	Intc	$Z^{\mathrm{RD+.3\,SG}}_t$		R^2	Intc	$Z^{\mathrm{RD+SG}}_t$
Z^{OCF}_{t+1}	0.71	0.04	0.64	Z^{OCF}_{t+1}	0.68	0.03	0.52	Z^{OCF}_{t+1}	0.55	0.05	0.31
Z^{OCF}_{t+2}	0.66	0.05	0.66	Z^{OCF}_{t+2}	0.65	0.04	0.55	Z^{OCF}_{t+2}	0.53	0.06	0.33
Z^{OCF}_{t+3}	0.62	0.07	0.68	Z^{OCF}_{t+3}	0.61	0.05	0.56	Z^{OCF}_{t+3}	0.51	0.07	0.34
Z^{OCF}_{t+4}	0.59	0.08	0.72	Z^{OCF}_{t+4}	0.59	0.06	0.59	Z^{OCF}_{t+4}	0.49	0.09	0.36
Z^{OCF}_{t+5}	0.56	0.10	0.74	Z^{OCF}_{t+5}	0.55	0.08	0.61	Z^{OCF}_{t+5}	0.46	0.10	0.37
Z^{OCF}_{t+6}	0.53	0.13	0.76	Z^{OCF}_{t+6}	0.53	0.10	0.63	Z^{OCF}_{t+6}	0.44	0.13	0.38
Z^{OCF}_{t+7}	0.48	0.14	0.80	Z^{OCF}_{t+7}	0.48	0.12	0.67	Z^{OCF}_{t+7}	0.40	0.14	0.40

Z^{OCF}_{t+8}	0.42	0.17	0.83	Z^{OCF}_{t+8}	0.42	0.14	0.70	Z^{OCF}_{t+8}	0.36	0.16	0.42
Z^{OCF}_{t+9}	0.40	0.19	0.86	Z^{OCF}_{t+9}	0.41	0.16	0.73	Z^{OCF}_{t+9}	0.35	0.18	0.44
Z^{OCF}_{t+10}	0.35	0.22	0.90	Z^{OCF}_{t+10}	0.37	0.18	0.77	Z^{OCF}_{t+10}	0.32	0.20	0.47
Panel D: t-statistics for Panel D											
	ΔR^2 (RD-FF)	Intc	Z^{RD}_t		ΔR^2 (RD+.3 SG-FF)	Intc	$Z^{\mathrm{RD+.3\,SG}}_t$		ΔR^2 (RD+SG-FF)	Intc	$Z^{\mathrm{RD+SG}}_t$
1	-7.04	3.77	23.23	1	-6.87	2.60	22.61	1	-10.93	5.25	16.91
2	-1.71	4.87	26.01	2	-2.74	3.35	23.39	2	-7.99	5.67	16.99
3	0.42	6.09	24.38	3	-0.67	4.48	23.52	3	-6.56	6.87	17.90
4	2.03	6.78	24.17	4	0.73	5.03	23.68	4	-5.35	7.31	17.85
5	4.22	9.12	21.57	5	1.86	7.40	22.57	5	-3.68	9.16	17.24
6	5.38	9.66	19.52	6	3.18	8.30	21.60	6	-2.88	9.68	16.93
7	4.75	8.75	16.27	7	3.13	6.94	16.63	7	-2.02	7.23	13.03
8	6.01	7.21	11.85	8	4.60	5.52	11.65	8	-1.10	5.80	9.42
9	7.03	9.78	15.70	9	5.72	8.26	15.87	9	0.00	8.63	13.00
10	6.52	9.16	16.14	10	5.39	8.95	18.99	10	0.82	9.55	15.67

Table 2: Cross-section multiple regression of future profitability (operating cash flows) at the portfolio level (1978 - 2022).

This table reports the time-series averages of the slope coefficients, t-statistics, and adjusted r-squared for the cross-section regressions of future operating cash flows (divided by current book value of equity) at the portfolio level. The operating cash flow is the Compustat data item OANCF which becomes available on Compustat after 1988. The explanatory variables include Fama-French operating profits divided by book equity ($Z^{\mathrm{FF}}_{t+\tau}$), R&D costs divided by book equity (RD_t), SG&A costs divided by book equity (SGA_t), change of total assets divided by one-year lagged total assets (AG_t), and book-to-market (BM_t). We separate R&D costs and SG&A costs following Peters and Taylor (2017). Firms are assigned into 5-by-10 industry and ROE(FF) groups in year t. Industry groups are based on Fama-French five industry classification and ROE(FF) cut points are NYSE ROE(FF) decile within each industry group. The portfolio-level earnings, R&D costs, SG&A costs, total assets, and book equity are the value-weighted for a given portfolio in year t. The portfolio-level future cash flows is computed following equation (4). We remove firms with total assets less than $12.5 million or book equity less than $25 million in year t.

Panel A: Average slopes							
	Adj. R^2	Intc	Z^{FF}_t	RD_t/B_t	SGA_t/B_t	AG_t	BM_t
Z^{OCF}_{t+1}	0.76	0.04	0.70	0.17	0.02	-0.12	0.03
Z^{OCF}_{t+2}	0.70	0.06	0.70	0.24	0.05	-0.15	0.02
Z^{OCF}_{t+3}	0.65	0.11	0.68	0.25	0.05	-0.13	-0.04
Z^{OCF}_{t+4}	0.61	0.14	0.69	0.29	0.05	-0.07	-0.07
Z^{OCF}_{t+5}	0.57	0.18	0.65	0.44	0.07	-0.09	-0.12
Z^{OCF}_{t+6}	0.55	0.21	0.65	0.48	0.08	-0.10	-0.13
Z^{OCF}_{t+7}	0.51	0.26	0.64	0.57	0.09	-0.09	-0.17
Z^{OCF}_{t+8}	0.44	0.32	0.60	0.66	0.09	-0.04	-0.24
Z^{OCF}_{t+9}	0.43	0.36	0.60	0.69	0.12	0.03	-0.27
Z^{OCF}_{t+10}	0.41	0.47	0.52	0.64	0.13	0.06	-0.35
Panel B: t-statistics							
		Intc	Z^{FF}_t	RD_t/B_t	SGA_t/B_t	AG_t	BM_t
1		2.76	24.23	4.56	1.21	-4.30	1.39
2		2.47	22.77	4.58	2.04	-5.78	0.51
3		3.93	16.34	4.26	2.35	-3.02	-1.09
4		5.56	17.29	5.57	2.40	-1.19	-3.11
5		5.88	12.59	6.44	2.95	-1.73	-3.98
6		5.45	12.14	4.97	3.06	-1.96	-4.21
7		5.01	10.50	4.11	2.74	-1.60	-4.39
8		5.42	7.49	5.30	3.23	-0.70	-5.38
9		6.25	9.19	6.76	3.28	0.33	-6.50
10		6.28	5.34	4.96	3.02	0.69	-7.47

Table 3: Cross-section univariate regression of future profitability (cash dividends) at the portfolio level (1975 - 2022).

This table reports the time-series averages of the slope coefficients, t-statistics, and r-squared for the cross-section regressions of future cash dividends (divided by current book value of equity) at the portfolio level. Cash dividends is Compustat data item DVC. Firms are assigned into 5-by-10 industry and ROE(FF) groups in year t. Industry groups are based on Fama-French five industry classification and ROE(FF) cut points are NYSE ROE(FF) decile within each industry group. For notation convenience, we denote $Z^X_{t+\tau}$ as $Y^X_{t+\tau}/B_t$ where X includes IB, NM, FF, RD, RD+.3 SG, and RD+SG. The portfolio-level earnings and book equity are the value-weighted earnings and book equity for a given portfolio in year t. The portfolio-level future cash flows is computed following equation (4). We also test the null hypothesis that the difference of the r-squared time series between $Z^{\text{NM}}_{t+\tau}$, $Z^{\text{FF}}_{t+\tau}$, $Z^{\text{RD}}_{t+\tau}$, $Z^{\text{RD+.3 SG}}_{t+\tau}$, $Z^{\text{RD+SG}}_{t+\tau}$ and $Z^{\text{IB}}_{t+\tau}$, $Z^{\text{NM}}_{t+\tau}$, $Z^{\text{FF}}_{t+\tau}$, $Z^{\text{FF}}_{t+\tau}$, $Z^{\text{FF}}_{t+\tau}$ is zero and report the t-statistics in the panel B and D. We remove firms with total assets less than \$12.5 million or book equity less than \$25 million in year t.

Panel A: Average slopes for Z^{IB}_t, Z^{NM}_t, and Z^{FF}_t

	R^2	Intc	Z^{IB}_t		R^2	Intc	Z^{NM}_t		R^2	Intc	Z^{FF}_t
Z^{DVC}_{t+1}	0.50	0.03	0.26	Z^{DVC}_{t+1}	0.45	0.00	0.07	Z^{DVC}_{t+1}	0.57	0.01	0.15
Z^{DVC}_{t+2}	0.51	0.03	0.29	Z^{DVC}_{t+2}	0.46	0.00	0.08	Z^{DVC}_{t+2}	0.57	0.01	0.17
Z^{DVC}_{t+3}	0.51	0.03	0.32	Z^{DVC}_{t+3}	0.45	0.00	0.08	Z^{DVC}_{t+3}	0.55	0.01	0.18
Z^{DVC}_{t+4}	0.50	0.03	0.36	Z^{DVC}_{t+4}	0.45	0.00	0.09	Z^{DVC}_{t+4}	0.53	0.01	0.20
Z^{DVC}_{t+5}	0.49	0.03	0.39	Z^{DVC}_{t+5}	0.45	0.00	0.10	Z^{DVC}_{t+5}	0.52	0.01	0.22
Z^{DVC}_{t+6}	0.49	0.03	0.43	Z^{DVC}_{t+6}	0.45	-0.01	0.11	Z^{DVC}_{t+6}	0.50	0.01	0.23
Z^{DVC}_{t+7}	0.47	0.03	0.48	Z^{DVC}_{t+7}	0.44	-0.01	0.12	Z^{DVC}_{t+7}	0.48	0.01	0.25
Z^{DVC}_{t+8}	0.46	0.03	0.53	Z^{DVC}_{t+8}	0.43	-0.01	0.13	Z^{DVC}_{t+8}	0.46	0.01	0.27
Z^{DVC}_{t+9}	0.44	0.03	0.57	Z^{DVC}_{t+9}	0.43	-0.01	0.14	Z^{DVC}_{t+9}	0.43	0.02	0.29
Z^{DVC}_{t+10}	0.42	0.04	0.60	Z^{DVC}_{t+10}	0.41	-0.01	0.15	Z^{DVC}_{t+10}	0.41	0.02	0.30

Panel B: t-statistics for Panel A

		Intc	Z^{IB}_t		ΔR^2 (NM-IB)	Intc	Z^{NM}_t		ΔR^2 (FF-NM)	Intc	Z^{FF}_t
1		6.00	16.47	1	-1.37	0.50	6.66	1	6.22	2.44	13.71
2		3.28	15.02	2	-1.66	0.10	5.20	2	6.28	1.18	11.32
3		5.26	13.47	3	-1.81	0.03	6.24	3	5.44	1.97	14.98
4		5.04	14.13	4	-1.75	-0.25	7.66	4	5.20	1.47	15.44
5		4.24	11.79	5	-1.38	-0.72	10.67	5	4.39	1.53	16.17
6		4.70	12.70	6	-1.42	-0.94	10.63	6	3.58	1.57	19.65
7		4.35	9.99	7	-1.21	-1.52	12.33	7	2.71	1.39	18.78
8		4.76	10.97	8	-1.24	-1.66	11.57	8	1.79	1.55	15.15
9		5.55	10.02	9	-0.64	-2.13	12.39	9	0.43	1.71	16.24
10		6.63	10.43	10	-0.40	-1.45	13.83	10	-0.43	2.38	17.64

Panel C: Average slopes for Z^{RD}_t, $Z^{\text{RD+.3 SG}}_t$, and $Z^{\text{RD+SG}}_t$

	R^2	Intc	Z^{RD}_t		R^2	Intc	$Z^{\text{RD+.3 SG}}_t$		R^2	Intc	$Z^{\text{RD+SG}}_t$
Z^{DVC}_{t+1}	0.58	0.00	0.15	Z^{DVC}_{t+1}	0.57	0.00	0.12	Z^{DVC}_{t+1}	0.45	0.01	0.07
Z^{DVC}_{t+2}	0.59	0.00	0.16	Z^{DVC}_{t+2}	0.58	0.00	0.13	Z^{DVC}_{t+2}	0.46	0.00	0.08
Z^{DVC}_{t+3}	0.58	0.00	0.17	Z^{DVC}_{t+3}	0.57	0.00	0.14	Z^{DVC}_{t+3}	0.45	0.00	0.08
Z^{DVC}_{t+4}	0.56	0.00	0.19	Z^{DVC}_{t+4}	0.56	0.00	0.16	Z^{DVC}_{t+4}	0.46	0.00	0.09
Z^{DVC}_{t+5}	0.55	0.00	0.21	Z^{DVC}_{t+5}	0.55	-0.01	0.17	Z^{DVC}_{t+5}	0.46	0.00	0.10
Z^{DVC}_{t+6}	0.54	0.00	0.22	Z^{DVC}_{t+6}	0.54	-0.01	0.19	Z^{DVC}_{t+6}	0.46	0.00	0.11
Z^{DVC}_{t+7}	0.52	0.00	0.24	Z^{DVC}_{t+7}	0.53	-0.01	0.21	Z^{DVC}_{t+7}	0.45	0.00	0.12
Z^{DVC}_{t+8}	0.50	0.00	0.26	Z^{DVC}_{t+8}	0.52	-0.01	0.23	Z^{DVC}_{t+8}	0.44	0.00	0.14

Z_{t+9}^{DVC}	0.48	0.00	0.28	Z_{t+9}^{DVC}	0.51	-0.01	0.24	Z_{t+9}^{DVC}	0.44	0.00	0.15
Z_{t+10}^{DVC}	0.46	0.01	0.30	Z_{t+10}^{DVC}	0.49	-0.01	0.26	Z_{t+10}^{DVC}	0.43	0.00	0.16

Panel D: t-statistics for Panel C

	ΔR^2 (RD-FF)	Intc	Z_t^{RD}		ΔR^2 (RD+.3 SG-FF)	Intc	$Z_t^{\mathrm{RD+.3\,SG}}$		ΔR^2 (RD+SG-FF)	Intc	$Z_t^{\mathrm{RD+SG}}$
1	3.38	0.81	9.20	1	0.21	0.01	9.09	1	-6.81	0.93	6.27
2	3.45	0.39	10.65	2	0.60	-0.34	7.84	2	-6.96	0.53	4.23
3	4.10	0.53	11.31	3	1.90	-0.47	9.06	3	-5.58	0.61	5.69
4	4.30	0.36	16.00	4	3.48	-0.74	14.16	4	-5.14	0.32	7.47
5	4.94	0.29	16.22	5	4.79	-1.20	16.17	5	-4.12	0.18	10.39
6	6.15	0.37	21.48	6	6.38	-1.41	19.79	6	-3.10	0.07	11.02
7	7.94	0.23	21.10	7	7.61	-1.90	20.14	7	-2.17	-0.25	12.15
8	8.56	0.22	16.57	8	8.58	-1.89	17.79	8	-1.05	-0.38	11.91
9	8.17	0.41	18.14	9	8.68	-2.09	18.56	9	0.26	-0.72	13.33
10	9.01	0.76	21.32	10	9.78	-1.69	20.71	10	1.41	-0.29	15.71

Table 4: Cross-section multiple regression of future profitability (cash dividends) at the portfolio level (1975 - 2022).

This table reports the time-series averages of the slope coefficients, t-statistics, and adjusted r-squared for the cross-section regressions of future operating cash flows (divided by current book value of equity) at the portfolio level. Cash dividends is Compustat data item DVC. The explanatory variables include Fama-French operating profits divided by book equity ($Z^{\text{FF}}_{t+\tau}$), R&D costs divided by book equity (RD_t), SG&A costs divided by book equity (SGA_t), change of total assets divided by one-year lagged total assets (AG_t), and book-to-market (BM_t). We separate R&D costs and SG&A costs following Peters and Taylor (2017). Firms are assigned into 5-by-10 industry and ROE(FF) groups in year t. Industry groups are based on Fama-French five industry classification and ROE(FF) cut points are NYSE ROE(FF) decile within each industry group. The portfolio-level earnings, R&D costs, SG&A costs, total assets, and book equity are the value-weighted for a given portfolio in year t. The portfolio-level future cash flows is computed following equation (4). We remove firms with total assets less than $12.5 million or book equity less than $25 million in year t.

Panel A: Average slopes							
	Adj. R^2	Intc	Z^{FF}_t	RD_t/B_t	SGA_t/B_t	AG_t	BM_t
Z^{DVC}_{t+1}	0.62	0.00	0.15	0.11	0.02	-0.06	0.01
Z^{DVC}_{t+2}	0.63	0.01	0.16	0.11	0.02	-0.06	0.01
Z^{DVC}_{t+3}	0.61	0.01	0.17	0.11	0.02	-0.06	0.01
Z^{DVC}_{t+4}	0.60	0.01	0.18	0.12	0.02	-0.06	0.00
Z^{DVC}_{t+5}	0.59	0.01	0.19	0.13	0.03	-0.07	0.00
Z^{DVC}_{t+6}	0.58	0.02	0.19	0.14	0.04	-0.07	-0.01
Z^{DVC}_{t+7}	0.56	0.03	0.20	0.16	0.04	-0.08	-0.02
Z^{DVC}_{t+8}	0.54	0.03	0.21	0.18	0.04	-0.08	-0.03
Z^{DVC}_{t+9}	0.53	0.04	0.21	0.17	0.05	-0.06	-0.05
Z^{DVC}_{t+10}	0.51	0.05	0.22	0.21	0.06	-0.06	-0.05
Panel B: t-statistics							
		Intc	Z^{FF}_t	RD_t/B_t	SGA_t/B_t	AG_t	BM_t
1		0.17	15.32	3.68	1.51	-4.74	1.34
2		0.35	13.61	2.94	1.22	-4.95	0.80
3		0.32	12.48	2.48	1.69	-4.57	0.82
4		0.53	10.94	2.39	2.31	-3.74	0.24
5		0.53	10.13	2.36	2.54	-3.48	-0.04
6		0.86	10.92	2.61	2.62	-3.63	-0.56
7		0.92	8.92	2.49	2.44	-3.20	-0.96
8		1.03	7.82	2.48	2.78	-3.01	-1.11
9		1.47	7.50	1.51	2.75	-2.76	-2.19
10		1.31	7.07	2.48	3.63	-2.00	-1.87

Table 5: Summary statistics for 2-by-3 ME-ROE(X) portfolios and factor percent returns (July 1976 - June 2023).

In the June end of each year t, firms are assigned into two size (ME) groups based on median June end market equity of NYSE firms, and assigned independently into three profitability (Y/B) groups based on the 30 and 70 percentiles profitability of all NYSE firms reported annual earnings in calendar year t-1. We report the time-series mean of the portfolio return over the one-month Treasury bill rate and the associated t-statistics. H-L is a portfolio that is long the high-ROE group and short the low-ROE group within each size group, and Mean(H-L) is the equal-weighted average of the two H-L portfolios. a(RMW) is the intercept and the associated t-statistics in the regression of Mean(H-L) on the profitability factor RMW of the Fama-French 5-factor model. All other statistics reported in this table are the time-series average of the monthly statistics of each portfolio. N(firms) is the number of firms in each portfolio. RD/B is the total R&D expense as fractions of total book equity. Overlap(FF) is the fraction of firms that are in both $Y(X)/B$ portfolio and the corresponding $Y(\text{FF})/B$ portfolio.

		Small				Big					
		Low	Med	High	H-L	Low	Med	High	H-L	Mean(H-L)	a(RMW)
Y(IB)/B	Mean	0.67	0.89	0.98	0.31	0.59	0.64	0.69	0.10	0.20	-0.13
	t-statistics	2.33	4.19	3.98	2.57	2.56	3.44	3.73	0.88	1.98	-2.86
	N(firms)	1732	1017	596		171	353	343			
Y(SPI)/B	Mean	0.63	0.92	1.00	0.36	0.53	0.64	0.70	0.17	0.26	-0.06
	t-statistics	2.27	4.28	3.99	3.24	2.27	3.47	3.78	1.41	2.68	-1.42
	N(firms)	1818	957	570		170	358	338			
Y(EBT)/B	Mean	0.62	0.90	1.00	0.38	0.50	0.61	0.73	0.23	0.30	-0.07
	t-statistics	2.16	4.08	4.25	2.75	2.14	3.28	3.95	1.97	2.72	-1.44
	N(firms)	1728	994	614		175	345	345			
Y(FF)/B	Mean	0.60	0.92	1.03	0.43	0.45	0.63	0.75	0.30	0.36	-0.01
	t-statistics	2.18	4.24	4.15	3.35	1.99	3.30	4.16	2.64	3.53	-0.47
	N(firms)	1808	969	560		188	356	321			
	RD/B	0.05	0.02	0.03		0.02	0.03	0.06			
Y(RD)/B	Mean	0.52	0.94	1.09	0.58	0.40	0.63	0.74	0.35	0.46	0.24
	t-statistics	1.96	4.28	4.19	6.80	1.81	3.39	3.98	3.28	6.08	5.12
	N(firms)	1690	1022	624		176	355	334			
	Overlap(FF)	0.90	0.83	0.88		0.87	0.84	0.90			
	RD/B	0.03	0.03	0.08		0.01	0.02	0.08			
Y(RD+.3SG)/B	Mean	0.50	0.93	1.10	0.60	0.43	0.62	0.77	0.34	0.47	0.28
	t-statistics	1.95	4.13	4.18	6.78	2.10	3.22	4.19	3.37	6.16	4.90
	N(firms)	1430	1131	775		216	355	295			
Y(RD+SG)/B	Mean	0.54	0.91	1.04	0.50	0.51	0.62	0.79	0.28	0.39	0.23
	t-statistics	2.16	3.87	4.00	4.77	2.62	3.18	4.33	2.79	4.74	3.21
	N(firms)	1111	1273	953		249	365	251			
Y(NM)/B	Mean	0.55	0.93	1.03	0.49	0.50	0.64	0.78	0.27	0.38	0.21
	t-statistics	2.24	3.94	3.93	4.72	2.59	3.32	4.15	2.72	4.65	3.06
	N(firms)	1198	1199	940		257	363	244			
Y(REVT)/B	Mean	0.64	0.93	1.02	0.38	0.57	0.70	0.76	0.19	0.28	0.09
	t-statistics	2.63	3.90	3.90	3.39	2.97	3.77	3.93	2.09	3.26	1.30
	N(firms)	1434	1004	899		334	334	197			

Table 6: Pricing tests on annual- and monthly-rebalanced profitability factors (July 1976 - June 2023). To construct annual-rebalanced profitability factors, r^{X}_{ROE}, in the June end of each year t, firms are assigned into two size groups based on the median June end market equity of NYSE firms and assigned independently into three profitability groups based on the 30 and 70 percentiles profitability of all NYSE firms reported annual earnings in calendar year t-1. To construct monthly-rebalanced profitability factors, r^{X}_{ROEQ}, at the end of each month, firms are assigned into two size (ME) groups based on median June end market equity of NYSE firms, and assigned independently into three profitability groups based on the 30 and 70 percentiles (most recent available) profitability of all NYSE firms. The profitability factor return in any given month is the average return of the portfolios that are long the high profitability portfolios and short the low profitability portfolios within each size group. Firms' most recent available profitability is updated at the month end of the most recent quarterly earnings announcement date (quarterly Compustat data item RDQ, exclude if missing). To remove stale earnings, we exclude a firm if its most recent available quarter-end is 6 months away. The quarterly R&D expenses (quarterly Compustat data item XRDQ) are only recorded after 1989 and some firms may report only once a year, we simply set missing XRDQ to zero with no further interpolation. The considered profitability measures include $Y(\text{IB})/B$, $Y(\text{SPI})/B$, $Y(\text{EBT})/B$, $Y(\text{FF})/B$, $Y(\text{RD})/B$, $Y(\text{RD} + .3\text{SG})/B$, $Y(\text{RD} + \text{SG})/B$, $Y(\text{NM})/B$, and $Y(\text{REVT})/B$. Long- and short-side of the factors is the equal-weighted average excess return of the high- and low-profitability portfolios in each month. We run time-series regressions of the constructed profitability factors (Table A-1 for annual-rebalanced long-minus-short, A-2 for monthly-rebalanced long-minus-short, B-1 for annual-reblanacing long side only, B-2 for monthly-reblanacing long side only, C-1 for annual-reblanacing short side only, and C-2 for monthly-reblanacing short side only) on the existing factor models and report their intercept and associated t-statistics in italic. The factor models include Fama-French five factor model (FF5), FF5 with RMW replaced by the cash-based operating profitability factor (FF5(C)), FF5 plus momentum factor (FF6), FF6 with the value factor, HML, replaced by the monthly-rebalanced value factor with most recent market equity, HML(Devil), from Asness and Frazzini (2013), FF6 with the value factor, HML, replaced by the intangible-adjusted value factor, HML(EKP), from Eisfeldt et al. (2022), Hou-Xue-Zhang four factor model (HXZ4) and Hou-Mo-Xue-Zhang five factor model (HMXZ5), Daniel-Hirshleifer-Sun behavioral three factor model (DHS3), and Stambaugh-Yuan four factor model (SY4). The data of the factor models is taken from the authors' websites. SY4 is available until December 2016, HML(EKP) are available until March 2022, HXZ4, HMXZ5, and DHS3 data are available until December 2022, FF5, FF6, and HML(Devil) are available until June 2023. Cash-based operating profitability factor is self constructed following the method in Fama and French (2018).

Moving from bottom-line to top-line profitability: left to right

Panel A: Annual-rebalanced factors

	r^{IB}_{ROE}	r^{SPI}_{ROE}	r^{EBT}_{ROE}	r^{FF}_{ROE}	r^{RD}_{ROE}	$r^{RD+.3SG}_{ROE}$	r^{RD+SG}_{ROE}	r^{NM}_{ROE}	r^{REVT}_{ROE}
Mean	0.20	0.26	0.30	0.36	0.46	0.47	0.39	0.38	0.28
	1.98	*2.68*	*2.72*	*3.53*	*6.08*	*6.16*	*4.74*	*4.65*	*3.26*
FF5	0.01	0.04	0.06	0.01	0.24	0.24	0.17	0.11	-0.12
	0.20	*0.90*	*1.27*	*0.50*	*6.18*	*5.35*	*2.69*	*1.94*	*-1.78*
FF5(C)	-0.03	0.00	0.02	-0.06	0.18	0.21	0.15	0.11	-0.13
	-0.45	*-0.02*	*0.34*	*-1.25*	*3.74*	*3.68*	*2.12*	*1.58*	*-1.73*
FF6	0.01	0.03	0.05	0.01	0.22	0.25	0.19	0.14	-0.09
	0.19	*0.79*	*1.20*	*0.42*	*5.77*	*5.43*	*3.17*	*2.42*	*-1.35*
FF6/Devil	0.01	0.04	0.04	-0.01	0.28	0.34	0.28	0.22	-0.08
	0.25	*1.04*	*0.91*	*-0.29*	*7.05*	*7.09*	*4.42*	*3.58*	*-1.28*
FF6/EKP	0.03	0.05	0.03	0.01	0.27	0.27	0.20	0.14	-0.16
	0.56	*1.05*	*0.65*	*0.45*	*5.99*	*4.92*	*2.82*	*2.10*	*-2.55*

HXZ4	-0.02	0.03	0.03	0.07	0.31	0.31	0.23	0.17	-0.09
	-0.27	*0.38*	*0.43*	*0.90*	*4.84*	*4.50*	*2.89*	*2.22*	*-1.15*
HMXZ5	-0.06	-0.03	-0.02	0.02	0.14	0.16	0.13	0.07	-0.06
	-0.83	*-0.34*	*-0.27*	*0.26*	*2.24*	*2.22*	*1.50*	*0.87*	*-0.64*
DHS3	-0.03	0.03	0.01	0.09	0.28	0.30	0.23	0.18	-0.01
	-0.37	*0.31*	*0.16*	*1.07*	*3.63*	*3.69*	*2.62*	*2.11*	*-0.09*
SY4	0.12	0.14	0.15	0.14	0.26	0.24	0.15	0.10	-0.10
	1.26	*1.54*	*1.42*	*1.40*	*3.93*	*3.26*	*1.66*	*1.15*	*-0.97*

Panel A-2: Monthly-rebalanced factors

	$r^{\mathrm{IB}}_{\mathrm{ROEQ}}$	$r^{\mathrm{SPI}}_{\mathrm{ROEQ}}$	$r^{\mathrm{EBT}}_{\mathrm{ROEQ}}$	$r^{\mathrm{FF}}_{\mathrm{ROEQ}}$	$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	$r^{\mathrm{RD+.3\,SG}}_{\mathrm{ROEQ}}$	$r^{\mathrm{RD+SG}}_{\mathrm{ROEQ}}$	$r^{\mathrm{NM}}_{\mathrm{ROEQ}}$	$r^{\mathrm{REVT}}_{\mathrm{ROEQ}}$
Mean	0.61	0.64	0.62	0.63	0.76	0.70	0.53	0.53	0.38
	5.08	*5.52*	*5.08*	*5.65*	*8.31*	*8.01*	*6.12*	*6.17*	*3.93*
FF5	0.35	0.36	0.33	0.29	0.55	0.46	0.28	0.25	-0.05
	4.37	*4.81*	*4.24*	*4.54*	*8.41*	*7.58*	*4.39*	*3.95*	*-0.71*
FF5(C)	0.30	0.31	0.28	0.22	0.49	0.42	0.27	0.25	-0.06
	3.21	*3.51*	*3.00*	*2.82*	*6.88*	*5.97*	*3.57*	*3.33*	*-0.69*
FF6	0.21	0.25	0.20	0.19	0.45	0.38	0.26	0.23	-0.05
	3.20	*3.76*	*3.09*	*3.45*	*7.95*	*6.88*	*3.99*	*3.58*	*-0.66*
FF6/Devil	0.26	0.29	0.24	0.21	0.51	0.46	0.34	0.30	-0.07
	4.09	*4.56*	*3.70*	*3.78*	*9.34*	*8.72*	*5.30*	*4.67*	*-0.92*
FF6/EKP	0.26	0.30	0.24	0.23	0.49	0.41	0.28	0.24	-0.13
	3.83	*4.52*	*3.68*	*4.05*	*8.36*	*6.88*	*4.02*	*3.47*	*-1.82*
HXZ4	0.05	0.11	0.06	0.12	0.37	0.33	0.23	0.20	-0.10
	1.34	*2.39*	*1.26*	*2.06*	*7.28*	*5.94*	*3.25*	*2.83*	*-1.21*
HMXZ5	0.07	0.12	0.06	0.11	0.26	0.24	0.16	0.13	-0.05
	1.77	*2.36*	*1.20*	*1.77*	*4.94*	*4.08*	*2.10*	*1.77*	*-0.55*
DHS3	0.21	0.28	0.24	0.28	0.50	0.44	0.31	0.28	0.05
	1.87	*2.52*	*2.04*	*2.63*	*5.25*	*4.71*	*3.33*	*3.13*	*0.57*
SY4	0.31	0.37	0.33	0.36	0.50	0.40	0.23	0.21	-0.04
	3.10	*3.79*	*3.11*	*3.65*	*7.12*	*5.36*	*2.66*	*2.37*	*-0.34*

Panel B-1: Long side of the annual-rebalanced factors

	$r^{\mathrm{IB}}_{\mathrm{ROE}}$	$r^{\mathrm{SPI}}_{\mathrm{ROE}}$	$r^{\mathrm{EBT}}_{\mathrm{ROE}}$	$r^{\mathrm{FF}}_{\mathrm{ROE}}$	$r^{\mathrm{RD}}_{\mathrm{ROE}}$	$r^{\mathrm{RD+.3\,SG}}_{\mathrm{ROE}}$	$r^{\mathrm{RD+SG}}_{\mathrm{ROE}}$	$r^{\mathrm{NM}}_{\mathrm{ROE}}$	$r^{\mathrm{REVT}}_{\mathrm{ROE}}$
Mean	0.83	0.85	0.86	0.89	0.92	0.94	0.92	0.91	0.89
	4.05	*4.08*	*4.31*	*4.35*	*4.30*	*4.39*	*4.34*	*4.23*	*4.07*
FF5	-0.01	-0.01	0.00	-0.02	0.09	0.08	0.05	0.02	-0.07
	-0.45	*-0.47*	*-0.20*	*-0.97*	*4.07*	*3.30*	*1.41*	*0.48*	*-1.69*
FF5(C)	-0.02	-0.02	-0.01	-0.04	0.08	0.08	0.05	0.02	-0.08
	-0.58	*-0.87*	*-0.45*	*-1.63*	*3.19*	*2.71*	*1.36*	*0.61*	*-1.68*
FF6	0.01	0.01	0.01	-0.01	0.09	0.09	0.06	0.03	-0.05
	0.36	*0.28*	*0.59*	*-0.43*	*4.08*	*3.52*	*1.78*	*0.90*	*-1.18*
FF6/Devil	0.01	0.01	0.00	-0.03	0.11	0.12	0.09	0.05	-0.07
	0.38	*0.41*	*-0.23*	*-1.68*	*4.66*	*4.53*	*2.42*	*1.44*	*-1.65*
FF6/EKP	0.00	0.00	-0.01	-0.03	0.10	0.09	0.06	0.02	-0.11
	0.13	*0.21*	*-0.35*	*-1.31*	*3.99*	*3.26*	*1.49*	*0.60*	*-2.74*
HXZ4	0.02	0.02	0.02	0.03	0.14	0.13	0.09	0.06	-0.04
	0.54	*0.64*	*0.55*	*0.63*	*4.51*	*3.78*	*2.15*	*1.40*	*-0.75*

HMXZ5	0.03	0.04	0.03	0.05	0.11	0.10	0.10	0.07	0.02
	0.85	*0.92*	*0.83*	*1.00*	*3.29*	*2.67*	*2.04*	*1.53*	*0.32*
DHS3	0.14	0.15	0.14	0.16	0.26	0.27	0.24	0.21	0.14
	2.61	*2.77*	*2.68*	*2.82*	*4.43*	*4.26*	*3.50*	*3.12*	*1.79*
SY4	0.01	0.02	0.02	0.02	0.07	0.05	0.01	-0.02	-0.09
	0.33	*0.48*	*0.51*	*0.46*	*2.29*	*1.36*	*0.21*	*-0.32*	*-1.50*

Panel B-2: Long side of the monthly-rebalanced factors

	$r^{\text{IB}}_{\text{ROEQ}}$	$r^{\text{SPI}}_{\text{ROEQ}}$	$r^{\text{EBT}}_{\text{ROEQ}}$	$r^{\text{FF}}_{\text{ROEQ}}$	$r^{\text{RD}}_{\text{ROEQ}}$	$r^{\text{RD+.3 SG}}_{\text{ROEQ}}$	$r^{\text{RD+SG}}_{\text{ROEQ}}$	$r^{\text{NM}}_{\text{ROEQ}}$	$r^{\text{REVT}}_{\text{ROEQ}}$
Mean	1.00	1.02	1.02	1.02	1.05	1.06	0.99	1.00	0.94
	4.75	*4.80*	*4.83*	*4.78*	*4.81*	*4.85*	*4.58*	*4.57*	*4.22*
FF5	0.14	0.15	0.14	0.12	0.23	0.22	0.14	0.13	-0.01
	4.53	*5.06*	*4.32*	*3.66*	*7.19*	*6.90*	*3.52*	*3.32*	*-0.25*
FF5(C)	0.13	0.14	0.12	0.09	0.20	0.21	0.14	0.14	-0.02
	3.63	*4.11*	*3.32*	*2.49*	*6.33*	*6.18*	*3.41*	*3.31*	*-0.47*
FF6	0.10	0.12	0.11	0.09	0.19	0.20	0.13	0.13	0.00
	3.60	*4.27*	*3.51*	*2.90*	*6.50*	*6.30*	*3.42*	*3.24*	*0.04*
FF6/Devil	0.12	0.14	0.13	0.10	0.22	0.23	0.16	0.15	-0.01
	4.27	*4.75*	*4.12*	*2.97*	*7.18*	*7.14*	*4.04*	*3.75*	*-0.28*
FF6/EKP	0.11	0.13	0.13	0.10	0.21	0.21	0.14	0.13	-0.05
	3.73	*4.34*	*3.98*	*3.14*	*6.62*	*6.20*	*3.38*	*3.09*	*-1.11*
HXZ4	0.07	0.09	0.09	0.09	0.19	0.20	0.15	0.14	0.00
	2.53	*3.32*	*2.65*	*2.33*	*5.92*	*5.82*	*3.48*	*3.26*	*0.03*
HMXZ5	0.09	0.11	0.11	0.12	0.18	0.19	0.15	0.14	0.06
	3.24	*3.54*	*3.20*	*2.99*	*5.08*	*5.09*	*3.22*	*3.17*	*1.03*
DHS3	0.25	0.28	0.27	0.27	0.36	0.36	0.32	0.31	0.20
	4.14	*4.51*	*4.29*	*4.25*	*5.87*	*5.57*	*4.53*	*4.43*	*2.47*
SY4	0.11	0.13	0.12	0.13	0.18	0.16	0.08	0.07	-0.03
	2.81	*3.39*	*2.99*	*2.83*	*4.99*	*4.35*	*1.60*	*1.56*	*-0.49*

Panel C-1: Short side of the annual-rebalanced factors

	$r^{\text{IB}}_{\text{ROE}}$	$r^{\text{SPI}}_{\text{ROE}}$	$r^{\text{EBT}}_{\text{ROE}}$	$r^{\text{FF}}_{\text{ROE}}$	$r^{\text{RD}}_{\text{ROE}}$	$r^{\text{RD+.3 SG}}_{\text{ROE}}$	$r^{\text{RD+SG}}_{\text{ROE}}$	$r^{\text{NM}}_{\text{ROE}}$	$r^{\text{REVT}}_{\text{ROE}}$
Mean	0.63	0.58	0.56	0.53	0.46	0.46	0.53	0.53	0.61
	2.51	*2.35*	*2.23*	*2.17*	*1.97*	*2.10*	*2.48*	*2.50*	*2.91*
FF5	-0.02	-0.05	-0.06	-0.03	-0.15	-0.16	-0.12	-0.10	0.05
	-0.46	*-1.26*	*-1.60*	*-1.37*	*-5.01*	*-5.27*	*-3.11*	*-2.83*	*1.46*
FF5(C)	0.01	-0.02	-0.04	0.02	-0.10	-0.13	-0.10	-0.09	0.06
	0.26	*-0.46*	*-0.73*	*0.47*	*-2.89*	*-3.58*	*-2.34*	*-2.18*	*1.47*
FF6	0.00	-0.03	-0.04	-0.02	-0.13	-0.16	-0.13	-0.11	0.04
	-0.02	*-0.75*	*-1.13*	*-0.79*	*-4.47*	*-5.20*	*-3.52*	*-3.20*	*1.22*
FF6/Devil	0.00	-0.04	-0.05	-0.03	-0.17	-0.22	-0.19	-0.17	0.02
	-0.09	*-0.96*	*-1.23*	*-1.29*	*-5.87*	*-6.95*	*-5.01*	*-4.63*	*0.49*
FF6/EKP	-0.02	-0.04	-0.04	-0.04	-0.17	-0.18	-0.14	-0.12	0.05
	-0.58	*-1.09*	*-0.98*	*-1.76*	*-5.25*	*-5.02*	*-3.26*	*-2.92*	*1.67*
HXZ4	0.04	0.00	-0.01	-0.04	-0.16	-0.18	-0.14	-0.11	0.06
	0.77	*-0.08*	*-0.24*	*-0.93*	*-3.78*	*-4.10*	*-2.92*	*-2.51*	*1.44*
HMXZ5	0.10	0.06	0.06	0.02	-0.04	-0.06	-0.03	0.00	0.07
	1.79	*1.14*	*1.03*	*0.51*	*-0.80*	*-1.33*	*-0.65*	*0.01*	*1.81*

DHS3	0.17	0.13	0.13	0.07	-0.03	-0.04	0.00	0.03	0.15
	1.89	*1.42*	*1.46*	*0.89*	*-0.33*	*-0.48*	*0.04*	*0.37*	*2.35*
SY4	-0.11	-0.12	-0.12	-0.11	-0.18	-0.19	-0.14	-0.12	0.00
	-1.53	*-1.81*	*-1.72*	*-1.81*	*-3.53*	*-3.65*	*-2.34*	*-2.09*	*0.10*

Panel C-2: Short side of the monthly-rebalanced factors

	$r^{\mathrm{IB}}_{\mathrm{ROEQ}}$	$r^{\mathrm{SPI}}_{\mathrm{ROEQ}}$	$r^{\mathrm{EBT}}_{\mathrm{ROEQ}}$	$r^{\mathrm{FF}}_{\mathrm{ROEQ}}$	$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	$r^{\mathrm{RD+.3\,SG}}_{\mathrm{ROEQ}}$	$r^{\mathrm{RD+SG}}_{\mathrm{ROEQ}}$	$r^{\mathrm{NM}}_{\mathrm{ROEQ}}$	$r^{\mathrm{REVT}}_{\mathrm{ROEQ}}$
Mean	0.39	0.38	0.39	0.38	0.29	0.36	0.46	0.47	0.56
	1.52	*1.50*	*1.53*	*1.55*	*1.23*	*1.58*	*2.15*	*2.21*	*2.59*
FF5	-0.21	-0.21	-0.18	-0.17	-0.33	-0.24	-0.15	-0.13	0.04
	-3.60	*-3.76*	*-3.22*	*-3.79*	*-7.05*	*-5.79*	*-3.65*	*-3.16*	*1.02*
FF5(C)	-0.17	-0.17	-0.15	-0.13	-0.29	-0.21	-0.12	-0.11	0.03
	-2.58	*-2.66*	*-2.30*	*-2.38*	*-5.60*	*-4.37*	*-2.70*	*-2.45*	*0.73*
FF6	-0.11	-0.12	-0.10	-0.10	-0.25	-0.18	-0.12	-0.10	0.05
	-2.26	*-2.55*	*-1.92*	*-2.58*	*-6.36*	*-4.92*	*-3.11*	*-2.64*	*1.29*
FF6/Devil	-0.14	-0.16	-0.11	-0.12	-0.29	-0.24	-0.18	-0.15	0.05
	-3.02	*-3.29*	*-2.27*	*-2.99*	*-7.67*	*-6.75*	*-4.66*	*-3.95*	*1.38*
FF6/EKP	-0.15	-0.17	-0.12	-0.13	-0.28	-0.21	-0.14	-0.11	0.08
	-3.03	*-3.52*	*-2.36*	*-3.19*	*-6.88*	*-5.19*	*-3.34*	*-2.71*	*2.02*
HXZ4	0.01	-0.02	0.02	-0.03	-0.18	-0.13	-0.08	-0.06	0.11
	0.45	*-0.46*	*0.58*	*-0.82*	*-5.06*	*-3.57*	*-2.00*	*-1.52*	*2.38*
HMXZ5	0.02	-0.01	0.05	0.01	-0.09	-0.05	-0.01	0.01	0.11
	0.49	*-0.24*	*1.10*	*0.24*	*-2.37*	*-1.35*	*-0.23*	*0.25*	*2.31*
DHS3	0.04	-0.01	0.03	-0.01	-0.14	-0.07	0.01	0.03	0.15
	0.35	*-0.06*	*0.31*	*-0.06*	*-1.46*	*-0.83*	*0.09*	*0.32*	*2.12*
SY4	-0.20	-0.24	-0.20	-0.23	-0.32	-0.23	-0.16	-0.14	0.01
	-2.62	*-3.16*	*-2.49*	*-3.19*	*-5.31*	*-3.91*	*-2.53*	*-2.18*	*0.11*

Table 7: Pricing tests on annual-rebalanced profitability factors varying the fraction of SGA (July 1976 - June 2023).

To construct annual-rebalanced profitability factors, in the June end of each year t, firms are assigned into two size groups based on median June end market equity of NYSE firms, and assigned independently into three profitability groups based on the 30 and 70 percentiles profitability of all NYSE firms reported annual earnings in calendar year t-1. The profitability factor return in any given month is the average return of the portfolios that are long the high profitability portfolios and short the low profitability portfolios within each size group. We consider the profitability measures of $Y(\gamma)/B$ and $Y(\gamma) = (\text{REVT}-\text{COGS}-\text{XSGA}-\text{XINT}+\frac{\gamma}{10}\text{SGA}+\text{XRD})$ where $\gamma = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. We run time-series regressions of the constructed profitability factors on the existing factor models and report their intercept and associated t-statistics in italic. The factor models include Fama-French five factor model (FF5), FF5 with RMW replaced by the cash-based operating profitability factor (FF5(C)), FF5 plus momentum factor (FF6), FF6 with the value factor, HML, replaced by the monthly-rebalanced value factor with most recent market equity, HML(Devil), from Asness and Frazzini (2013), FF6 with the value factor, HML, replaced by the intangible-adjusted value factor, HML(EKP), from Eisfeldt et al. (2022), Hou-Xue-Zhang four factor model (HXZ4) and Hou-Mo-Xue-Zhang five factor model (HMXZ5), Daniel-Hirshleifer-Sun behavioral three factor model (DHS3), and Stambaugh-Yuan four factor model (SY4). The data of the factor models is taken from the authors' websites. Except for SY4 which is available only until December 2016, HML(EKP) are available until March 2022, HXZ4, HMXZ5, and DHS3 data are available until December 2022, FF5, FF6, and HML(Devil) are available until June 2023. Cash-based operating profitability factor is self constructed following the method in Fama and French (2018).

	0% ($r^{\text{RD}}_{\text{ROE}}$)	10%	20%	30% ($r^{\text{RD}+.3\,\text{SG}}_{\text{ROE}}$)	40%	50%	60%	70%	80%	90%	100% ($\approx r^{\text{RD+SG}}_{\text{ROE}}$)
Mean	0.46	0.47	0.47	0.47	0.46	0.46	0.43	0.44	0.43	0.42	0.40
	6.08	*6.28*	*6.32*	*6.16*	*5.90*	*5.76*	*5.41*	*5.39*	*5.22*	*5.06*	*4.87*
FF5	0.24	0.24	0.25	0.24	0.22	0.22	0.20	0.20	0.20	0.19	0.18
	6.18	*6.09*	*5.96*	*5.35*	*4.67*	*4.36*	*3.74*	*3.67*	*3.44*	*3.16*	*2.93*
FF5(C)	0.18	0.19	0.21	0.21	0.19	0.20	0.18	0.19	0.18	0.17	0.17
	3.74	*3.79*	*3.92*	*3.68*	*3.21*	*3.12*	*2.72*	*2.78*	*2.65*	*2.50*	*2.35*
FF6	0.22	0.23	0.25	0.25	0.23	0.24	0.21	0.22	0.22	0.21	0.21
	5.77	*5.87*	*5.93*	*5.43*	*4.83*	*4.61*	*4.01*	*4.01*	*3.78*	*3.53*	*3.38*
FF6/Devil	0.28	0.30	0.33	0.34	0.32	0.33	0.30	0.31	0.31	0.30	0.30
	7.05	*7.42*	*7.54*	*7.09*	*6.43*	*6.18*	*5.55*	*5.43*	*5.16*	*4.92*	*4.69*
FF6/EKP	0.27	0.27	0.28	0.27	0.26	0.25	0.23	0.24	0.23	0.22	0.21
	5.99	*5.60*	*5.46*	*4.92*	*4.40*	*4.17*	*3.65*	*3.65*	*3.44*	*3.21*	*3.03*
HXZ4	0.31	0.31	0.32	0.31	0.29	0.28	0.26	0.26	0.26	0.25	0.24
	4.84	*4.77*	*4.83*	*4.50*	*4.04*	*3.89*	*3.48*	*3.50*	*3.37*	*3.17*	*3.09*
HMXZ5	0.14	0.15	0.17	0.16	0.15	0.16	0.13	0.15	0.15	0.13	0.13
	2.24	*2.28*	*2.48*	*2.22*	*2.01*	*2.04*	*1.73*	*1.83*	*1.81*	*1.64*	*1.61*
DHS3	0.28	0.30	0.31	0.30	0.29	0.28	0.26	0.27	0.26	0.25	0.25
	3.63	*3.76*	*3.84*	*3.69*	*3.39*	*3.31*	*3.05*	*3.11*	*2.96*	*2.85*	*2.77*
SY4	0.26	0.26	0.26	0.24	0.21	0.21	0.18	0.19	0.18	0.16	0.16
	3.93	*3.79*	*3.75*	*3.26*	*2.74*	*2.63*	*2.28*	*2.23*	*2.07*	*1.88*	*1.79*

Table 8: Mean and t-statistics of the monthly factor returns for the full sample and two subsamples (July 1976 - June 2023).

RMW, robust-minus-weak, is the annual-rebalanced operating profitability factor in Fama-French five factor model. $r^{\text{FF}}_{\text{CROE}}$ is self-constructed cash based operating profitability factor based on Fama and French (2018). $r^{\text{NM}}_{\text{ROA}}$ is self-constructed gross profitability factor based on Novy-Marx (2013). $r^{\text{BGLN}}_{\text{ROA}}$ is self-constructed operating profitability factor based on Ball et al. (2015). $r^{\text{BGLN}}_{\text{CROA}}$ is self-constructed cash based operating profitability factor based on Ball et al. (2016). R_ROE is the monthly-rebalnacing profitability factor in Hou-Xue-Zhang four factor model. PERF is the performance factor from Stambaugh-Yuan four factor model (SY4). UMD is the momentum factor taken from Kenneth French's data library. QMJ is the monthly-rebalanced quality-minus-junk factor by Asness et al. (2019). PEAD is the post earnings announcement drift factor from Daniel-Hirshleifer-Sun behavioral three factor model. HML(EKP) is the intangible-adjusted value factor by Eisfeldt et al. (2022). $r^{\text{RD}}_{\text{ROE}}$ and $r^{\text{RD}}_{\text{ROEQ}}$ are self constructed annual- and monthly-rebalanced profitability factors based on the profitability measures $Y(\text{RD})/B$ where $Y(\text{RD}) = \text{REVT} - \text{COGS} - \text{XSGA} - \text{XINT} + \text{XRD}$ for annual-rebalanced factor and $Y(\text{RD}) = \text{REVTQ} - \text{COGSQ} - \text{XSGAQ} - \text{XINTQ} + \text{XRDQ}$ for monthly-rebalanced factor. RMW, R_ROE, PEAD, PERF, UMD, QMJ are taken from the authors' websites. PERF is available until December 2016, HML(EKP) is available until March 2022, R_ROE and PEAD are available until December 2022, RMW, UMD, and QMJ are available until June 2023.

	RMW	$r^{\text{FF}}_{\text{CROE}}$	$r^{\text{NM}}_{\text{ROA}}$	$r^{\text{BGLN}}_{\text{ROA}}$	$r^{\text{BGLN}}_{\text{CROA}}$	R_ROE	PERF	UMD	QMJ	PEAD	HML(EKP)	$r^{\text{RD}}_{\text{ROE}}$	$r^{\text{RD}}_{\text{ROEQ}}$
	Full Sample: July 1976 to June 2023												
Mean	0.36	0.37	0.27	0.33	0.38	0.56	0.68	0.57	0.43	0.54	0.45	0.46	0.76
t-stat	*3.69*	*4.21*	*2.77*	*3.56*	*4.60*	*5.05*	*3.79*	*3.10*	*4.28*	*6.71*	*5.33*	*6.08*	*8.31*
	July 1976 to December 1999												
Mean	0.27	0.35	0.18	0.22	0.35	0.77	0.64	1.00	0.41	0.79	0.45	0.34	0.79
t-stat	*2.92*	*3.84*	*1.32*	*1.84*	*3.51*	*6.62*	*3.67*	*5.07*	*4.16*	*8.51*	*4.58*	*3.92*	*7.21*
	January 2000 to June 2023												
Mean	0.46	0.40	0.36	0.43	0.42	0.34	0.74	0.15	0.45	0.29	0.45	0.58	0.72
t-stat	*2.65*	*2.62*	*2.57*	*3.11*	*3.12*	*1.83*	*2.08*	*0.47*	*2.57*	*2.20*	*3.23*	*4.69*	*4.96*

Table 9: Pricing tests of $r^{\text{RD}}_{\text{ROE}}$ on related factors and anomalies and the pricing tests of related factors and anomalies on $r^{\text{RD}}_{\text{ROE}}$ (July 1976 - June 2023).

RMW, robust-minus-weak, is the annual-rebalanced operating profitability factor in Fama-French five factor model. $r^{\text{FF}}_{\text{CROE}}$ is self-constructed cash based operating profitability factor based on Fama and French (2018). $r^{\text{NM}}_{\text{ROA}}$ is self-constructed gross profitability factor based on Novy-Marx (2013). $r^{\text{BGLN}}_{\text{ROA}}$ is self-constructed operating profitability factor based on Ball et al. (2015). $r^{\text{BGLN}}_{\text{CROA}}$ is self-constructed cash based operating profitability factor based on Ball et al. (2016). R_ROE is the monthly-rebalnacing profitability factor in Hou-Xue-Zhang four factor model. PERF is the performance factor from Stambaugh-Yuan four factor model (SY4). UMD is the momentum factor taken from Kenneth French's data library. QMJ is the monthly-rebalanced quality-minus-junk factor by Asness et al. (2019). PEAD is the post earnings announcement drift factor from Daniel-Hirshleifer-Sun behavioral three factor model. HML(EKP) is the intangible-adjusted value factor by Eisfeldt et al. (2022). $r^{\text{RD}}_{\text{ROE}}$ and $r^{\text{RD}}_{\text{ROEQ}}$ are self constructed annual- and monthly-rebalanced profitability factors based on the profitability measures $Y(\text{RD})/B$ where $Y(\text{RD}) = \text{REVT} - \text{COGS} - \text{XSGA} - \text{XINT} + \text{XRD}$ for annual-rebalanced factor and $Y(\text{RD}) = \text{REVTQ} - \text{COGSQ} - \text{XSGAQ} - \text{XINTQ} + \text{XRDQ}$ for monthly-rebalanced factor. RMW, R_ROE, PEAD, PERF, UMD, QMJ are taken from the authors' websites. PERF is available until December 2016, HML(EKP) is available until March 2022, R_ROE and PEAD are available until December 2022, RMW, UMD, and QMJ are available until June 2023. The table reports the intercept and slope coefficient with the associated t-statistics and (adjusted) r-squared of the univariate and multiple time-series regressions. % drop in Table B-2 reports the percentage drop of alpha in adjusted FF5 relative to FF5.

	RMW	$r^{\text{FF}}_{\text{CROE}}$	$r^{\text{NM}}_{\text{ROA}}$	$r^{\text{BGLN}}_{\text{ROA}}$	$r^{\text{BGLN}}_{\text{CROA}}$	R_ROE	PERF	UMD	QMJ	PEAD	HML(EKP)
Panel A-1: $X = a + b\,\text{RMW} + \epsilon$											
a		0.10	0.21	0.22	0.30	0.28	0.43	0.52	0.17	0.57	0.42
		2.00	*2.17*	*2.51*	*3.73*	*3.46*	*2.65*	*2.78*	*2.33*	*6.94*	*4.91*
b		0.77	0.16	0.29	0.22	0.76	0.71	0.15	0.72	-0.07	0.09
		37.99	*3.85*	*7.70*	*6.29*	*21.88*	*10.37*	*1.87*	*23.63*	*-1.89*	*2.62*
R^2		0.72	0.03	0.10	0.07	0.46	0.18	0.01	0.50	0.01	0.01
Panel A-2: $X = a + b_1\text{RMRF} + b_2\text{SMB} + b_3\text{HML} + b_4\text{CMA} + b_5\text{RMW} + \epsilon$											
a		0.18	0.19	0.23	0.33	0.37	0.60	0.55	0.37	0.63	0.20
		4.30	*2.79*	*3.68*	*5.34*	*4.56*	*4.26*	*3.03*	*6.12*	*7.64*	*4.09*
b_1		-0.08	0.02	-0.02	-0.05	-0.05	-0.20	-0.15	-0.20	-0.07	0.04
		-8.65	*1.06*	*-1.44*	*-3.10*	*-2.61*	*-5.85*	*-3.44*	*-14.09*	*-3.49*	*3.36*
b_2		-0.10	0.11	0.07	0.03	-0.09	0.11	0.12	-0.10	0.01	0.19
		-6.96	*4.31*	*3.11*	*1.34*	*-3.12*	*2.21*	*1.77*	*-4.56*	*0.30*	*11.05*
b_3		0.05	-0.61	-0.58	-0.52	-0.20	-0.89	-0.60	-0.19	-0.18	0.41
		2.68	*-20.27*	*-21.10*	*-19.64*	*-5.70*	*-13.70*	*-7.71*	*-7.31*	*-5.01*	*19.56*
b_4		0.10	0.13	0.19	0.24	0.07	0.56	0.49	0.10	0.06	0.26
		3.61	*2.91*	*4.40*	*5.91*	*1.29*	*5.56*	*4.05*	*2.51*	*1.11*	*7.90*
b_5		0.65	0.38	0.46	0.33	0.74	0.85	0.26	0.63	-0.05	0.08
		35.02	*11.78*	*15.62*	*11.64*	*19.94*	*13.15*	*3.09*	*23.00*	*-1.24*	*3.45*
Adj. R^2		0.81	0.54	0.57	0.49	0.51	0.45	0.12	0.68	0.07	0.71
Panel B-1: $X = a + b\,r^{\text{RD}}_{\text{ROE}} + \epsilon$											
a	-0.11	0.02	-0.03	-0.06	0.04	0.16	0.08	0.32	0.01	0.51	0.51
	-1.81	*0.25*	*-0.33*	*-0.92*	*0.69*	*1.72*	*0.57*	*1.72*	*0.11*	*6.12*	*5.95*
b	1.03	0.77	0.65	0.85	0.73	0.88	1.46	0.55	0.92	0.07	-0.14
	30.97	*20.82*	*13.80*	*23.03*	*21.30*	*17.97*	*18.02*	*5.47*	*22.62*	*1.63*	*-3.03*

R^2	0.63	0.44	0.25	0.49	0.45	0.37	0.40	0.05	0.48	0.00	0.02
Panel B-2: $X = a + b_1\text{RMRF} + b_2\text{SMB} + b_3\text{HML} + b_4\text{CMA} + b_5 r^{\text{RD}}_{\text{ROE}} + \epsilon$											
a	-0.13	0.06	0.05	0.02	0.13	0.27	0.31	0.43	0.22	0.60	0.16
	-2.66	*1.23*	*0.79*	*0.41*	*2.50*	*2.93*	*2.27*	*2.29*	*3.55*	*7.09*	*3.29*
b_1	-0.01	-0.09	0.02	-0.01	-0.03	-0.06	-0.18	-0.14	-0.20	-0.06	0.04
	-0.89	*-7.86*	*1.39*	*-1.07*	*-2.95*	*-2.80*	*-5.58*	*-3.34*	*-14.29*	*-3.30*	*3.54*
b_2	-0.12	-0.17	0.09	0.06	0.04	-0.18	0.05	0.11	-0.16	0.02	0.19
	-7.21	*-10.54*	*3.97*	*3.51*	*2.16*	*-5.85*	*1.17*	*1.81*	*-7.50*	*0.86*	*11.43*
b_3	0.31	0.26	-0.47	-0.40	-0.38	0.03	-0.55	-0.50	0.03	-0.18	0.44
	15.27	*12.88*	*-16.67*	*-18.20*	*-17.82*	*0.89*	*-8.82*	*-6.47*	*1.02*	*-5.22*	*21.28*
b_4	-0.11	0.02	0.09	0.14	0.21	-0.01	0.41	0.46	0.03	0.07	0.25
	-3.54	*0.78*	*2.14*	*4.00*	*6.22*	*-0.24*	*4.37*	*3.86*	*0.74*	*1.22*	*7.77*
b_5	1.06	0.75	0.56	0.76	0.64	0.79	1.26	0.45	0.79	0.01	0.13
	38.83	*28.00*	*15.18*	*26.15*	*22.42*	*15.73*	*15.82*	*4.33*	*23.55*	*0.30*	*4.59*
Adj. R^2	0.79	0.75	0.60	0.72	0.67	0.42	0.51	0.13	0.69	0.07	0.71
% drop		0.67	0.74	0.91	0.61	0.27	0.48	0.22	0.41	0.05	0.20
Panel C-1: $r^{\text{RD}}_{\text{ROE}} = a + b\,X + \epsilon$											
a	0.24	0.25	0.36	0.27	0.23	0.22	0.22	0.41	0.24	0.42	0.50
	5.12	*4.33*	*5.39*	*4.97*	*3.97*	*3.60*	*3.64*	*5.48*	*4.25*	*5.33*	*6.48*
b	0.61	0.56	0.39	0.57	0.61	0.42	0.28	0.09	0.52	0.06	-0.12
	30.97	*20.82*	*13.80*	*23.03*	*21.30*	*17.97*	*18.02*	*5.47*	*22.62*	*1.63*	*-3.03*
R^2	0.63	0.44	0.25	0.49	0.45	0.37	0.40	0.05	0.48	0.00	0.02
Panel C-2: $r^{\text{RD}}_{\text{ROE}} = a + b_1\text{RMRF} + b_2\text{SMB} + b_3\text{HML} + b_4\text{CMA} + b_5\text{RMW} + b_6 X + \epsilon$											
a		0.21	0.20	0.15	0.10	0.24	0.18	0.22	0.17	0.21	0.23
		5.36	*5.52*	*4.91*	*3.42*	*6.10*	*4.70*	*5.77*	*4.42*	*5.16*	*5.70*
b_1		0.01	-0.01	0.00	0.01	0.00	0.00	0.00	0.03	0.00	-0.01
		0.97	*-1.08*	*0.31*	*1.87*	*-0.54*	*0.11*	*-0.19*	*3.15*	*-0.17*	*-1.33*
b_2		0.06	0.02	0.01	0.03	0.04	0.03	0.04	0.06	0.04	0.02
		3.99	*1.33*	*1.02*	*2.48*	*2.83*	*2.04*	*2.52*	*4.21*	*2.82*	*1.23*
b_3		-0.26	-0.14	-0.03	-0.03	-0.25	-0.18	-0.23	-0.21	-0.24	-0.29
		-15.78	*-6.57*	*-1.81*	*-2.05*	*-14.50*	*-8.76*	*-13.46*	*-12.91*	*-14.17*	*-13.41*
b_4		0.06	0.05	0.01	-0.02	0.08	0.06	0.06	0.06	0.07	0.04
		2.32	*2.11*	*0.31*	*-1.22*	*2.90*	*2.08*	*2.44*	*2.35*	*2.80*	*1.56*
b_5		0.57	0.62	0.52	0.55	0.69	0.57	0.68	0.57	0.69	0.67
		18.26	*33.11*	*31.03*	*37.36*	*29.47*	*27.57*	*38.40*	*24.00*	*39.11*	*37.08*
b_6		0.18	0.19	0.38	0.41	0.00	0.09	0.03	0.19	0.05	0.11
		4.64	*8.61*	*18.90*	*21.11*	*0.16*	*7.61*	*3.22*	*7.30*	*2.64*	*3.09*
Adj. R^2		0.77	0.79	0.86	0.87	0.76	0.78	0.77	0.78	0.77	0.76

Table 10: Summary statistics and performance of within-industry profitability factors (July 1976 - June 2023).

Firms are assigned into industry groups following Fama-French five industry definition. In Panel A, N(firms) is the time-series average of the number of firms, RD>0 is the time-series average fraction of firms that report positive R&D expense, RD/B is the time-series average total R&D expense divided by total book equity, RD/REVT is the time-series average total R&D expense divided by total revenue. In Panel B, we form within-industry profitability factors based on various profitability measures including Y(IB)/B, Y(SPI)/B, Y(EBT)/B, Y(FF)/B, Y(RD)/B, Y(RD + .3 SG)/B, Y(RD+SG)/B, Y(NM)/B, and Y(REVT)/B from the bottom-line to the top-line earnings. Specifically, in the June end of each year t, firms within each industry are assigned into two size (ME) groups based on median June end market equity of NYSE firms within the given industry, and assigned independently into three profitability groups based on the 30 and 70 percentiles profitability of all NYSE firms within the given industry reported annual earnings in calendar year t-1. The profitability factor return within each industry in any given month is the average return of the portfolios that are long the high profitability portfolios and short the low profitability portfolios within each size group.

Panel A: Summary statistics

	N(firms)	RD > 0	RD/B	RD/REVT
Cnsmr	778	0.21	0.04	0.01
Manuf	905	0.42	0.02	0.01
HiTec	719	0.69	0.07	0.04
Hlth	381	0.75	0.12	0.08
Other	1102	0.08	0.00	0.00

Panel B: Profitability factors formed on within-industry Y(X)/B

Monthly Percentage Returns and Associated *t-statistics*

	IB	SPI	EBT	FF	RD	RD+.3SG	RD+SG	NM	REVT
Cnsmr	0.16	0.10	0.26	0.39	0.46	0.46	0.40	0.37	0.30
	1.24	*0.74*	*1.95*	*3.70*	*4.80*	*5.18*	*3.85*	*3.58*	*3.22*
Manuf	0.13	0.19	0.29	0.31	0.36	0.38	0.35	0.34	0.36
	1.02	*1.53*	*2.32*	*2.91*	*3.42*	*3.36*	*2.69*	*2.69*	*3.01*
HiTec	0.14	0.14	0.07	0.23	0.41	0.30	0.28	0.23	0.25
	0.95	*0.97*	*0.45*	*1.44*	*3.32*	*2.66*	*2.59*	*2.22*	*2.11*
Hlth	-0.05	0.06	0.04	0.12	0.36	0.49	0.41	0.44	0.19
	-0.25	*0.30*	*0.21*	*0.58*	*2.16*	*2.87*	*2.11*	*2.33*	*1.03*
Other	0.33	0.34	0.38	0.37	0.39	0.36	0.27	0.27	0.22
	3.34	*3.55*	*3.47*	*4.24*	*4.63*	*4.37*	*3.31*	*3.10*	*1.97*

Table 11: Performance of profitability factors with various denominators (July 1976 - June 2023).

In the June end of each year t, firms are assigned into two size (ME) groups based on median June end market equity of NYSE firms, and assigned independently into three profitability groups based on the 30 and 70 percentiles profitability of all NYSE firms reported annual earnings in calendar year t-1. The profitability factor return in any given month is the average return of the portfolios that are long the high profitability portfolios and short the low profitability portfolios within each size group. The profitability is measured using various numerators and various denominators. The included numerators are Y(IB), Y(SPI), Y(EBT), Y(FF), Y(RD), Y(RD + .3 SG), Y(RD + SG), Y(NM), and Y(REVT) from the bottom-line to the top-line earnings. The included denominators are total revenue (REVT), book equity (B), total assets (AT), intangible-adjusted book value by Peters and Taylor (2017) (B^{PT}), and market equity (ME). The intangible-adjusted book value is the physical capital plus the intangible capital where the physical capital is measured as the gross property, plant, and equipment and the intangible is measured by capitalizing 100% of R&D expenses and 30% of the SG&A expenses plus the balance sheet intangibles. The data on B^{PT} is taken from Peters and Taylor Total Q on Wharton Research Data Services which is available from June 1950 to May 2022 at the time we access the data. In each panel, we report the time-series average of the monthly percentage returns, Fama-French five factor alpha, and Hou-Xue-Zhang four factor alpha and the associated t-statistics (in italic) of each factor.

Moving from bottom-line to top-line profitability: left to right									
Panel A: Y(X)/REVT									
	IB	SPI	EBT	FF	RD	RD+.3 SG	RD+SG	NM	REVT
Mean	0.02	0.05	0.09	0.07	0.18	0.17	0.08	0.02	
	0.15	*0.50*	*0.84*	*0.68*	*2.28*	*2.42*	*1.18*	*0.34*	
FF5	-0.02	0.02	0.03	0.01	0.26	0.30	0.22	0.18	
	-0.27	*0.37*	*0.51*	*0.18*	*3.58*	*4.70*	*3.84*	*3.19*	
HXZ4	-0.09	-0.03	-0.02	-0.02	0.26	0.31	0.24	0.18	
	-1.18	*-0.46*	*-0.31*	*-0.27*	*3.42*	*4.68*	*3.87*	*2.98*	
Panel B: Y(X)/B									
	IB	SPI	EBT	FF	RD	RD+.3 SG	RD+SG	NM	REVT
Mean	0.22	0.28	0.32	0.38	0.47	0.48	0.40	0.38	0.28
	2.12	*2.81*	*2.88*	*3.66*	*6.24*	*6.30*	*4.83*	*4.72*	*3.29*
FF5	0.01	0.04	0.06	0.01	0.25	0.25	0.17	0.12	-0.11
	0.19	*0.89*	*1.34*	*0.63*	*6.41*	*5.58*	*2.79*	*2.02*	*-1.75*
HXZ4	-0.02	0.03	0.03	0.07	0.31	0.31	0.23	0.17	-0.09
	-0.27	*0.38*	*0.43*	*0.90*	*4.84*	*4.50*	*2.89*	*2.22*	*-1.15*
Panel B: Y(X)/AT									
	IB	SPI	EBT	FF	RD	RD+.3 SG	RD+SG	NM	REVT
Mean	0.14	0.18	0.23	0.24	0.33	0.33	0.28	0.28	0.26
	1.43	*1.94*	*2.32*	*2.52*	*3.39*	*3.51*	*2.81*	*2.82*	*3.03*
FF5	0.03	0.07	0.10	0.08	0.26	0.26	0.22	0.20	0.02
	0.46	*1.20*	*1.76*	*1.34*	*3.95*	*4.03*	*3.24*	*2.95*	*0.22*
HXZ4	0.06	0.11	0.13	0.14	0.33	0.31	0.30	0.27	0.04
	0.74	*1.46*	*1.70*	*1.74*	*3.75*	*3.65*	*3.24*	*3.02*	*0.47*
Panel B: Y(X)/B^{PT}									

	IB	SPI	EBT	FF	RD	RD+.3 SG	RD+SG	NM	REVT
Mean	0.15	0.18	0.21	0.22	0.31	0.29	0.25	0.23	0.17
	1.47	*1.87*	*2.07*	*2.28*	*3.48*	*3.00*	*2.40*	*2.14*	*1.74*
FF5	-0.01	0.01	0.04	0.02	0.21	0.22	0.24	0.20	-0.07
	-0.15	*0.12*	*0.69*	*0.32*	*3.04*	*2.74*	*2.61*	*2.16*	*-0.92*
HXZ4	-0.10	-0.08	-0.03	-0.03	0.18	0.16	0.20	0.16	-0.14
	-1.44	*-1.07*	*-0.43*	*-0.43*	*2.39*	*1.88*	*2.02*	*1.60*	*-1.51*

Panel B: Y(X)/ME

	IB	SPI	EBT	FF	RD	RD+.3 SG	RD+SG	NM	REVT
Mean	0.27	0.35	0.32	0.36	0.48	0.47	0.45	0.43	0.38
	1.89	*2.51*	*2.14*	*2.52*	*3.95*	*3.95*	*3.93*	*3.77*	*3.23*
FF5	0.01	0.05	-0.01	-0.01	0.16	0.08	0.04	0.01	-0.09
	0.11	*0.94*	*-0.16*	*-0.14*	*3.42*	*1.64*	*0.68*	*0.23*	*-1.21*
HXZ4	0.01	0.07	0.00	0.03	0.19	0.13	0.09	0.05	-0.06
	0.10	*0.61*	*-0.03*	*0.26*	*1.88*	*1.25*	*0.89*	*0.55*	*-0.57*

Table 12: Model comparison tests.

The R&D-adjusted profitability factors (annual-rebalanced $r^{\mathrm{RD}}_{\mathrm{ROE}}$ and monthly-rebalanced $r^{\mathrm{RD}}_{\mathrm{ROEQ}}$) are incorporated into the existing models in two ways. First, for models with annual- or monthly-rebalanced profitability related factors, the original profitability factor in the model is replaced with the annual- or monthly-rebalanced R&D-adjusted profitability factors respectively. Second, all models are augmented with the R&D-adjusted profitability factors. The squared Sharpe ratio test by Barillas et al. (2020) is used to compare the original models and the models with the R&D-adjusted profitability factors. The first row of each panel is the difference in the squared Sharpe ratios between the column model and the row model, and the second row report the p-value of the test with the null hypothesis that the difference in the squared Sharpe ratio is zero. The factor models include Fama-French five factor model (FF5), FF5 with RMW replaced by the cash-based operating profitability factor (FF5(C)), FF5 plus momentum factor (FF6), FF6 with the value factor, HML, replaced by the monthly-rebalanced value factor with most recent market equity, HML(Devil), from Asness and Frazzini (2013) ($\mathrm{FF6}^1$), FF6 with the value factor, HML, replaced by the intangible-adjusted value factor, HML(EKP), from Eisfeldt et al. (2022) ($\mathrm{FF6}^2$), Hou-Xue-Zhang four factor model (HXZ4) and Hou-Mo-Xue-Zhang five factor model (HMXZ5), Daniel-Hirshleifer-Sun behavioral three factor model (DHS3), and Stambaugh-Yuan four factor model (SY4). The data of the factor models is taken from the authors' websites. SY4 is available until December 2016, HML(EKP) are available until March 2022, HXZ4, HMXZ5, and DHS3 data are available until December 2022, FF5, FF6, and HML(Devil) are available until June 2023. Cash-based operating profitability factor is self constructed following the method in Fama and French (2018).

July 1976 - June 2023			July 1976 - December 1999			January 2000 - June 2023		
	FF5/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF5+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		FF5/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF5+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		FF5/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF5+$r^{\mathrm{RD}}_{\mathrm{ROE}}$
FF5	0.060	0.073	FF5	0.070	0.069	FF5	0.059	0.062
	0.000	*0.000*		*0.023*	*0.000*		*0.010*	*0.000*
	FF5(C)/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF5(C)+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		FF5(C)/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF5(C)+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		FF5(C)/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF5(C)+$r^{\mathrm{RD}}_{\mathrm{ROE}}$
FF5(C)	0.023	0.063	FF5(C)	0.017	0.042	FF5(C)	0.058	0.069
	0.212	*0.000*		*0.473*	*0.007*		*0.036*	*0.000*
	FF6/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF6+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		FF6/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF6+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		FF6/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	FF6+$r^{\mathrm{RD}}_{\mathrm{ROE}}$
FF6	0.053	0.064	FF6	0.060	0.058	FF6	0.058	0.061
	0.000	*0.000*		*0.039*	*0.001*		*0.008*	*0.000*
	$\mathrm{FF6}^1/r^{\mathrm{RD}}_{\mathrm{ROE}}$	$\mathrm{FF6}^1+r^{\mathrm{RD}}_{\mathrm{ROE}}$		$\mathrm{FF6}^1/r^{\mathrm{RD}}_{\mathrm{ROE}}$	$\mathrm{FF6}^1+r^{\mathrm{RD}}_{\mathrm{ROE}}$		$\mathrm{FF6}^1/r^{\mathrm{RD}}_{\mathrm{ROE}}$	$\mathrm{FF6}^1+r^{\mathrm{RD}}_{\mathrm{ROE}}$
$\mathrm{FF6}^1$	0.072	0.098	$\mathrm{FF6}^1$	0.078	0.075	$\mathrm{FF6}^1$	0.075	0.092
	0.000	*0.000*		*0.034*	*0.000*		*0.001*	*0.000*
July 1976 - March 2022			July 1976 - December 1999			January 2000 - March 2022		
	$\mathrm{FF6}^2/r^{\mathrm{RD}}_{\mathrm{ROE}}$	$\mathrm{FF6}^2+r^{\mathrm{RD}}_{\mathrm{ROE}}$		$\mathrm{FF6}^2/r^{\mathrm{RD}}_{\mathrm{ROE}}$	$\mathrm{FF6}^2+r^{\mathrm{RD}}_{\mathrm{ROE}}$		$\mathrm{FF6}^2/r^{\mathrm{RD}}_{\mathrm{ROE}}$	$\mathrm{FF6}^2+r^{\mathrm{RD}}_{\mathrm{ROE}}$
$\mathrm{FF6}^2$	0.059	0.073	$\mathrm{FF6}^2$	0.036	0.032	$\mathrm{FF6}^2$	0.086	0.103
	0.000	*0.000*		*0.112*	*0.007*		*0.001*	*0.000*
July 1976 - December 2022			July 1976 - December 1999			January 2000 - December 2022		
	HXZ4/$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	HXZ4+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$		HXZ4/$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	HXZ4+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$		HXZ4/$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	HXZ4+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$
HXZ4	0.091	0.107	HXZ4	0.114	0.110	HXZ4	0.106	0.145
	0.000	*0.000*		*0.015*	*0.000*		*0.001*	*0.000*
	HMXZ5/$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	HMXZ5+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$		HMXZ5/$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	HMXZ5+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$		HMXZ5/$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$	HMXZ5+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$
HMXZ5	0.030	0.054	HMXZ5	0.040	0.037	HMXZ5	0.053	0.094
	0.044	*0.000*		*0.151*	*0.012*		*0.053*	*0.000*

	DHS3+$r^{\mathrm{RD}}_{\mathrm{ROE}}$	DHS3+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$		DHS3+$r^{\mathrm{RD}}_{\mathrm{ROE}}$	DHS3+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$		DHS3+$r^{\mathrm{RD}}_{\mathrm{ROE}}$	DHS3+$r^{\mathrm{RD}}_{\mathrm{ROEQ}}$
DHS3	0.026	0.057	DHS3	0.046	0.134	DHS3	0.048	0.068
	0.001	*0.000*		*0.003*	*0.000*		*0.000*	*0.000*
July 1976 - December 2016			July 1976 - December 1999			January 2000 - December 2016		
	SY4/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	SY4+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		SY4/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	SY4+$r^{\mathrm{RD}}_{\mathrm{ROE}}$		SY4/$r^{\mathrm{RD}}_{\mathrm{ROE}}$	SY4+$r^{\mathrm{RD}}_{\mathrm{ROE}}$
SY4	0.121	0.126	SY4	0.279	0.274	SY4	0.095	0.096
	0.000	*0.000*		*0.001*	*0.000*		*0.016*	*0.000*

Table 13: Maximum drawdown of the profitability factors (July 1976 - June 2023).

To compute the drawdowns for a factor, we keep tracking the cumulative returns of the factor portfolio and identify the episodes where the cumulative returns are below a historical peak. The episodes are identified independently across factors. The drawdowns are the cumulative percentage returns from the historical peak to the trough within each episode. Panel A and B report the ten maximum drawdowns of each annual- and monthly-rebalanced profitability factors respectively.

Panel A: Annual-rebalanced profitability factors

Rank	$r^{\text{IB}}_{\text{ROE}}$	$r^{\text{SPI}}_{\text{ROE}}$	$r^{\text{EBT}}_{\text{ROE}}$	$r^{\text{FF}}_{\text{ROE}}$	$r^{\text{RD}}_{\text{ROE}}$	$r^{\text{RD+.3 SG}}_{\text{ROE}}$	$r^{\text{RD+SG}}_{\text{ROE}}$	$r^{\text{NM}}_{\text{ROE}}$	$r^{\text{REVT}}_{\text{ROE}}$
1	-36.86	-34.12	-45.52	-42.01	-14.94	-12.22	-20.40	-21.11	-42.13
2	-25.84	-23.90	-27.08	-24.43	-12.11	-12.13	-19.83	-19.40	-14.19
3	-20.59	-18.45	-20.74	-12.96	-10.54	-10.57	-15.74	-13.81	-11.32
4	-18.70	-17.96	-18.32	-12.51	-9.51	-10.49	-12.22	-11.22	-9.71
5	-17.51	-17.24	-15.76	-12.43	-8.29	-9.02	-10.60	-9.97	-9.25
6	-17.43	-11.36	-13.82	-9.97	-6.77	-7.61	-8.75	-9.70	-8.18
7	-12.46	-10.85	-9.47	-6.75	-6.35	-7.46	-7.84	-7.40	-6.94
8	-8.64	-8.58	-9.27	-6.11	-5.96	-6.95	-7.12	-7.18	-5.64
9	-8.56	-7.09	-8.58	-5.68	-5.80	-6.94	-6.42	-6.95	-5.10
10	-8.24	-5.69	-7.61	-5.66	-5.64	-6.91	-6.41	-6.93	-4.74

Panel B: Monthly-rebalanced profitability factors

Rank	$r^{\text{IB}}_{\text{ROEQ}}$	$r^{\text{SPI}}_{\text{ROEQ}}$	$r^{\text{EBT}}_{\text{ROEQ}}$	$r^{\text{FF}}_{\text{ROEQ}}$	$r^{\text{RD}}_{\text{ROEQ}}$	$r^{\text{RD+.3 SG}}_{\text{ROEQ}}$	$r^{\text{RD+SG}}_{\text{ROEQ}}$	$r^{\text{NM}}_{\text{ROEQ}}$	$r^{\text{REVT}}_{\text{ROEQ}}$
1	-33.28	-29.07	-38.86	-36.47	-19.12	-13.78	-21.43	-21.36	-42.50
2	-31.60	-25.92	-30.62	-23.78	-17.13	-13.38	-13.52	-12.04	-13.06
3	-28.06	-24.80	-27.11	-21.43	-12.28	-11.41	-11.34	-9.28	-12.76
4	-25.12	-23.34	-26.51	-20.61	-11.39	-11.31	-10.17	-9.20	-11.63
5	-14.20	-13.67	-13.69	-11.39	-10.59	-9.47	-8.19	-8.51	-10.66
6	-9.22	-11.64	-11.17	-10.66	-7.80	-7.57	-7.28	-8.51	-8.27
7	-9.12	-8.68	-10.76	-10.25	-7.63	-6.70	-7.26	-8.07	-8.12
8	-6.91	-8.68	-8.17	-8.58	-6.97	-5.06	-6.87	-7.47	-7.35
9	-6.62	-8.15	-7.73	-5.54	-5.60	-4.69	-6.77	-7.17	-6.45
10	-6.32	-7.48	-7.31	-5.21	-4.70	-4.69	-6.69	-7.12	-6.04

Table 14: Co-tail risk of factors with the market (July 1976 - June 2023).

Panel A reports the time-series average of the monthly percentage return of each factor and the associated t-statistics conditional on the market performance. r_{MKT} is the market excess return plus the risk-free rate in the Ken French's data library. N(months) is the total number of months that satisfy the given condition on market performance. The considered market conditions include months with market returns that are > 0, < 0, < -5, < -7, and < -10. The considered factors include the annual- and monthly-rebalanced profitability factors formed based on $Y(\text{IB})/B$, $Y(\text{FF})/B$, $Y(\text{RD})/B$, and $Y(\text{NM})/B$. Panel B reports the percentage return of each factor in each of the nine months in which the market return is lower than -10%.

Panel A: Subsamples

	N(months)	$r^{\text{IB}}_{\text{ROE}}$	$r^{\text{FF}}_{\text{ROE}}$	$r^{\text{RD}}_{\text{ROE}}$	$r^{\text{NM}}_{\text{ROE}}$	$r^{\text{IB}}_{\text{ROEQ}}$	$r^{\text{FF}}_{\text{ROEQ}}$	$r^{\text{RD}}_{\text{ROEQ}}$	$r^{\text{NM}}_{\text{ROEQ}}$
Full sample	564	0.20	0.36	0.46	0.38	0.61	0.63	0.76	0.53
		1.98	*3.53*	*6.08*	*4.65*	*5.08*	*5.65*	*8.31*	*6.17*
$r_{\text{MKT}} > 0$	361	-0.22	-0.04	0.33	0.46	0.27	0.35	0.68	0.65
		-1.82	*-0.35*	*3.82*	*4.68*	*1.84*	*2.61*	*6.11*	*6.23*
$r_{\text{MKT}} < 0$	203	0.96	1.08	0.70	0.23	1.21	1.13	0.90	0.31
		5.42	*6.07*	*4.85*	*1.61*	*6.00*	*5.85*	*5.66*	*2.10*
$r_{\text{MKT}} < -5$	51	2.27	2.34	1.59	0.41	2.86	2.30	1.79	0.47
		5.99	*5.53*	*5.21*	*1.27*	*5.60*	*4.56*	*4.31*	*1.30*
$r_{\text{MKT}} < -7$	30	2.29	2.35	1.87	0.65	3.13	2.26	1.99	0.43
		4.29	*3.77*	*4.52*	*1.46*	*4.27*	*3.20*	*3.35*	*0.84*
$r_{\text{MKT}} < -10$	9	3.93	3.34	2.87	0.82	3.58	2.60	2.48	0.29
		3.48	*2.21*	*3.11*	*0.67*	*2.40*	*1.75*	*1.96*	*0.22*

Panel B: Months with $r_{\text{MKT}} < -10$

Date	r_{MKT}	$r^{\text{IB}}_{\text{ROE}}$	$r^{\text{FF}}_{\text{ROE}}$	$r^{\text{RD}}_{\text{ROE}}$	$r^{\text{NM}}_{\text{ROE}}$	$r^{\text{IB}}_{\text{ROEQ}}$	$r^{\text{FF}}_{\text{ROEQ}}$	$r^{\text{RD}}_{\text{ROEQ}}$	$r^{\text{NM}}_{\text{ROEQ}}$
197810	-11.23	1.16	0.64	-0.32	-3.25	-1.53	-0.86	-0.86	-4.51
198003	-11.69	5.18	1.60	1.87	2.52	-2.03	-2.34	-2.34	-0.10
198710	-22.64	3.89	4.44	1.99	-3.83	0.83	1.28	1.28	-5.51
199808	-15.65	2.91	3.49	2.22	0.12	1.91	-0.20	-1.04	-1.83
200011	-10.21	11.54	14.23	9.47	7.89	12.72	12.83	9.20	7.10
200209	-10.21	3.51	3.10	1.08	-0.06	4.71	2.69	1.24	0.69
200810	-17.15	5.67	3.26	3.95	-0.94	5.13	4.75	6.05	1.05
200902	-10.09	0.58	1.42	3.55	4.01	5.53	4.02	5.02	4.94
202003	-13.26	0.91	-2.13	2.05	0.88	4.99	1.23	3.80	0.82

Table 15: Henriksson and Merton regression (July 1976 - June 2023).

$r^{i,t}_{\text{ROE}} = \alpha^s_i + \beta^U_i \text{RMRF} + \beta^{U-D}_i \max[-\text{RMRF}, 0] + \epsilon_{i,t}$, where i is IB, FF, RD, and NM; RMRF is the market excess return. The last term in the regression aims to capture the put option feature embedded in the factor returns. We report the regression results for the long-short portfolio (factor return), long side only, and short side only separately. t-statistics with heteroscedasticity-corrected standard errors is reported.

$$r^{i,t}_{\text{ROE}} = \alpha^s_i + \beta^U_i \text{RMRF} + \beta^{U-D}_i \max[-\text{RMRF}, 0] + \epsilon_{i,t}$$

	Long - Short				Long Side				Short Side		
	α^s_i	β^U_i	β^{U-D}_i		α^s_i	β^U_i	β^{U-D}_i		α^s_i	β^U_i	β^{U-D}_i
IB	0.12	-0.10	0.11	IB	0.17	1.04	-0.02	IB	0.05	1.15	-0.13
t-stat	*0.79*	*-2.49*	*1.54*	*t-stat*	*2.20*	*49.85*	*-0.71*	*t-stat*	*0.31*	*27.24*	*-1.88*
White t-stat	*0.81*	*-2.13*	*1.39*	*White t-stat*	*2.27*	*44.98*	*-0.63*	*White t-stat*	*0.31*	*22.15*	*-1.74*
FF	0.29	-0.10	0.10	FF	0.29	1.01	-0.06	FF	0.01	1.11	-0.15
t-stat	*1.82*	*-2.26*	*1.42*	*t-stat*	*3.46*	*44.78*	*-1.51*	*t-stat*	*0.04*	*28.86*	*-2.44*
White t-stat	*2.04*	*-2.04*	*1.37*	*White t-stat*	*3.63*	*36.33*	*-1.19*	*White t-stat*	*0.05*	*25.55*	*-2.48*
RD	0.24	0.02	0.14	RD	0.24	1.07	-0.03	RD	0.00	1.05	-0.17
t-stat	*2.02*	*0.67*	*2.79*	*t-stat*	*2.84*	*46.91*	*-0.81*	*t-stat*	*0.03*	*29.88*	*-3.03*
White t-stat	*2.07*	*0.55*	*2.55*	*White t-stat*	*2.96*	*40.51*	*-0.78*	*White t-stat*	*0.03*	*24.44*	*-2.80*
NM	0.17	0.10	0.10	NM	0.26	1.06	-0.04	NM	0.09	0.96	-0.14
t-stat	*1.33*	*2.84*	*1.78*	*t-stat*	*2.61*	*40.57*	*-0.98*	*t-stat*	*0.73*	*31.01*	*-2.79*
White t-stat	*1.25*	*2.55*	*1.40*	*White t-stat*	*2.34*	*34.45*	*-0.79*	*White t-stat*	*0.77*	*27.99*	*-2.52*

Table 16: Variable definitions.

Variable	Definition
REVT	Revenue (quarterly equivalent: REVTQ), Data source: Compustat
COGS	Cost of good sold (quarterly equivalent: COGSQ), Data source: Compustat
XSGA	Selling, general and administrative expense (quarterly equivalent: XSGAQ), Data source: Compustat
XRD	Research and development expense (quarterly equivalent: XRDQ), Data source: Compustat
DP	Depreciation (quarterly equivalent: DPQ), Data source: Compustat
XINT	Interest and related expense (quarterly equivalent: XINTQ), Data source: Compustat
SPI	Special items (quarterly equivalent: SPIQ), Data source: Compustat
IB	Income before extraordinary items (quarterly equivalent: IBQ), Data source: Compustat
CEQ	Common/ordinary equity, Data source: Compustat
AT	Total assets, Data source: Compustat
DVC	Common/ordinary dividends, Data source: Compustat
RDQ	Report date of quarterly earnings, Data source: Compustat
PRIMEXCH	Primary exchange, Data source: CRSP
SHRCD	Share code, Data source: CRSP
SGA	XSGA excluding research and development related expense following Peters and Taylor (2017)
ME	Market equity, price times number of shares outstanding. For factor construction and cross-section return regressions, we use June-end market equity; while for cross-section earnings forecasting regressions, we use December-end market equity
B	Book equity, equal to CEQ
BQ	Quarterly book equity computed following Hou et al. (2015)
B^{PT}	Intangible-adjusted book equity following Peters and Taylor (2017), data source: `https://wrds-www.wharton.upenn.edu/`, date visited: December 2, 2023
BM	Book-to-market ratio, B/ME
AG	Asset growth, dAT(t)/AT(t-1)
$Y(\mathrm{REVT})$	REVT
$Y(\mathrm{NM})$	REVT-COGS, following Novy-Marx (2013)
$Y(\mathrm{RD} + \mathrm{SG})$	REVT-COGS-XINT, following Eisfeldt et al. (2022) definition of investment in intangibles
$Y(\mathrm{RD} + .3\mathrm{SG})$	REVT-COGS-XSGA-XINT+0.3*SGA+XRD, following Peters and Taylor (2017) definition of investment in intangibles
$Y(\mathrm{RD})$	REVT-COGS-XSGA-XINT+XRD
$Y(\mathrm{FF})$	REVT-COGS-XSGA-XINT, following Fama and French (2015) instruction of RMW factor
$Y(\mathrm{EBT})$	REVT-COGS-XSGA-XINT-DP
$Y(\mathrm{SPI})$	IB-SPI
$Y(\mathrm{IB})$	IB, following Fama and French (2006) using annual earnings IB and following Hou et al. (2015) using quarterly earnings IBQ
FF5	Fama-French five factor model, Fama and French (2015), data source: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`, date visited: December 2, 2023
FF5(C)	Fama-French five factor model with RMW replaced by the cash-based operating profitability factor, Fama and French (2018)
FF6	Fama-French five factor model plus momentum factor, Fama and French (2016)
$\mathrm{FF6}^1$	FF6 with HML replaced by HML(Devil)
$\mathrm{FF6}^2$	FF6 with HML replaced by HML(EKP)
HXZ4	Hou-Xue-Zhang four factor model, Hou et al. (2015), data source: `https://global-q.org/factors.html`, date visited: December 2, 2023
HMXZ5	Hou-Mo-Xue-Zhang five factor model, HXZ4 plus expected growth factor, Hou et al. (2021), data source: `https://global-q.org/factors.html`, date visited: December 2, 2023
DHS3	Daniel-Hirshleifer-Sun three factor model, Daniel et al. (2020), data source: `https://sites.google.com/view/linsunhome`, date visited: December 2, 2023
SY4	Stambaugh-Yuan four factor model Stambaugh and Yuan (2017), available up to the end of 2016, data source: `https://finance.wharton.upenn.edu/~stambaug/`, date visited: December 2, 2023
UMD	Momentum factor, data source: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`, date visited: December 2, 2023
RMW	Robust-minus-weak factor in FF5
HML(EKP)	Intangible-adjusted value factor, Eisfeldt et al. (2022), data source: `https://github.com/edwardtkim/intangiblevalue`, date visited: December 2, 2023
HML(Devil)	Monthly-rebalanced HML factor with most recent market equity, Asness and Frazzini (2013), data source: `https://www.aqr.com/Insights/Datasets/The-Devil-in-HMLs-Details-Factors-Monthly`, date visited: December 2, 2023

Internet Appendix for "Intangibles Investment and Asset Quality"

Ravi Jagannathan*
Robert A. Korajczyk†
Kai Wang‡

First time available online: November 13, 2023

May 28, 2024

*Kellogg School of Management, Northwestern University, and NBER, Email: rjaganna@kellogg.northwestern.edu; Tel: 1-847-491-8338.

†Kellogg School of Management, Northwestern University, Email: r-korajczyk@kellogg.northwestern.edu; Tel: 1-847-491-8336.

‡Kellogg School of Management, Northwestern University, Email: kai.wang2@kellogg.northwestern.edu.

Contents

1 Additional Figures and Tables

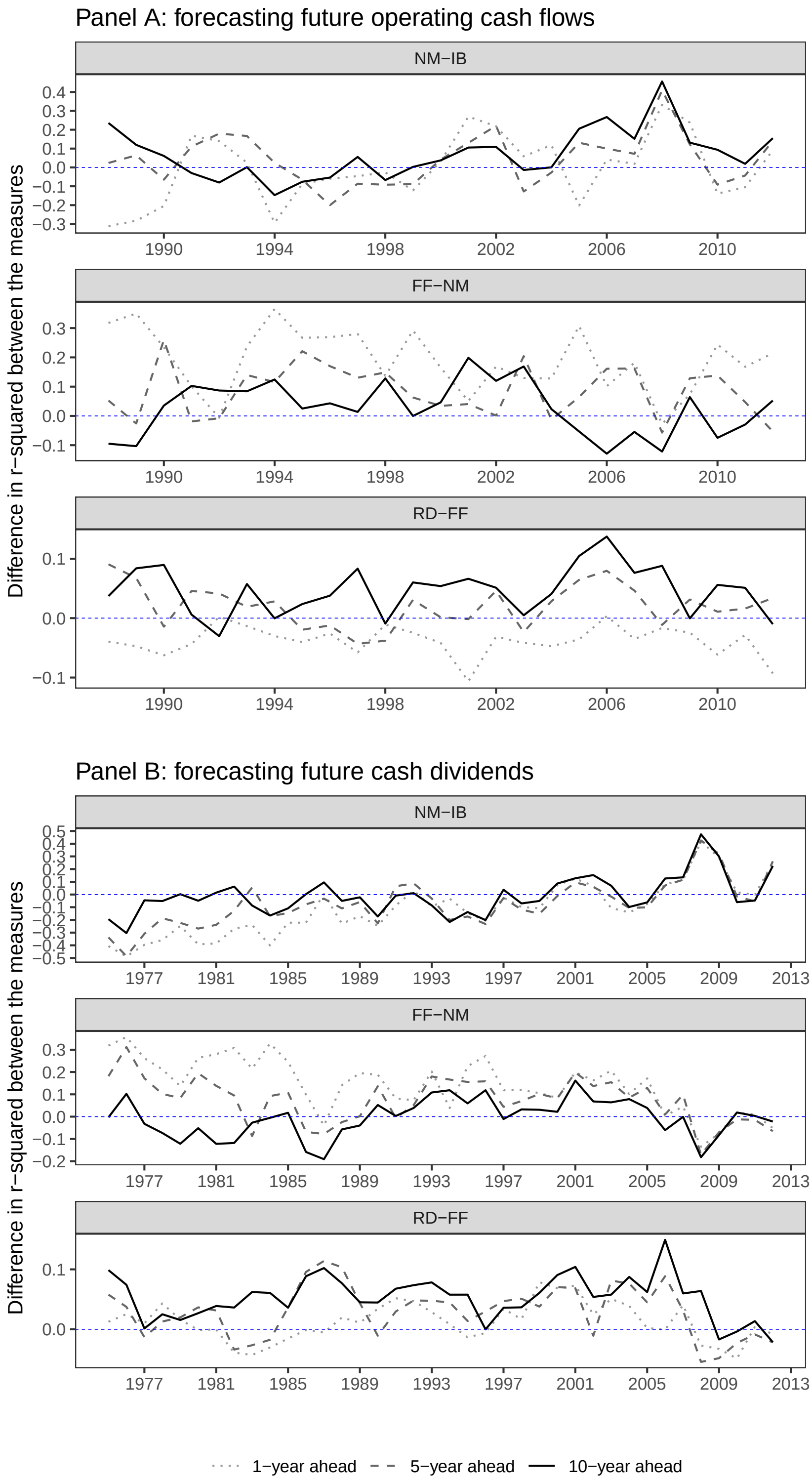

Figure 1: The differences in R^2 of the cross section regressions when forecasting future cash flows using different profitability measures at 1-, 5-, 10-year horizons.

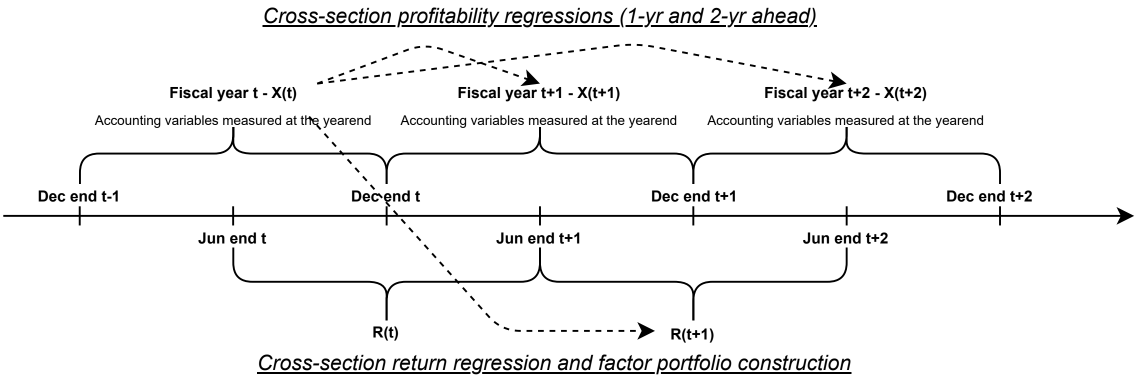

Figure 2: Timeline for portfolio formation and cross-section profitability regressions.

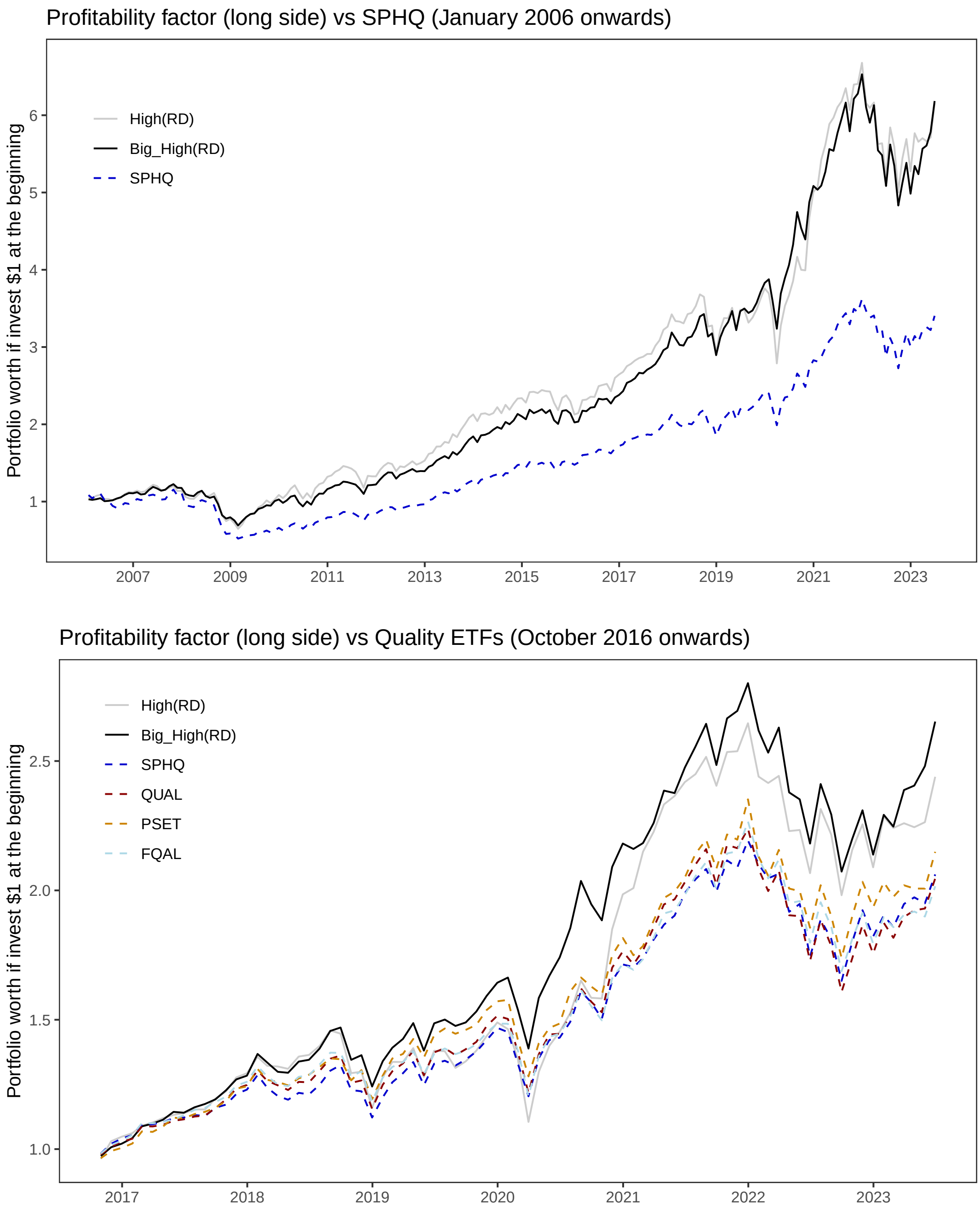

Figure 3: Cumulative return of annual-rebalanced high Y(RD)/B portfolio and quality ETFs. The considered quality ETFs include Invesco S&P 500 Quality ETF (SPHQ, available from January 2006), iShares MSCI USA Quality Factor ETF (QUAL, available from August 2013), Principal Quality ETF (PSET, available from April 2016), and Fidelity Quality Factor ETF (FQAL, available from October 2016). Since some of the quality ETFs are concentrated on large firms, we show the cumulative return of the big-high Y(RD)/B portfolio separately.

Table 1: Fractions of firms delisted after τ years (1975 - 2022).

This table reports the time-series average of the fractions of firms that are delisted after $\tau = 1, 2, ..., 10$ years, $(N_0 - N_\tau)/N_0$. The reasons of the delisting correspond to the first digit of of delisting code (CRSP data item DLSTCD). To be included, a firm must be listed in one of the three major exchanges and have a share code of 10 or 11. Merger, Dropped, Exchanges, and Liquidation are the fractions of firms that are delisted for the given reason divided by the total number of firms delisted. We remove firms with total assets less than \$12.5 million or book equity less than \$25 million in year t.

τ	0	1	2	3	4	5	6	7	8	9	10
Panel A: All firms											
N_τ	3515	3323	3092	2876	2679	2495	2330	2181	2045	1921	1808
$(N_0 - N_\tau)/N_0$		0.05	0.11	0.17	0.23	0.28	0.32	0.37	0.40	0.44	0.47
Panel B: Fraction of firms delisted due to various reasons											
Merger		0.04	0.09	0.13	0.17	0.21	0.24	0.27	0.30	0.33	0.35
Dropped		0.01	0.02	0.03	0.05	0.06	0.07	0.08	0.08	0.09	0.10
Exchanges		0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02
Liquidation		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01

2 Dividend Discount Model, Present Value Relation, and Characteristics-based Factors

Fama and French (2006, 2015) use the dividend discount model (DDM) to motivate the characteristics-based five-factor model. DDM states that the market price of one share of stock is the sum of the discounted value of expected dividends per share,

$$P_{i,t} = \sum_{\tau=1}^{\infty} \mathbb{E}_t \left(D_{i,t+\tau} \right) / (1 + r_{i,t})^{\tau}, \tag{1}$$

where $P_{i,t}$ is the market value of one share of firm i's equity at time t, $\mathbb{E}_t \left(D_{i,t+\tau} \right)$ is the expectation at time t for the dividend per share in period $t + \tau$, and $r_{i,t}$ is the discount rate for firm i's all future dividends at time t. Note that DDM is a valuation model that takes $\mathbb{E}_t \left(D_{i,t+\tau} \right)$ and $r_{i,t}$ as inputs, and $P_{i,t}$ as output, and it is an assumption to apply the same discount rate to all future dividends. DDM implies that if two firms have the same stream of expected dividends per share, the firm with riskier expected dividends should have higher discount rate and thereby lower price per share. Therefore, if pricing is rational, a strategy that long low price stocks and short high price stocks, controlling for the expected dividends, should have positive expected return (though zero risk-adjusted return).[1]

Equation (1) has another interpretation as a present value relation (PVR), which is a mathematical identity that is used to compute the internal rate of return, implied cost of capital, or stock yield (analogous to bond yield). It takes the observed stock price and estimates of expected dividends as inputs, and yields r as output. The PVR implies that a higher yield is obtained from buying the same expected stream of dividends at a lower price. This should hold both ex post and ex ante. Therefore, the same long-short strategy has positive expected return irrelevant of rational or irrational pricing (as noted in Fama and French (2015)). Campbell and Shiller (1988) use PVR to motivate dividend-to-price as a natural candidate to predict stock returns, and Gebhardt et al. (2001) compute the implied cost of capital based on PVR and find it strongly correlated with the future stock returns.

By applying the fact that dividend is the after-tax earnings minus investment, and the clearing surplus accounting, Fama and French (2015) rewrite equation (1) as follows:

$$\frac{P_{i,t}}{B_{i,t}} = \sum_{\tau=1}^{\infty} \mathbb{E} \left(\frac{Y_{i,t+\tau}}{B_{i,t}} - \frac{dB_{i,t+\tau}}{B_{i,t}} \right) \Big/ (1 + r_{i,t})^{\tau}. \tag{2}$$

where $Y_{i,t+\tau}$ is the after-tax earnings of firm i in year $t+\tau$, $B_{i,t+\tau}$ is the book value of equity,

[1] The dividend discount model, however, is silent about the source of the systematic risk, i.e., it does not endogenously specify why some firms have higher expected return than others.

$dB_{i,t+\tau}$ is the change of the book equity from $t+\tau-1$ to $t+\tau$, $P_{i,t}$ is the market value of the equity, and $r_{i,t}$ is internal rate of return. This motives the four characteristics included in the Fama-French five factor model: size, book-to-market equity ratio, profitability, and investment rate.

In this section, we use the Gordon growth model (GGM), a DDM with perpetual profitability, investment rate, and discount rate, to relate the DDM/PVR to the characteristics-based linear factor model.[2]. The GGM (by suppressing the time subscript) is given as follows:

$$\frac{P_i}{B_i} = \frac{\mathrm{ROE}_i \times (1 - \mathrm{PBR}_i)}{r_i - g_i} \tag{3}$$

where r_i is the discount rate, ROE_i is the return on equity, PBR_i is the equity plowback rate, and $g_i = \mathrm{ROE}_i \times \mathrm{PBR}_i$ is the growth rate. Our simulated model economy assumes three exogenous variables, ROE, expected return, and PBR. The price-to-book ratio is determined by the present value relation in equation (3).[3] Note that the Fama-French investment factor is constructed based on the asset growth, measured by the annual change of total assets divided by the lagged total assets. In the GGM, this measure of investment is equivalent to the growth rate rather than PBR. Therefore, we consider four characteristics that are potentially related to the cross section of expected return: ROE, P/B, PBR, and growth rate.

In the simulation, we first focus on the cross section relation between characteristics and expected returns. We find significant relations between expected return and single characteristics except for ROE while ROE becomes significant when controlling for other characteristics. The growth rate is significant when holding PBR fixed but not vice versa. We therefore use growth rate rather than PBR when constructing factors. In our simulation, the PBR and growth rate are positively related to the expected return whereas it is negative in real data. We will argue that this contradiction is due to the interaction between future ROE and PBR which leaves the relation between the expected returns and PBR (and thereby growth rate) undetermined.

We then construct a model economy where returns are generated according to a linear latent factor structure in which the expected return of firms and their factor loadings are matched by construction. We examine whether in this example economy characteristics based factors are able to explain the cross section of returns, and whether they do as well as the Connor-Korajczyk (CK) factors in explaining the cross section of returns. We find that the CK factors do much better than characteristics-based factors, which contradicts what is found in the actual return data. We argue that this can be due

[2]We choose to use Gordon growth model for the simulation because it is enough to clarify the relation between linear factor model and DDM/PVR without explicit modeling of the term structures of the variables.

[3]In this simple example, we assume that PBR is exogenously given, however, rational corporate investment behavior can be easily included: firms will always invest at the maximum or minimum level based on the relation between ROE and discount rate.

to time varying factor loadings. Therefore, we construct another economy with time variation in risk exposures. We find characteristics-based factors outperform the CK factors in that economy.

2.1 Calibration

We construct a cross section of firms with expected perpetual ROE, PBR, and r. To calibrate discount rates, we first estimate the market beta for each firm with complete observations from 2013 to 2022 in the CRSP/Compustat merged sample. Such a filtering process biases the sample toward high quality firms but helps eliminate outliers which is particularly important when calibrating for perpetual ROE. The firms must also have a NYSE, AMEX, or NASDAQ primary exchange code and a share code of 10 or 11 to be included.[4] We take the mean (1.10) and standard deviation (0.51) of the market beta across firms, and then compute the discount rate by multiplying a 6% annual risk premium by the market beta, assuming, without loss of generality, that the riskless rate is 0. The (annual) expected returns r_i of the 4,000 firms are therefore drawn from a normal distribution with mean 6.62% and standard deviation 3.06%.

To calibrate for ROE, we compute the 10-year average ROE[5] for each firm and exclude firms with negative sum of book equity or ROE below 0% or above 60%[6]. The expected ROE of the 4,000 simulated firms are drawn from a normal distribution with mean 0.25 and standard deviation 0.13. The top and bottom panels of Figure 4 plot the empirical distribution of ROE and expected return.

The GGM is, obviously, a toy model, which assumes r_i and ROE_i and PBR_i are constant in perpetuity. In that setting certain combinations of parameters lead to implausibly high (infinite) valuations. In oreder to avoid these situations we place the following restrictions on the parameters. We exclude the draws of ROE_i outside the range [0%, 60%] and draw of r_i outside the range [2%, 12%]. Then we draw plowback rate from uniform distribution over $[0, \frac{r_i}{2\,\text{ROE}_i}]$, subject to the restriction that $\text{PBR}_i < 1$. Though we start with 4,000 firms, there are, on average, 3,413 firms (with a minimum of 3,349 and a maximum of 3,476) left after applying all the restrictions for the 100 independent cross sections we generated.

Imposing equation (3) to the randomly generated ROE, PBR, and the expected returns, we can compute the P/B for firms. In our setup, ROE and expected return are exogenous variables, and a ratio of them, $\frac{r_i}{2\,\text{ROE}_i}$, determines the upper bound of the PBR. This does not make PBR endogenous because firms with ROE greater (lower) than expected return will always want to invest at maximum (minimum) level. To keep the

[4]There are 1,853 out of total 6,016 firms with complete observations in the CRSP/Compustat merged sample from 2013 to 2022.

[5]$\text{ROE} = \frac{1}{10}\sum_{t=2013}^{2022} \frac{\text{REVT}_t - \text{COGS}_t - \text{XSGA}_t - \text{XINT}_t + \text{RD}_t}{\text{CEQ}_t}$.

[6]This leaves a final sample of 1,586 firms

example easy to follow, we choose not to specify a model to endogenize the corporate investment behavior.

2.2 Relation between Expected Returns and Characteristics

We rewrite GGM to derive the relations between expected return and ROE, P/B, and PBR respectively as follows:

$$r_i = \left(\mathrm{PBR}_i + \frac{1 - \mathrm{PBR}_i}{\mathrm{P}_i/\mathrm{B}_i}\right)\mathrm{ROE}_i, \tag{4}$$

$$r_i = \mathrm{ROE}_i\,\mathrm{PBR}_i + \mathrm{ROE}_i(1 - \mathrm{PBR}_i)\frac{1}{\mathrm{P}_i/\mathrm{B}_i}, \tag{5}$$

$$r_i = \frac{\mathrm{ROE}_i}{\mathrm{P}_i/\mathrm{B}_i} + \left(1 - \frac{1}{\mathrm{P}_i/\mathrm{B}_i}\right)\mathrm{ROE}_i\,\mathrm{PBR}_i. \tag{6}$$

According to equations (4) and (5), there is a negative relation between r_i and P/B and a positive relation between r_i and ROE when controlling for other characteristics, however, the relation between r_i and PBR depends on whether P/B is above or below one (see equation (6)). A firm with $\mathrm{P}_i/\mathrm{B}_i > 1$ has $\mathrm{ROE}_i > r_i$, and therefore any extra investment earn positive net value for the equity holders. In this case, to get the same P/B ratio with a same ROE, a firm with higher expected return must have higher PBR. Therefore, we could expect a positive relation between the expected return and PBR (and thereby growth rate) when holding other characteristics fixed if the sample is dominated by firms with P/B above one.[7]

Though the potential characteristics are identified, whether all characteristic provide incremental explanatory power in explaining the cross section of expected returns remains a question. We use two methods for this question: (1) run regressions of expected returns on the characteristics (results are reported in Panel A of Table 2); (2) assign firms into groups based on characteristics and form long-short portfolios (results are reported in Panel B and C of Table 2). The DDM motivation for characteristics-based factors requires controlling for other characteristics which naturally relies on conditional sorting procedure to form portfolios while the conventional approach is to form portfolios independently. Therefore, we run both univariate and multiple regressions and apply both independent and conditional sorting in portfolio assignment. We can expect that the results will be similar when the correlation between characteristics is low.

In the univariate regressions, all characteristics are strongly associated with the expected returns except for ROE.[8] However, ROE turns significant in the multiple regres-

[7]The expected return of a high-minus-low PBR portfolio can be different from the expected return of a high-minus-low growth rate portfolio if PBR and ROE do not have a zero correlation within each ROE group.

[8]We generate the cross section independently 100 times. The t-statistics of ROE in the univariate regression ranges from −0.8 to 4.0 with a mean of 1.2 while the absolute value of the t-statistics of other

sions with P/B and PBR which shed light on the importance of the keeping other things fixed in the factor construction. Because the asset growth measure in Fama and French (2015) is equivalent to the growth rate measure, $g_i = \text{ROE}_i \times \text{PBR}_i$, in the GGM, we include the growth rate as an independent variable in the regression. We find that PBR becomes less significant when the growth rate, g, is added.[9] We obtain similar coefficients and t-statistics for remaining characteristics in the multiple regression without PBR.

For the portfolio sorts, we assign firms into groups based on (1) univariate decile sort, (2) 2-by-2-by-3 P/B, PBR, and ROE sort, (3) 2-by-2-by-3 P/B, ROE, and growth rate sort, and (3) 2-by-2-by-2-by-3 P/B, PBR, ROE, and growth rate sort of given characteristics. The return of each group is the equal weighted average return and the return for each long-short strategy is the mean return of the top groups minus the mean return of the bottom groups based on the given characteristic. Panel B and C in Table 2 apply independent sorts and conditional sort, respectively, and report the mean, minimum, and maximum of the strategy returns based on the 100 independently generated cross sections. Using the 2-by-2-by-3 P/B, ROE, and PBR conditional sort strategy as an example, we first assign firms into 2-by-2 groups based on median P/B and median ROE, and then assign firms with each group further into three groups based on the 30 and 70 percentiles of PBR. The monthly percentage expected return of this strategy is the average return of the four high PBR groups minus the average return of the four low PBR groups. We also report the performance of similar strategies using CRSP/Compustat merged data from July 1963 to June 2023 in Panel D in Table 2. To make comparison easier, we convert the annual expected returns into monthly percentage expected returns for the simulation results.[10]

The results using portfolio sorts are consistent with the findings in the regression analysis. A univiariate sort on ROE performs poorly with a 0.02 monthly percentage expected return and the minimum and maximum values take opposite signs, while the other three characteristics show persistent results. With both independent and conditional sorts, the performance on ROE strategy improves and becomes persistent when controlling for other characteristics. We then use all four characteristics to form 2-by-2-by-2-by-3 portfolios. The return of the PBR strategy becomes weak and removing it does not attenuate the performance of the other three characteristics.[11] More interestingly, the strategy performance of the independent sorts is better than the conditional sort across all characteristics. However, it remains an empirical question whether the better performance of independent sort holds in the real data.

---

characteristics is not less than 15.

[9]The t-statistics of PBR ranges from -3.1 to 1.1 with a mean of -0.9 while the absolute value of the t-statistics of other characteristics is not less than 24.

[10]The mean, minimum, and maximum of the equal weighted average expected (monthly percentage) return are 0.562, 0.555, and 0.571 respectively.

[11]The 2-by-2-by-2-by-3 independent sort does not always guarantee 24 groups with due to the high correlation between PBR and growth rate.

We then compute the performance of the similar strategies using CRSP/Compustat merged data from July 1963 to June 2023.[12] Note that P/B, ROE, and growth rate in the GGM corresponds to the (opposite of) book-to-market, operating profitability, and investment rate in Fama and French (2015) respectively. For all three characteristics, we compute the strategy performance using the mean return of the top groups minus the mean return of the bottom groups for the given characteristic. P/B and ROE are not significant with the univariate sort, and growth rate is not significant with 2-by-2-by-3 sort.[13] There is some evidence that independent sort performs better than the conditional sort for ROE, but no evidence for P/B and growth rate. More importantly, the sign of the growth rate strategy contradicts the simulation results.

To understand the relation between the expected return and growth rate, first note that when controlling for ROE and P/B, this relation is determined by the relation between the expected return and PBR.[14] There are two sources of ambiguity on the sign of this relation. First, whether P/B > 1 is a dominant feature in the population. Second, the impact of PBR on the book value of equity. In the GGM, ROE is the expected perpetual return on the one-period lagged book value of equity. As the firm invests more at $t+\tau$, the firm will have more book equity at $t+\tau+1$ and thereby higher earnings at $t+\tau+1$ with that same ROE. In equation (2), however, future profitability is measured by future earnings divided by the book equity at time t. With such a profitability measure, the more a firm invests at t the higher the after-tax earnings at $t+1$ will be if ROE is constant and book equity grows with investment. Thus, one cannot vary $dB_{t+\tau}/B_t$ while keeping $Y_{t+\tau+1}/B_t$ fixed in equation (2). We will need additional assumptions beyond the present value relation to explain the empirically discovered negative relation between expected return and investment rate.[15]

2.3 Linear Latent Factor Structures

We wish to determine whether, in a world in which asset returns follow a factor structure, whether the characteristic-based factors can mimic the true underlying factors. That

[12] We measure book equity following Fama and French (2006), size as the market equity at the end of June in year $t-1$, growth, g, as asset growth and ROE as the operating profitability following Fama and French (2015). We use Compustat data item CEQ to measure book equity when constructing intangible-adjusted profitability factor because its coverage is broad enough after 1975 when the R&D expense data became valid.

[13] Our results for the univariate sort and 2-by-3 Size-X sort are close to the results in Kenneth French's data library. In Kenneth French's data library, the time-series average (monthly percentage) return gaps between top 10% and bottom 10% for book-to-market, operating profitability, and investment are 0.34, 0.30, and -0.40, and the associate t-statistics are 1.86, 1.86, and -3.11 respectively from July 1963 to June 2023.

[14] $\frac{P_i}{B_i} = \frac{\mathrm{ROE}_i - \mathrm{ROE}_i\,\mathrm{PBR}_i}{r_i - \mathrm{ROE}_i\,\mathrm{PBR}_i}$

[15] A mechanic fix to match the empirically found negative relation between expected return and investment rate is to assume higher expected returns for firms with lower investment rate potentially due to behavioral reasons.

is, we construct a world where the cost of capital for assets is determined by four latent factors. We study the performance of the characteristic-based factors (long/short portfolios) relative to latent factors estimated by asymptotic principal components Connor and Korajczyk (1986). We assume the following linear (latent) factor model for returns:

$$R_{i,t} = \underbrace{\sum_{k=1}^{K} \beta_{i,k}\lambda_k}_{r_i} + \underbrace{\sum_{k=1}^{K} \beta_{i,k}F_{k,t}}_{\beta_i F_t} + \epsilon_{i,t}, \tag{7}$$

where λ_k is the expected return of the kth factor, $F_{k,t}$ is the realized value of the kth factor at time t, and it is normally distributed with mean zero and standard deviation σ_k, $\beta_{i,k}$ is the firm i's loading on the kth factor, and $\epsilon_{i,t}$ is the noise term for firm i at time t, and $K = \{1, 2, 3, 4\}$ is the total number of latent factors. We generate 600 monthly observations for the factors and 600 monthly noise term independently. For simplicity of the notation, we assume a risk-free rate of zero. When $K = 4$, we refer the four factors as the market, ROE, P/B, and growth rate. The three characteristics, ROE, P/B, and growth rate, are chosen because they show strong correlation with the firms' expected returns. When $K \leq 4$, we calibrate for only the first K factors of the market, ROE, P/B, and growth rate factors.

We choose the first cross section (out of the 100 cross sections) generated in the last section to fix the firms' characteristics and expected returns.[16] The chosen cross section has 4,000 firms, in which 564 firms are removed due to the restrictions we impose. When $K = 4$, the true expected returns, λ_k, of the four factors, market, ROE, P/B, and growth rate, are naturally the expected returns of the corresponding factors for the given cross section, which are 0.56, 0.22, -0.40, and 0.19 in monthly percentage respectively. The standard deviations of the four latent factors, $F_{k,t}$, are set equal to the standard deviations of the Fama-French RMW, HML, and CMA in the last 50 years (from January 1973 to December 2022), which are 4.64, 2.33, 3.10, and 2.04 in monthly percentage respectively. The noise term $\epsilon_{i,t}$ is assumed to be normally distributed with mean zero and the standard deviation 10 in monthly percentage, which is set equal to the empirical standard deviations of the CAPM residuals of firms from 2013 to 2022.[17]

The firms' factor loadings, $\beta_{i,k}$, are generated as follows. When $K = 1$, $\beta_{i,1}^{K=1} = r_i/\lambda_1$, where r_i is the expected return of firm i and λ_1 is the equal-weighted expected return of all firms. In this case, the factor loading $\beta_{i,1}^{K=1}$ is perfectly correlated with the true expected return of firm i. When $K = 2$, we randomly draw $\beta_{i,1}^{K=1}$ without replacement

[16]Randomly choosing a cross section should have no impact on the simulation results as the quantitative relations derived from the static cross sections are stable.

[17]For the firms that have complete monthly observations from January 2013 to December 2022, the mean/median of the standard deviation of the monthly return is 9.9/12.2. The mean/median of the standard deviation of the CAPM residual is 8.5/10.9. The Fama-French five factor model bring these numbers down to 7.9/10.2.

and set them as $\beta_{i,1}^{K=2}$. Then the firm i's loading on factor 2 is $\beta_{i,2}^{K=2} = \left(r_i - \beta_{i,1}^{K=2}\,\lambda_1\right)/\lambda_2$ where λ_2 is the expected return of the ROE factor for the chosen cross section. In this case, $\beta_{i,2}^{K=2}$ is perfectly correlated with the residual expected return of the first factor for firm i. Similarly, when $K = 3$, we set $\beta_{i,1}^{K=3} = \beta_{i,1}^{K=2}$, $\beta_{i,2}^{K=3}$ to be the randomly drawn $\beta_{i,2}^{K=2}$ (without replacement), and $\beta_{i,3}^{K=3} = \left(r_i - \beta_{i,1}^{K=3}\,\lambda_1 - \beta_{i,2}^{K=3}\,\lambda_2\right)/\lambda_3$ where λ_3 is the expected return of the P/B factor. Finally, wen $K = 4$, we set $\beta_{i,1}^{K=4} = \beta_{i,1}^{K=3}$, $\beta_{i,2}^{K=4} = \beta_{i,2}^{K=3}$, $\beta_{i,3}^{K=4}$ to be the randomly drawn $\beta_{i,3}^{K=3}$ (without replacement), and $\beta_{i,4}^{K=4} = \left(r_i - \beta_{i,1}^{K=4}\,\lambda_1 - \beta_{i,2}^{K=4}\,\lambda_2 - \beta_{i,3}^{K=4}\,\lambda_3\right)/\lambda_4$ where λ_4 is the expected return of the growth rate factor.

2.4 Characteristic-based vs. Statistical Factors

We repeat the simulation 100 times. The firms' characteristics and expected returns are fixed for not only the 100 simulations but also the 600 months within each simulation. The realization of the factors and noise terms are randomly generated based on the same given distribution for each simulation. Within each simulation we construct the market factor and the characteristics-based factors, and examine their performance and explanatory power for the cross section of stock returns. The market factor in a month is the equal-weighted average realized return of all firms in the given month. To construct the ROE factors, firms are assigned into 2-by-2-by-3 P/B, growth rate, and ROE groups independently based on median P/B, median growth rate, and 30 and 70 percentiles of ROE. We then take the equal-weighted average return for each group, and then take the equal-weighted average return for the four high ROE groups and four low ROE groups. The return of ROE factor in a month is the return difference between the high ROE groups and the low ROE groups in the given month. The P/B and growth rate factors are constructed using the same procedure except that firms are assigned into 30 and 70 percentiles of P/B and growth rate respectively. We do not rebalance the groups through the 600 months in each simulation as the firms' characteristics do not change over time. We also take the negative of the P/B factor return so that it has a positive expected return.

Figure 5 plots the histogram of the time-series average return of the market factor and three characteristics-based factors for the 100 simulations. Columns one to four correspond to the cases where the total number of latent factors equal to one to four, $K = 1, 2, 3, 4$, respectively. All characteristic-based factors have positive expected returns even when the true model has a one-factor structure, $K = 1$. This is not surprising characteristics are conditionally correlated with the cost of capital, hence correlated with market beta when the economy is priced by a single factor. We run regressions of the three characteristics-based factors on the market factor and report the intercepts in Figure 6. Interestingly, the intercepts of the factors are clearly centered around 0 when there is only one latent factor, $K = 1$ (see the first column). This implies that the characteristics-based factors do not seem to provide independent information when controlling for the market factor when the true model has only one factor even though they all have expected returns significantly different from zero. When $K > 1$, however, the characteristics-based factor alphas become significantly different from zero against the market factor but the intercepts do not increase as the total number of factors grows.

We also construct the 3-by-3-by-3 conditionally sorted ROE, P/B, and growth rate portfolios and use them as test assets to examine the explanatory power of the factors.[18] Figure 7 plots the histogram of the p-value of the GRS test of the 27 test assets on the market and the characteristics-based factors for the 100 simulations. The total number of latent factors can be 1 to 4, $K = 1, 2, 3, 4$ (for each row), and the total number of factors included in the factor model can be 0 to 4, $J = 0, 1, 2, 3, 4$ (for each column). When $J = 0$, the GRS test is equivalent to test whether the test assets have a maximum squared Sharpe ratio of zero. When $J > 0$, we always add the factors into the model in the following order: market, ROE, P/B, and growth rate. We provide the frequency of p-values being less or equal to 10% for each plot. If the factor model can fully explain the returns of the test assets, we should expect this frequency to be near 10%. This is clearly not the case when no factor is included in the model, $J = 0$ (see the first column). When the true model has one factor, $K = 1$, the market factor effectively lowers the rejection frequency from 44% to 13%, and adding more factors does not have any impact. This is implies that the market factor can be good enough to describe the cross section of stock returns when $K = 1$. When $K > 1$, however, the market factor is no longer sufficient. Adding more factors into the model does lower the chance of rejection, but the improvement seems to stop after adding the ROE factor and remains high relative to the significance level of 10% level.

We then estimate statistical factors following Connor and Korajczyk (1986) and examine their performance relative to the characteristics-based factors. Figure 8 plots the average eigenvalues of the first six principal components across the 100 simulations for $K = 1, 2, 3, 4$. We regress the three characteristics-based factors on the first C-K factor when $K = 1$, on the first two C-K factors when $K = 2$, on the first three C-K factors when $K = 3$, and on the first four C-K factors when $K = 4$. Figure 9 shows that the intercepts of the characteristics-based factors are clearly not significantly different from zero. Moreover, we compare the maximum squared Sharpe ratios of the statistical factors with the characteristics-based models (with a market factor) in Figure 10. When $K = 1$, the C-K model has a higher Sharpe ratio than the four-factor characteristics-based model for 63% of the simulations. This frequency becomes 61%, 88%, and 99% when $K = 2$, $K = 3$, and $K = 4$ respectively. The C-K models are also tested against the same test

[18] We do not use independent sort to construct the test assets because they do not guarantee 27 portfolios due to the high correlations among characteristics in extreme cases.

assets and the results are shown in Figure 11. The frequency of rejecting zero alphas for the test assets is generally close to, but slightly higher than, the stated 10% level of the test when the right C-K model is used.

This finding that C-K models can dominate the characteristic-based models contradicts what is typically found in actual data. We argue that this may be due to time-varying factor loadings in the actual data. In the next section, we show that characteristics-based factors have advantages over the C-K factors if the firms' factor loadings change over time.

2.5 Time-varying Factor Loadings

We generate time variations in factor loadings by repeating the generating process for the factor loadings described in section 2.3 every 24 months. When there is only one latent factor, the factor loadings will remain constant over the entire 600 months. When there are more than one latent factor, the factor loadings will change every 24 months. We construct both the characteristic-based factors and C-K factors over the entire sample following the same procedure as above. The performance of the characteristic-based factors and their intercepts against the market factor are very similar to those reported earlier in Figures 5 and 6. However, Figure 12 shows that the performance of the characteristics-based factors in describing the cross section improves substantially especially when the correct number of factors are chosen for a given number of latent factors.

Figure 13 shows that the eigenvalues of the C-K factors drop much faster after the first principal component when the factor loadings are time-varying relative to the simulations with static factor loadings. Figure 13 suggests choosing a two-factor model except when $K = 1$. For a more reasonable comparison to the static factor loading case, we use the correct number of C-K factors for a given K even though the third and fourth factors may not contribute to the performance of the model. The intercepts of the characteristics-based factors against the "correct" C-K model are given in Figure 14. As the number of latent factors grows, the C-K models become less capable of explaining the performance of the characteristics-based factors. Comparing the maximum squared Sharpe ratios, Figure 15 shows that the characteristics-based model can outperform the "correct" C-K model though when $K > 2$. The GRS test p-values of the test assets against the C-K models in Figure 16 provides additional evidence that the marginal benefit of adding the C-K factors become close to zero after the second principal component.

2.6 Summary

In this section/appendix, we argue that the DDM/PVR provides a solid motivation to construct the Fama-French five factors. It can be given a "casual" interpretation when viewed as a valuation model though the results hold irrespective of rational or irrational pricing because PVR is a mathematical identity. In the simulation, we take ROE, the discount rate, and PBR as given and let DDM formula determine the P/B ratio. We find that the relation between expected return and growth rate (investment rate in the context of the Fama-French five factor model) is not clear mainly due to the interaction between future investment rate and profitability. In our linear latent factor structure, the characteristic-based factors have positive expected returns and, hence, some explanatory power for the cross section of stock returns. When the factor loadings are static over the entire sample, the benefit of adding more than one characteristic-based factor, in addition to the market factor, seems limited whereas the asymptotic PCA factors dominate the characteristics-based factor models when the number of latent factors for simulated economies with multiple factors. However, when the factor loadings become time-varying, we find the opposite results: the benefit of having more than two asymptotic PCA factors seems limited and the characteristics-based factor models do better in describing the cross section of the stock returns and outperform C-K models.

References

Campbell, John Y, Robert J Shiller. 1988. Stock prices, earnings, and expected dividends. *the Journal of Finance* **43**(3) 661–676.

Connor, Gregory, Robert A. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* **15**(3) 373–394. doi:https://doi.org/10.1016/0304-405X(86)90027-9. URL `https://www.sciencedirect.com/science/article/pii/0304405X86900279`.

Fama, Eugene F, Kenneth R French. 2006. Profitability, investment and average returns. *Journal of Financial Economics* **82**(3) 491–518.

Fama, Eugene F, Kenneth R French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* **116**(1) 1–22.

Gebhardt, William R, Charles MC Lee, Bhaskaran Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Research* **39**(1) 135–176.

Tables and Figures

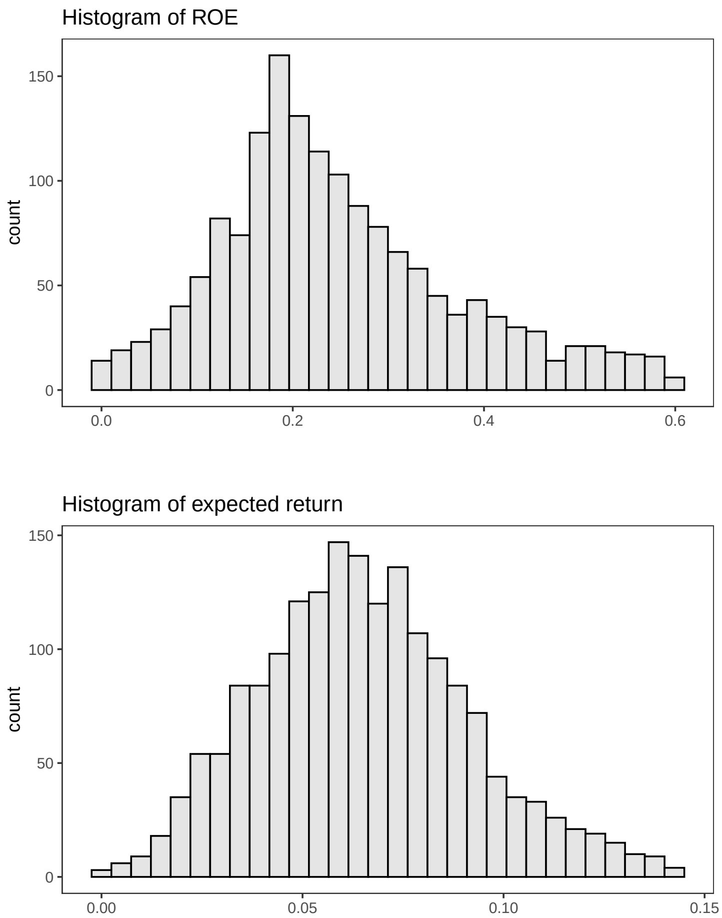

Figure 4: Sample of ROE and expected returns.

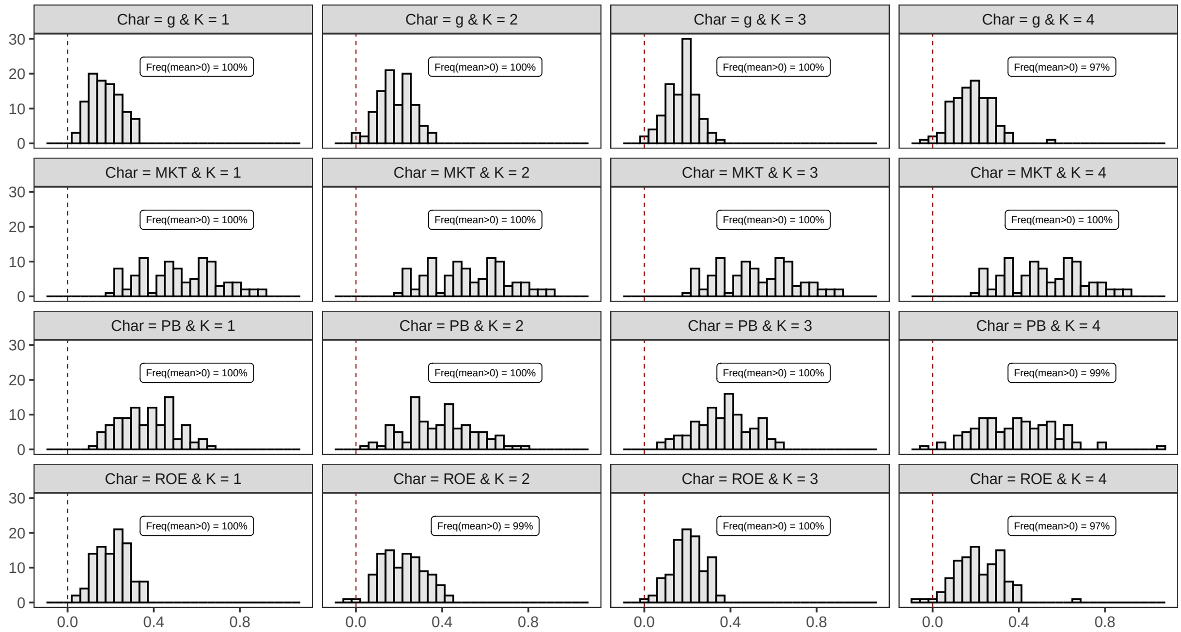

Figure 5: Histogram of time-series average returns of market and characteristics-based factors when K = 1, 2, 3, and 4.

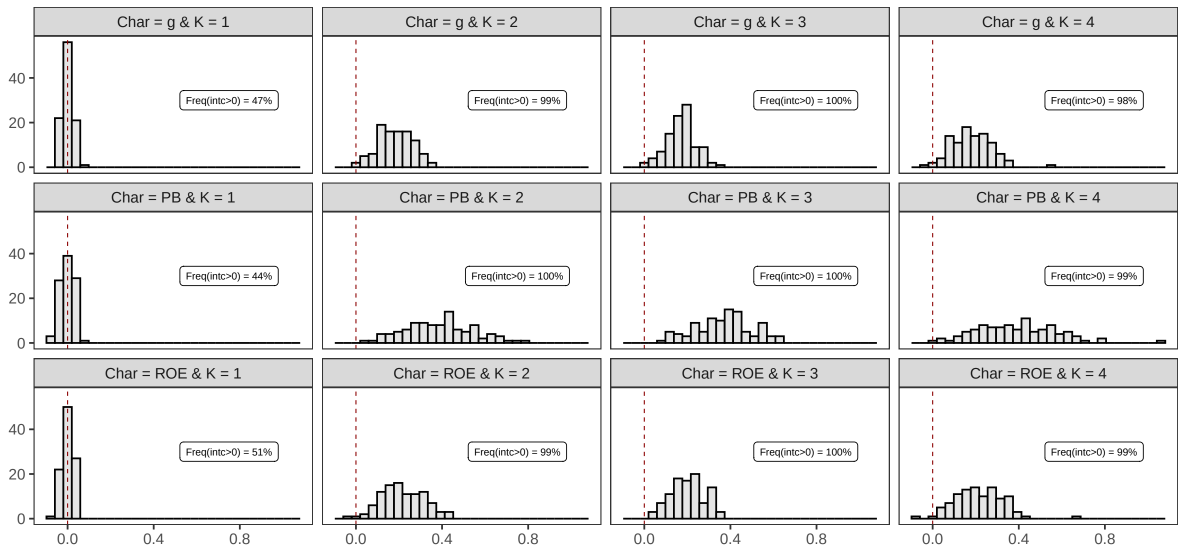

Figure 6: Histogram of intercept of characteristics-based factors on market factor when K = 1, 2, 3, and 4.

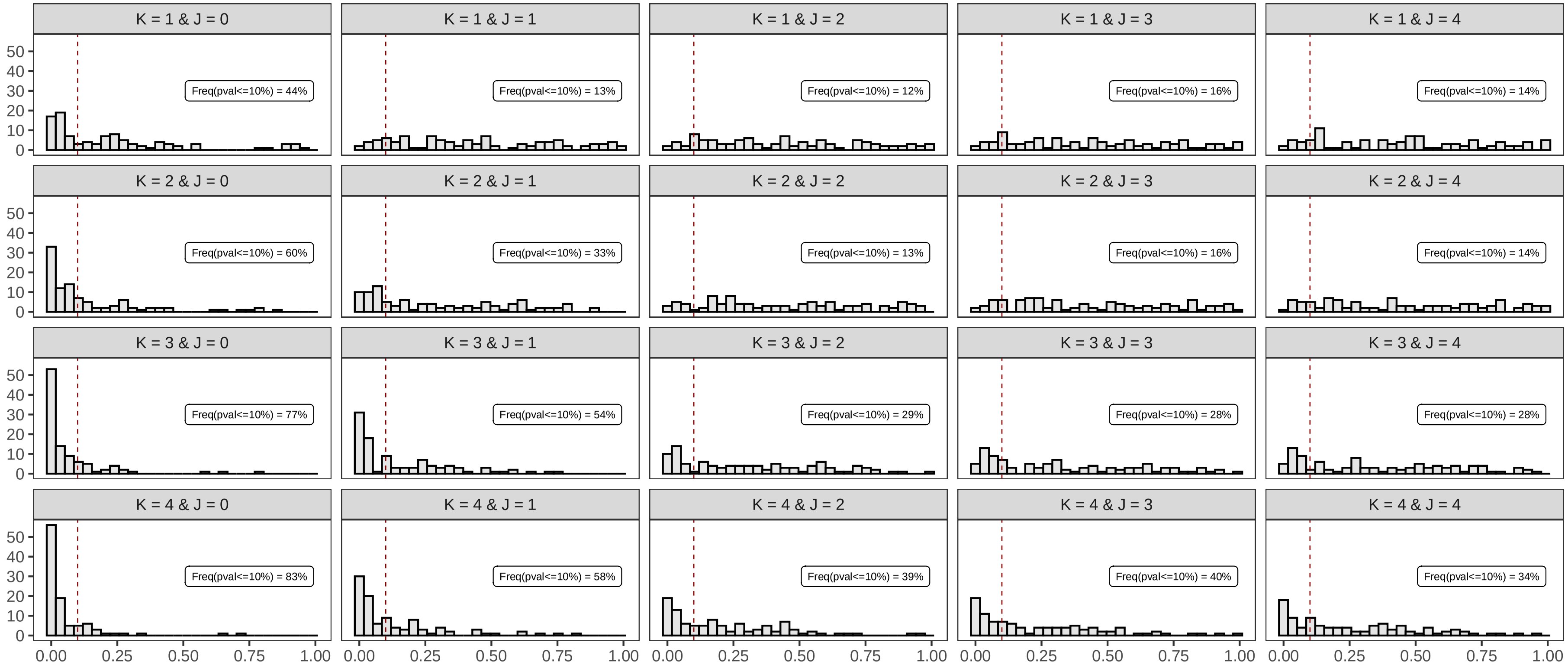

Figure 7: Histogram of GRS test p-value of the 3-by-3-by-3 ROE, P/B, and growth rate portfolios on market and characteristics-based factors when K = 1, 2, 3, and 4 and J = 1, 2, 3, and 4.

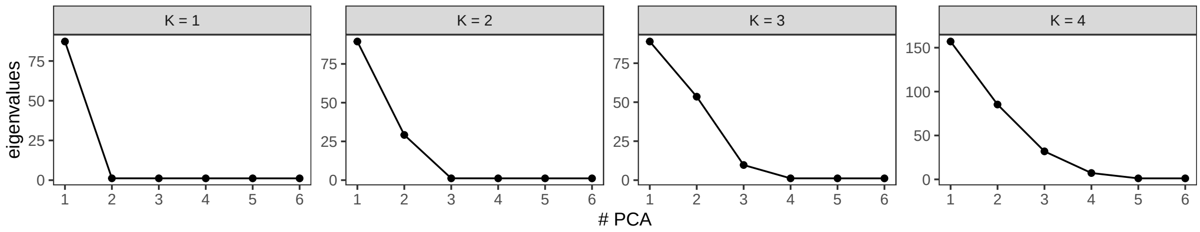

Figure 8: Eigenvalues of the Connor-Korajczyk factors when K = 1, 2, 3, and 4.

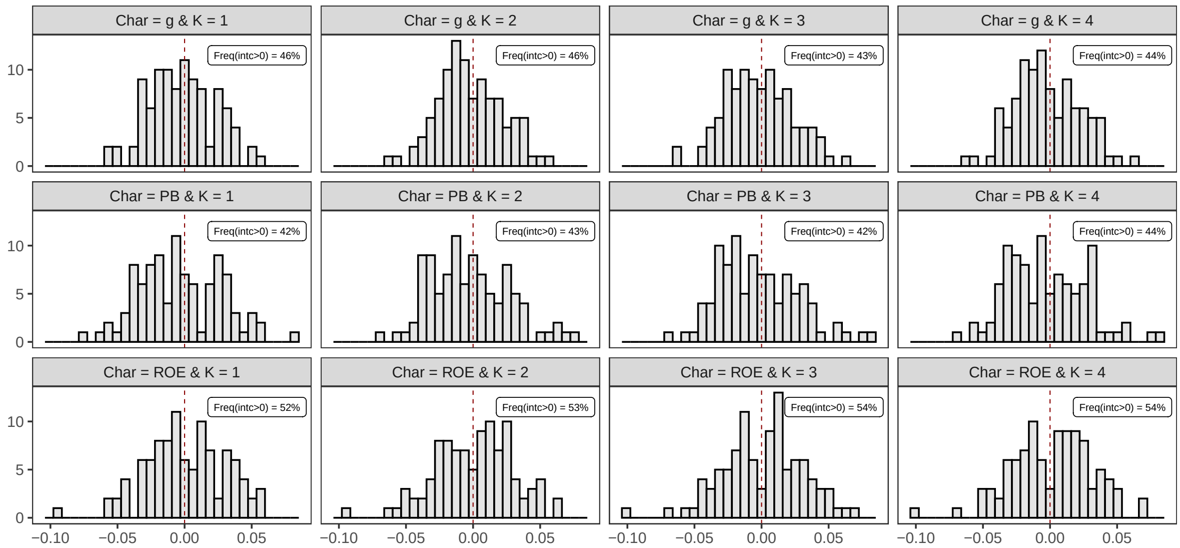

Figure 9: Histogram of the intercept of characteristics-based factors on the Connor-Korajczyk factors when K = 1, 2, 3, and 4.

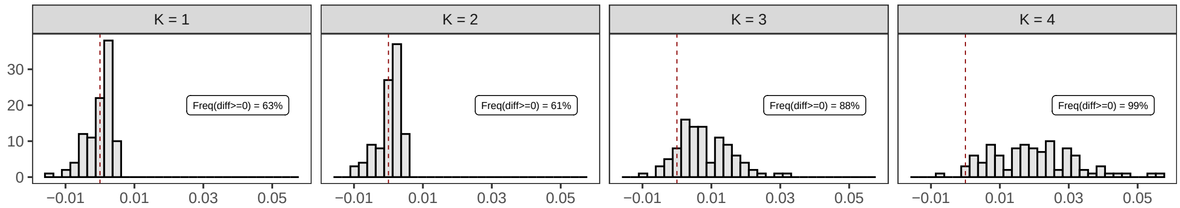

Figure 10: Histogram of the difference in squared Sharpe ratio between Connor-Korajczyk factors and market and characteristics-based factors when K = 1, 2, 3, and 4.

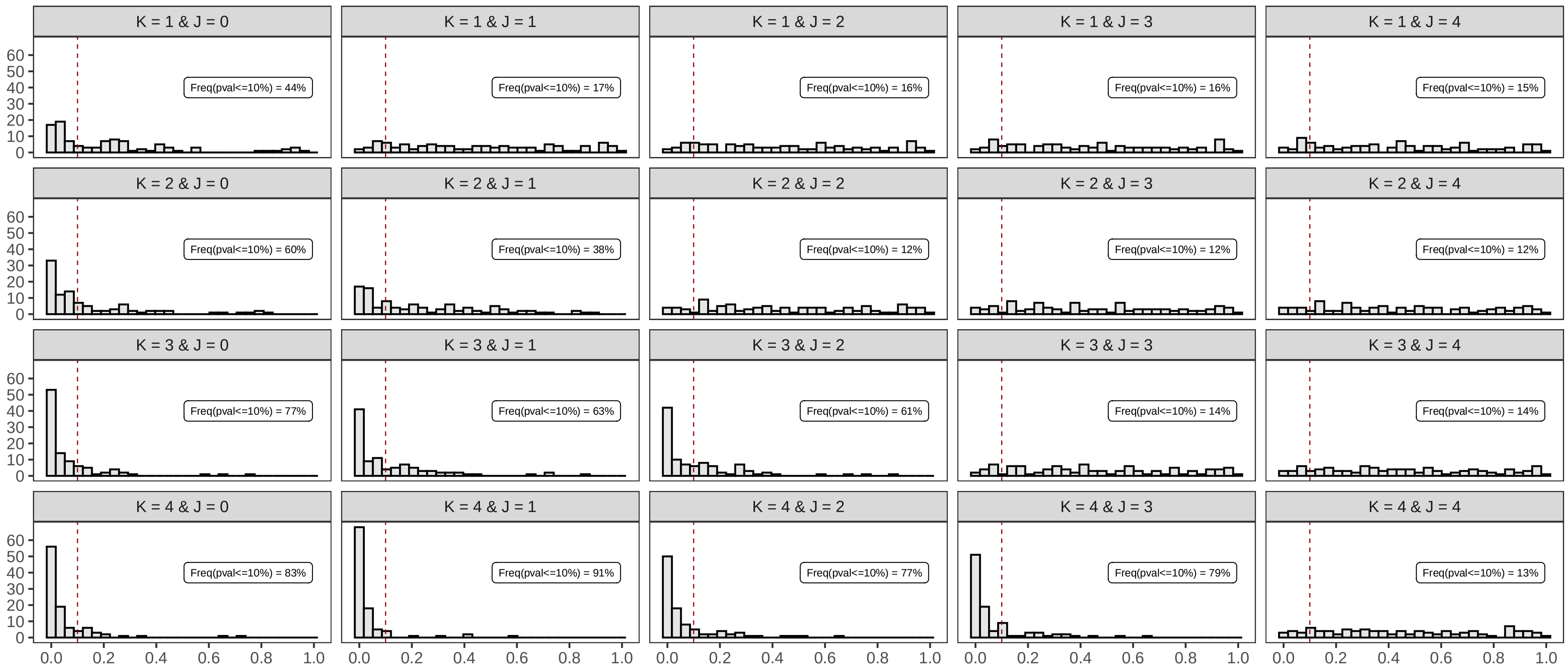

Figure 11: Histogram of GRS test p-value of the 3-by-3-by-3 ROE, P/B, and growth rate portfolios on Connor-Korajczyk factors when K = 1, 2, 3, and 4 and J = 0, 1, 2, 3, and 4.

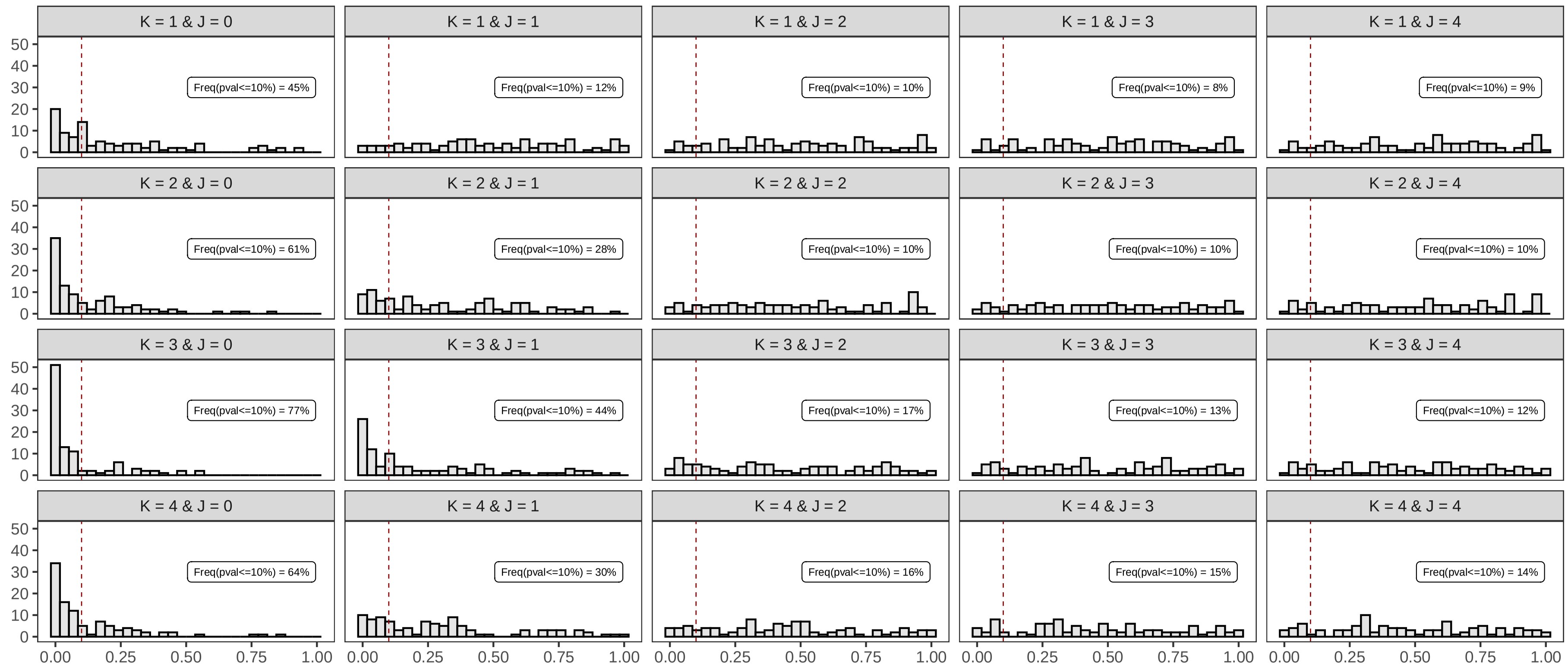

Figure 12: Histogram of GRS test p-value of the 3-by-3-by-3 ROE, P/B, and growth rate portfolios on market and characteristics-based factors when factor loadings are time-varying and K = 1, 2, 3, and 4 and J = 1, 2, 3, and 4.

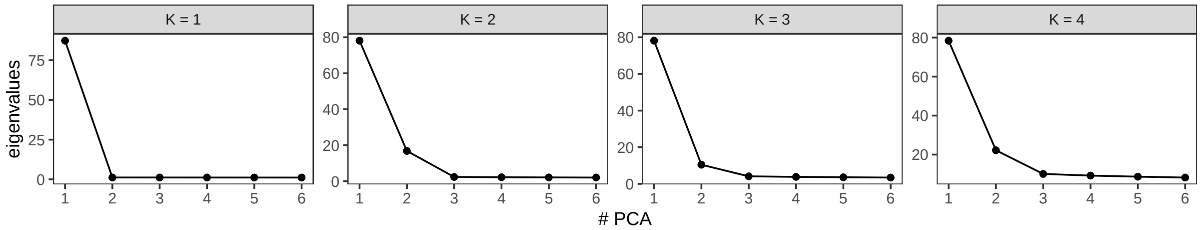

Figure 13: Eigenvalues of the Connor-Korajczyk factors when factor loadings are time-varying and K = 1, 2, 3, and 4.

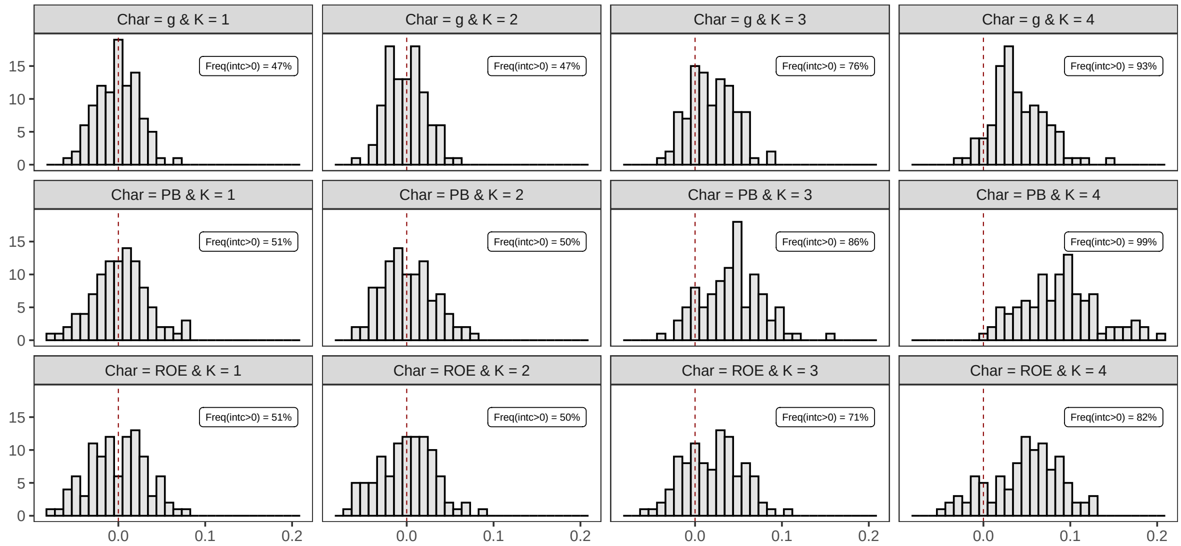

Figure 14: Histogram of the intercept of characteristics-based factors on the Connor-Korajczyk factors when factor loadings are time-varying and K = 1, 2, 3, and 4.

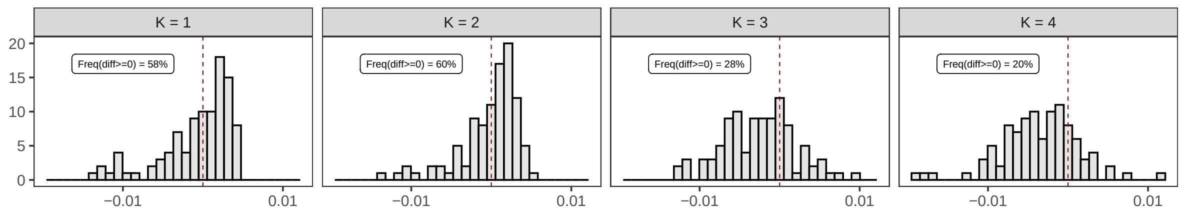

Figure 15: Histogram of the difference in squared Sharpe ratio between Connor-Korajczyk factors and market and characteristics-based factors when factor loadings are time-varying and K = 1, 2, 3, and 4.

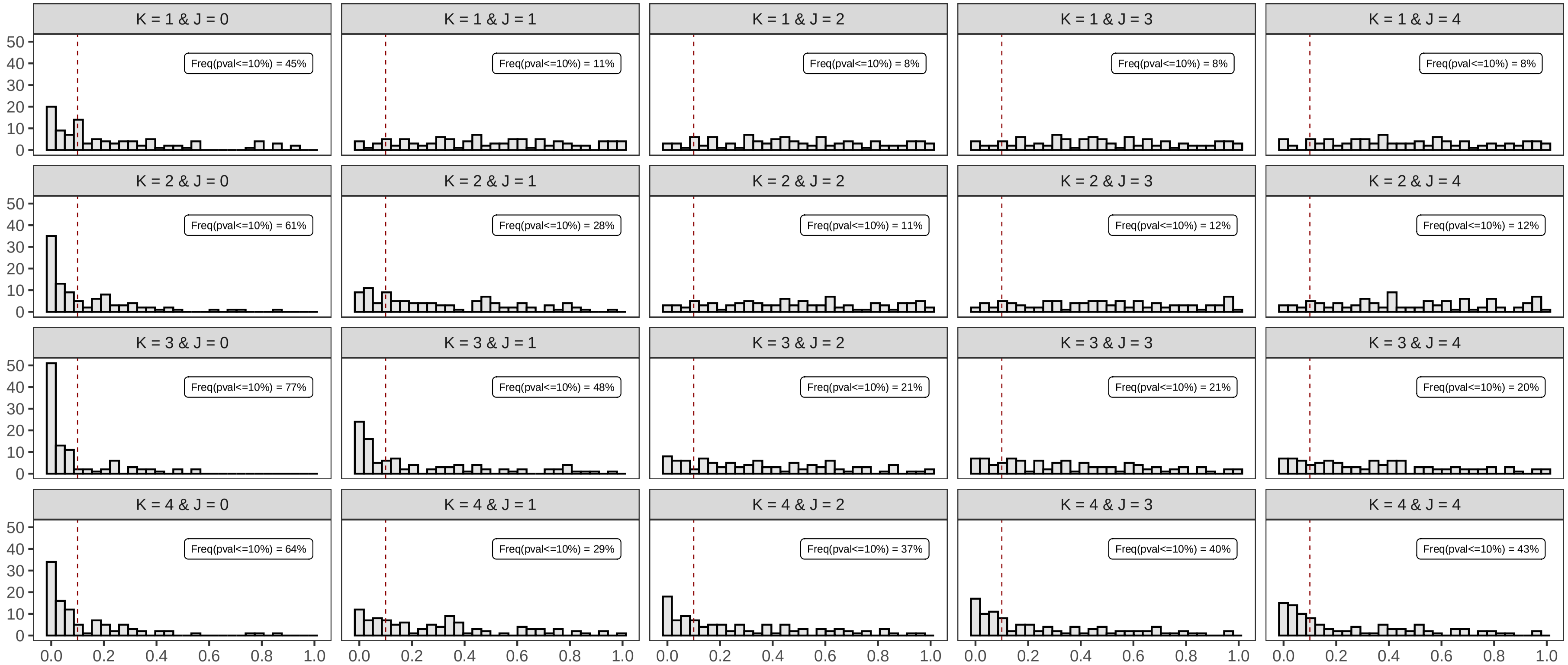

Figure 16: Histogram of GRS test p-value of the 3-by-3-by-3 ROE, P/B, and growth rate portfolios on Connor-Korajczyk factors when factor loadings are time-varying, K = 1, 2, 3, and 4 and J = 0, 1, 2, 3, and 4.

Table 2: Static relation between (monthly percentage) expected returns and characteristics

Panel A, B, and C are based on the simulation with 100 independently generated cross sections. Each cross section has 4,000 firms with randomly generated characteristics including ROE, expected return, and PBR, where ROE and expected return are drawn from normal distributions with means and standard deviations taken from firms from 2013 to 2022 in the CRSP/Compustat merge dataset, and PBR is drawn from uniform distributions from 0 to one half of the expected return divided by ROE. We remove firms with ROE outside 0% to 60% range or (annual) expected return outside 2% to 12% range, and we further exclude firms with (annual) expected return greater than two times of its ROE. The price-to-book (P/B) is computed based on GGM and the perpetual growth rate is ROE times PBR. Panel A summarizes the univariate and multiple regressions of the expected returns on the characteristics. In panel B and C, the firms are then assigned into groups based on univariate decile sort, 2-by-2-by-3 P/B, PBR, and ROE sort, 2-by-2-by-3 P/B, ROE, and g sort, and 2-by-2-by-2-by-3 P/B, PBR, ROE, and g sort of characteristics. The return of each group is the equal weighted average return and the return for each sorting strategy is the mean return of the top groups minus the mean return of the bottom groups based on the given characteristic. Panel B and C apply independent sort and conditional sort respectively, and report the mean, minimum, and maximum of the strategy returns based on the 100 independently generated cross sections. Panel D computes the average monthly percentage returns of similar strategies based on CRSP/Compustat merge data from July 1976 to June 2023. Fama and French (2015) measure investment rate as the change of total assets divided by lagged total assets (asset growth) which is close to ROE times PBR, equal to growth rate g, in the GGM formula. We use g to represent the characteristic asset growth. The book equity is measured following Fama and French (2006).

Panel A: Regressions of expected returns on the characteristics

	Univariate				Multiple			Multiple			Multiple			
	P/B	PBR	ROE	g	P/B	PBR	ROE	P/B	ROE	g	P/B	PBR	ROE	g
Mean	-0.03	0.64	0.04	8.40	-0.04	0.66	1.34	-0.04	1.00	6.14	-0.04	-0.03	0.99	6.30
Min	-0.03	0.57	-0.02	7.77	-0.05	0.58	1.26	-0.04	0.94	5.62	-0.04	-0.11	0.93	5.87
Max	-0.03	0.75	0.13	8.86	-0.04	0.75	1.43	-0.04	1.08	6.57	-0.04	0.04	1.07	6.69

Panel B: Independent sort

	Univariate decile sort				2-by-2-by-3 sort			2-by-2-by-3 sort			2-by-2-by-2-by-3 sort			
	P/B	PBR	ROE	g	P/B	PBR	ROE	P/B	ROE	g	P/B	PBR	ROE	g
Mean	-0.38	0.20	0.02	0.32	-0.39	0.17	0.28	-0.40	0.23	0.19	-0.39	0.06	0.26	0.25
Min	-0.41	0.15	-0.01	0.28	-0.42	0.14	0.26	-0.42	0.21	0.17	-0.41	-0.02	0.25	0.16
Max	-0.36	0.23	0.06	0.36	-0.38	0.19	0.30	-0.39	0.24	0.21	-0.37	0.11	0.28	0.30

Panel C: Conditional Sort

	2-by-2-by-3 sort			2-by-2-by-3 sort			2-by-2-by-2-by-3 sort			
	P/B	PBR	ROE	P/B	ROE	g	P/B	PBR	ROE	g
Mean	-0.32	0.15	0.23	-0.33	0.18	0.19	-0.32	0.04	0.16	0.14
Min	-0.34	0.13	0.22	-0.35	0.17	0.17	-0.35	0.03	0.14	0.12
Max	-0.31	0.17	0.25	-0.32	0.19	0.20	-0.27	0.06	0.18	0.15

Panel D: Results using CRSP/Compustat merged data (July 1963 to June 2023)

	Univariate decile sort			2-by-2-by-3 independent sort			2-by-2-by-3 conditional sort			2-by-3 Size-X independent sort		
	P/B	ROE	g	P/B	ROE	g	P/B	ROE	g	P/B	ROE	g
Mean	-0.27	0.29	-0.39	-0.31	0.42	-0.14	-0.27	0.26	-0.15	-0.27	0.30	-0.26
SD	4.75	4.18	3.46	3.35	2.27	2.13	3.00	1.90	2.15	3.02	2.21	2.05
t-stat	-1.53	1.83	-3.01	-2.46	4.94	-1.82	-2.45	3.68	-1.84	-2.40	3.65	-3.41