R&D, Expected Profitability, and Expected Returns

Amit Goyal **Sunil Wahal***

January 2023

Abstract

Current R&D expenditures forecast cash-based operating profitability up to three years in the future and sometimes as much as ten years, but do not forecast asset growth. High R&D firms have positive loadings on a cash-based operating profitability factor, and zero alphas. Capitalizing R&D to augment book values with intangible assets is unnecessary for asset pricing, so long as expected profitability is explicitly recognized as a determinant of expected returns.

JEL classification: G11, G12, G13

Keywords: Research & Development, Intangibles, Profitability, Asset Pricing, Expected Returns, Accruals

*Amit Goyal is from Swiss Finance Institute at the University of Lausanne, email: Amit.Goyal@unil.ch and Sunil Wahal is from WP Carey School of Business, Arizona State University, email: Sunil.Wahal@asu.edu. Sunil Wahal is a consultant to Avantis Investors. Avantis did not provide data or funding for this research. We thank the Center for Responsible Investing at ASU for financial support. We are grateful to Natalie Burclaff of the Library of Congress for providing access to archives of historical financial statements, and to Charles Kessler at S&P Global for resolving data issues associated with R&D in Compustat. We thank Jinsung Bach, Ethan Kibsey, Mathew Tanczos, and Alexandre Tilly for their diligent data collection efforts. We thank Juhani Linnainmaa for sharing data on cash-based operating profitability factor, Ken French for making data on Fama-French factors available on his website, and Lu Zhang for making data on q/q^5 factors available on his website. We thank Hank Bessembinder, Garen Markarian, and Seth Pruitt for helpful comments.

1 Introduction

Research and Development (R&D) is critically important to economic growth, most of which is done in the corporate sector. In fact, the National Science Foundation reports that in 2020, 73% of aggregate R&D expenditures were due to the business sector (see Table 6 at https://ncses.nsf.gov/pubs/nsf22320). Yet, the stock market's valuation of corporate R&D remains unsettled and controversial. An extensive literature argues that stock markets misprice R&D and innovation generated by R&D, which has potentially profound implications for allocative efficiency and aggregate growth.[1] The prototypical mispricing story is that investors underappreciate the value of R&D, resulting in low prices and high future (realized) returns. An equally large literature in both finance and accounting, starting with Lev and Sougiannis (1996), investigates issues surrounding the (lack of) capitalization of R&D. The oft-expressed concern is that since R&D is expensed, it does not show up in firm-level balance sheets and is, therefore, mispriced by markets. Capitalization of R&D, part of knowledge capital, is also central to the more recent intangibles literature which asks whether non-capitalized but risky investment expenditures are related to future returns and economic growth (see Corrado, Haskel, Jona-Lasinio, and Iommi (2022) for a summary of the issues in the growth literature).[2]

In this paper, we study the asset pricing implications of R&D. To provide discipline to the analysis, we invoke two "umbrella" models grounded in economic theory: the dividend discount model and the neoclassical q-theory of investment. The models are central to both the macroeconomic perspective as well as asset pricing. With dividend irrelevance (Miller and Modigliani (1961)) and clean surplus accounting, the dividend discount model can be written as:

$$\frac{M_t}{B_t} = \frac{\sum_{\tau=1}^{\infty} \mathrm{E}(Y_{t+\tau} - dB_{t+\tau})/(1+r)^{\tau}}{B_t}. \tag{1}$$

where M_t is market value at time t, $\mathrm{E}(Y_{t+\tau})$ is expected profitability, $\mathrm{E}(dB_{t+\tau})$ is expected investment (change in book value), and r is the stock's internal rate of return. In this model, the ability of any sorting variable to explain the cross-section of returns should be solely

[1] See, for example, Chan, Lakonishok, and Sougiannis (2001), Cohen, Diether, and Malloy (2013), Dong, Hirshleifer, and Teoh (2021), Eberhart, Maxwell, and Siddique (2004), Hirshleifer, Hsu, and Li (2013), and others.

[2] A partial list of papers that investigate issues related to capitalization and returns includes Arnott, Harvey, Kalesnik, and Linnainmaa (2021), Bongaerts, Kang, and van Dijk (2022), Eisfeldt, Kim, and Papanikolaou (2020), Eisfeldt and Papanikolaou (2013), Ewens, Peters, and Wang (2019), Gulen, Li, Peters, and Zekhnini (2021), Iqbal, Rajgopal, Srivastava, and Zhao (2022), Kapons (2020), Kazemi (2021), Lev and Srivastava (2019), Oswald, Simpson, and Zarowin (2022), and Park (2019). Many, but not all, of these papers ask whether intangible assets are responsible for the degradation in observed performance of high book-to-market ratio stocks. Peters and Taylor (2017) are not concerned with returns but their methodological approach to investigate the investment-q relationship is frequently used by others.

through its joint influence on book-to-market ratios, expected profitability, and expected investment. That is at the heart of the Fama and French (2015) five-factor model's ability to explain anomalies (Fama and French (2016)).

The q-theory of investment delivers similar intuition even though the primitives and derivation are different. Because firms invest until the expected return on investment is equal to the cost of capital, q-theory links expected profitability, investment, and expected returns. Relying on the production analogue of Cochrane (1991), the two-period model of Hou, Xue, and Zhang (2015) is written as:

$$R_{i,t+1} = \frac{X_{it+1} + 1 - \delta}{1 + aI_{it}/A_{it}}, \tag{2}$$

where X_{it} is the expected return on assets (which, multiplied by total assets, A_{it}, delivers expected operating profits), I_{it} is investment, a is an adjustment cost parameter, and δ is a depreciation rate. q-theory is agnostic to the nature of the investment and, in fact, Peters and Taylor (2017) show that both tangible and intangible investment such as R&D are equally well-explained by q-theory.

In a cross-sectional asset pricing framework, the shared common intuition of these models says that if firms invest in R&D with the expectation of increasing future profits, then this should be reflected in covariances (loadings) with a profitability factor. And if monetizing R&D requires future investment, then this should similarly be reflected in loadings on an investment factor. Given the underlying theoretical constructs, factor model alphas should be indistinguishable from zero. Mispricing implies non-zero alphas although, as always, zero alphas do not necessarily exclude mispricing because of the joint testing problem. In the above framework, there is no independent channel for R&D to be related to expected returns — beyond measurement error, all of the cross-sectional variation in returns should be absorbed by loadings on the profitability and/or the investment factor(s).

Empirical tests are burdened with two problems, bad data and bad models. We start with the bad data problem. Under the 1974 Statement of Financial Accounting Standards (SFAS2), firms are required to expense R&D in the year the expenditure is incurred. There is, however, considerable latitude for expense shifting so that firms can choose to bundle R&D into selling, general, and administrative (SG&A) expenses and not report it separately. The issue is well known; Koh and Reeb (2015) document that firms which do not separately report R&D are non-random and do so strategically. Non-reporting of R&D can, therefore, add both bias and noise to asset pricing tests. Indeed, in our 1975-2021 sample period, 51%

of firms (representing 40% of the aggregate market capitalization) do not report R&D.[3] We address the issue in two ways. First, we engage in an extensive manual data collection effort from electronic and microfiche financial statements. This allows us to fill in data holes in Compustat and more importantly, to unwind strategic non-reporting by identifying firms that conduct R&D but do not report it.[4] We then impute R&D for such firms using size and industry groupings. Our motivation for using these groupings is that it ties the imputation process to well-established industrial organization models in which firms invest in R&D in order to win patent races, stay competitive within the industry, or merely continue to exist (see Tirole (1988, Chapter 10) for a summary of this literature). As a practical matter, imputation allows us to deploy the entire cross-section in asset pricing tests, increasing power.

Our tests start with annual cross-sectional regressions that forecast firm-level future profitability and investment in the spirit of Fama and French (2006), with the adjustments recommended by Aharoni, Grundy, and Zeng (2013). The motivation is that current R&D should be related to future profitability and/or future investment. In these regressions, current industry-adjusted R&D scaled by book equity forecasts future cash profitability up to 3 years in the future, controlling for past profitability, book-to-market ratios, accruals, investment, and other such predictors. Scaling by market equity, as is done in studies that consider R&D to be mispriced, the predictability extends up to ten years. For large stocks, which conduct 80% of the aggregate value of corporate R&D and represent 90% of the aggregate market capitalization, this forecastability also extends up to 10 years in the future. In short, there is robust evidence that R&D fulfils its primary purpose of generating future profits, and does so over long horizons. In contrast, regressions of future investment on current R&D show no meaningful relation, regardless of the scaling variable (book or market equity) or horizon.[5]

We then move to asset pricing tests, confronting both the five-factor and q models head-on. Original versions of both models fail. Using the five-factor model, a portfolio of industry-adjusted high R&D/Book equity stocks has an annualized alpha of 1.92% with a t-statistic of 2.55; the alpha using the q model is 2.32% with a t-statistic of 2.70. The annualized alpha

[3] Bryzgalova, Lerner, Lettau, and Pelger (2022), Chen and McCoy (2022), and Freyberger, Höppner, Neuhierl, and Weber (2021) argue that dealing with missingness is important for asset pricing tests.

[4] We do so by reading financial statements and identifying firms that have research facilities or personnel that are part of ongoing operations, carry R&D tax credits on their books, report patents, and self-report the importance of R&D to the business. Details of procedure are in Appendix B.

[5] There are two types of R&D. Product R&D attempts to generate new products and improve the quality of existing ones. Process R&D aims to lower the cost of existing products. For either type, it does not automatically follow that converting R&D to future profits requires incremental investment in tangible capital because of variation in the nature of the production function, capacity utilization, etc. (see, for example, Cohen and Klepper (1996)).

of the high R&D/Market equity portfolio using the five-factor (q) model is 2.74% (4.15%), with a t-statistic of 2.77 (4.12). And in long-short portfolios of high-minus-zero R&D/Book equity firms, the alphas are larger still, 3.67% and 4.62% per year with t-statistics of 3.32 and 3.31 respectively. For the five-factor model, the 'culprit' is the loading on profitability. High R&D stocks are highly profitable but have a negative loading on RMW, the factor used to measure expected profitability. For the q model, there are two culprits. High R&D stocks have a negative loading on ROE, the parallel factor used to measure expected profitability. Equally damaging is the loading on I/A, the factor used to measure investment. Despite the fact that high R&D stocks have low current investment, the loading on I/A is negative. In both models, it is these loadings that drives a wedge between realized and model implied returns.

Are these symptoms of mispricing or a bad model problem? For the five-factor model, the answer resides in accruals. High R&D firms have substantially larger (negative) accruals. Ball, Gerakos, Linnainmaa, and Nikolaev (2016) purge accruals from operating profits and point out that the cash-based component of profitability explains the cross-section of expected returns better than gross profit (Novy-Marx (2013)), or operating profitability (used in the construction of RMW). Using the Barillas and Shanken (2017) method, Fama and French (2018) come to the same conclusion, showing that RMW constructed from cash profitability (RMW_{CP}) outperforms other constructions. Replacing RMW with RMW_{CP}, we observe two facts. First, the profitability loading for high R&D firms flips from being negative for RMW to positive for RMW_{CP}, mirroring portfolio characteristics. Second, the alphas are indistinguishable from zero. The message, in this case, is that the bad model problem is not that bad, so long as one uses cash profitability to represent economic profitability.

The answer for the q model is more complex. With respect to expected profitability, ROE is measured using income before extraordinary items and is also contaminated by accruals.[6] Replacing ROE with RMW_{CP}, the loading on profitability changes sign and the alphas are indistinguishable from zero. There remains, however, the mismatch between current investment and the loading on the I/A factor. The two-period version of the model focuses on marginal investment relative to current investment, but R&D has implications for investment into the more distant future, pointing to a limitation of the motivating theoretical model. The multiperiod extension that appears in Hou, Mo, Xue, and Zhang (2021), referred to as the q^5 model, accommodates this intuition by adding an expected investment growth term that represents the growth in marginal q. Evaluating R&D relative to such an augmented

[6] In their Table 4, Hou, Xue, and Zhang (2015) report that alpha for the accruals anomaly using the q model is -0.56 with a t-statistic of -3.90, and point out that the "pattern in ROE factor loadings goes in the wrong direction."

model lays bare the limitations of the two-period version as well as the advantage of the multi-period setting. Consider, for example, the case of a high R&D firm attempting to monetize the fruit of R&D. If doing so requires investment in tangible capital beyond the first year, the two-period q model is unable to capture investment growth (unlike the dividend discount model which includes infinite horizon expected profitability and investment). The multi-period q^5 model attempts to do precisely that.

Augmenting the q model with the expected growth factor helps resolve the original q model's mispricing — to a degree. For high R&D portfolios, the loading on expected investment growth is sharply positive, offseting the lower expected returns generated by ROE and I/A, and rendering the alpha indistinguishable from zero. That is good news for the q^5 model. But vindication is incomplete. High R&D firms continue to have a negative loading on ROE despite being high profitability firms. As before, we can resurrect this aspect of the model by replacing ROE with RMW_{CP}. But the positive loading on expected investment growth still does not match the annual cross-sectional regressions described earlier in which current R&D is unrelated to future investment. The q^5 model thus performs better than the q model, but using a channel that is inconsistent with the data.

There remains the issue of capitalization. A typical approach of the studies focused on intangibles (cited in footnote 2) is to compute an intangible-adjusted book value, and then compare the performance of an adjusted value factor (iHML) to the original value factor (HML). The underlying idea is that capitalization fixes the mismeasurement in book values, thereby improving the performance of the value factor. But this approach ignores the possibility that R&D can affect discount rates through future profitability, and most importantly, that expected returns in these models are jointly determined by combinations of book-to-market ratios, expected profitability and expected investment. In fact, Fama and French (2006, pg. 494) make this point explicitly, stating that it is important to "emphasize the importance of *joint* controls for the three variables" (emphasis ours). Penman (2009) makes a related observation, reiterating that a fundamental aspect of accounting is that under clean surplus principles, the income statement mitigates poor balance sheet information via the "cancelling error property," and reminding readers that "there is also an income statement."[7] Following this line of reasoning, we capitalize R&D, and then ask whether the intercepts of capitalized R&D portfolios are non-zero after controlling for cash-based operating profitability. The answer is no: loadings on RMW_{CP} line up with the profitability of capitalized R&D portfolios and drive alphas to zero. We conclude that capitalizing R&D

[7]Penman's pithy statement echoes Graham and Meredith (1937, pg. 36), who state "... intangibles may have very large value indeed, but it is the income account and not the balance sheet that offers the clue to this value. In other words, it is the earning power of these intangibles, rather than their balance-sheet valuations, that really counts."

is unnecessary for asset pricing so long as one explicitly recognizes (cash) profitability as a determinant of expected returns.

Finally, how can one reconcile our results with the mispricing literature that uses (non-capitalized) R&D intensity? Chan, Lakonishok, and Sougiannis (2001) find that high R&D stocks (scaled by market equity) have high future excess returns and interpret their results as mispricing. Eberhart, Maxwell, and Siddique (2004) find positive long-run returns associated with firms that increase R&D expenditures and similarly interpret their results as evidence of mispricing. Li (2011) argues that the relation between R&D intensity and future returns is driven entirely by financially constrained firms, while Resutek (2022) says that the cause is the fixed cost of R&D. Donelson and Resutek (2012) suggest that the high returns of R&D firms are not due to R&D at all but because investors incorporate other value-relevant information into price. There is also a strand of the literature that looks at a non-profitability measure of the output of R&D, viz. innovation. Hirshleifer, Hsu, and Li (2013) report large positive future returns to portfolios formed on patents scaled by R&D, which they attribute to mispricing (see also, Dong, Hirshleifer, and Teoh (2021)). Cohen, Diether, and Malloy (2013) differentiate between "good" and "bad" R&D firms by estimating firm-by-firm regressions of sales growth on lagged R&D; because good R&D firms earn higher returns than bad R&D firms, they conclude that the market misvalues innovation. In each of these approaches and studies, what is missing from pricing equations is expected profitability. From an asset pricing perspective, it matters.

The remainder of the paper is organized as follows. Section 2 briefly describes the issues associated with R&D data as well as our fixes. Details of R&D calculations are relegated to Appendix B. Section 3 reports on annual cross-sectional regressions that forecast future profitability and investment. Section 4 describes the baseline asset pricing tests assessing the models as well as amended versions. Section 5 addresses the issue of whether capitalizing R&D is a fruitful from an asset pricing perspective. Section 6 addresses robustness issues and Section 8 concludes.

2 Data

2.1 Sample Construction

Our sample consists of all NYSE, Amex, and Nasdaq firms with share codes 10 and 11 from the CRSP-Compustat Merged Database using LinkTypes `LC` and `LU`, and requiring that the fiscal period end date be within the link date range. To ensure that we correctly reflect

Compustat's data collection processes described in Appendix B (particularly those that deal with the materiality of R&D), we use information from data items, supplementary data codes, and footnote codes. The sample begins in July 1975 (after the issuance of SFAS2) and ends in December 2021.

We construct book values (B) in a manner similar to Fama and French (1993). We compute operating profitability (OP) as sales minus cost of goods sold minus SG&A. We compute cash-based operating profitability (CP) following the balance sheet method of Ball, Gerakos, Linnainmaa, and Nikolaev (2016). We compute accruals (AC) using balance sheet and income statements items following Sloan (1996). Further details on variable construction are provided in Appendix A. All accounting information is assumed to be known six months after the fiscal year-end.

Our tests require industry adjustments. CRSP and Compustat suffer from missing SIC codes and, even with non-missing data, frequently disagree. Kahle and Walkling (1996) find Compustat codes to be more reliable than CRSP codes. Ken French also uses Compustat SIC codes to construct industry portfolios available on his website. Therefore, we construct the full time series of SIC codes for each firm, giving preference to Compustat's historical code (`SICH`), Compustat's header code (`SIC`), CRSP's historical code (`SICCD`), and CRSP's header code (`HSICCD`), in that order.

2.2 R&D Reporting Rules and Biases

The 1974 Statement of Financial Accounting Standards No. 2 (SFAS2) requires that firms expense R&D in the year that the expenditure is incurred. Disclosure of material R&D expenditures is required, where the materiality threshold is generally 1% of sales. Disclosure can take place in the income statement, a footnote, and sometimes in the Notes to Financial Statements. Despite this intent, Generally Accepted Accounting Principles (GAAP) allow for discretion in the classification of expenses so that firms can, should they choose to, classify R&D expenses inside core expenses such as SG&A. As McVay (2006) points out, such classification shifting is not the same as accruals management because it does not alter GAAP earnings. Nonetheless, classification shifting can be strategic in nature, implying that not reporting R&D could be non-random. Sun (2021) reports that classification shifting can have important valuation consequences because reductions in SG&A are viewed positively but reductions in R&D are viewed negatively. More directly, Koh and Reeb (2015) show that the number of firms that do not separately report R&D is large, and that such non-reporting R&D is strategic in nature: non-reporting firms file 14 times more patent applications than zero R&D reporting firms, and when subject to an exogenously imposed change in auditors, non-reporting R&D firms initiate reporting.

Classification shifting and selective but strategic disclosure has important, but largely ignored, implications for asset pricing tests. In studies that utilize R&D, researchers either restrict their attention to stocks with valid R&D (e.g., Cohen, Diether, and Malloy (2013), Lev and Sougiannis (1996), and others), or simply replace missing values with zeroes (e.g., Peters and Taylor (2017), and others who use their method). Bryzgalova, Lerner, Lettau, and Pelger (2022), Chen and McCoy (2022), and Freyberger, Höppner, Neuhierl, and Weber (2021) point out that non-random missing data can cause inference problems because stock returns often vary systematically with missingness. These authors develop econometric solutions suited for high dimensional machine learning purposes, where missingness is especially vexing because it occurs in a hundreds of explanatory variables. Our problem is confined to a single variable for which we have strong economic priors. The priors come from an extensive and long-standing literature in industrial economics in which firms engage in R&D to generate temporary monopolies, with patent races, compete with other firms in their industry, or ensure survival. The essence of this literature, which originates with Schumpeter (1943), is that *within-industry* competition is critical to understanding R&D and innovation. In the Aghion, Bloom, Blundell, Griffith, and Howitt (2005) model, for example, incentives to spend on R&D depend crucially on the difference between post-innovation and pre-innovation rents within an industry. Since the fraction of industries with such "neck-and-neck" competition is itself endogenous, R&D intensities vary systematically across industries. Similarly, in the Agarwal and Gort (2002) model, a firm increases its likelihood of survival within an industry by adding to its endowment via R&D. Given this type of micro-foundation, we rely on a supplemental data collection effort that either directly corrects the data, or provides the foundation for an imputation process linked to industry-specific information.

2.3 Addressing Missingness

We address the data missingness approach using a two pronged approach. The details are in Appendix B, but we provide a brief summary of the processes below.

In the first step, we hand collect R&D data from financial statements downloaded from the Mergent database and using the microfiche records at the Library of Congress (LOC). We hand-collect R&D expenditures when they are available, and when they are not, record evidence of the presence of R&D activity.[8] This evidence appears in the form of language

[8] Financial statements often provide R&D information for three years adjacent to the reporting year. Compustat sometimes only records data as of the fiscal year of the financial statement, ignoring adjacent years. That allows us to fill some holes in the time series.

discussing the presence of research facilities, research scientists and personnel, specific R&D tax credits, verbiage that emphasizes the importance of research to the business, the presence of patents, etc. Upon observing such markers, we flag the firm for the second step.

In the second step, we first compute the median R&D/Sales ratio for small and large stocks within an industry-year (we use the NYSE median as a size breakpoint). We primarily use two-digit SIC codes to compute medians. If the aforementioned flag is recorded, we replace missing values with the above-calculated median. If the median is not calculable, we replace missing values with zero, mirroring the rest of the R&D literature. For firm-years for which the flag is not recorded, we follow the same procedure except that we require a minimum number of firms to calculate the median (10 and 5 firms for small and large stocks respectively).

Table 1 illustrates the importance of addressing the non-reporting of R&D. For each two-digit SIC code with at least 10 firms, we report the average number of firms with non-missing R&D, the number of firms after the addition of supplementary data, and the number of firms after imputation. It is clear that for some industries, addressing the missingness problem is consequential. For example, in the Communications industry (SIC = 48), on average, 22 firms report R&D expenditures. Our supplementary efforts take the sample up to 56 firms, and after imputation, we are able to deploy 114 firms. However, we caution readers to not interpret this process as always adding non-zero R&D. Consider, for example, Depository Institutions (SIC = 60). On average, there is only one firm that reports R&D, and our supplementary efforts only add two more firms. The imputation of an additional 415 firms to take the sample to 419 essentially adds zero R&D to firms in this SIC code.

2.4 Measurement and Scaling Issues

The early accounting literature typically scales R&D by sales, as do industry reports and data providers like S&P. Because of this, Chan, Lakonishok, and Sougiannis (2001) also initially scale R&D by sales but then argue for scaling R&D by the market value of equity. In doing so, they note that scaling by market equity "highlight(s) stocks that have large R&D spending and at the same time relatively depressed market values." The majority of subsequent studies follow their lead, even though scaling by price induces a correlation with future expected returns.

In our primary tests, we deflate R&D by book equity. We do so for two reasons. First, the dividend discount model in equation (1) scales both sides of the equation by book equity. Since R&D is presumably related to future profitability, taking the model seriously demands that we deflate by book rather than market equity. Second, to the extent that the left hand side of the model (i.e., book-to-market ratios), already captures "depressed market values," deflating R&D by book equity avoids double counting (a similar argument applies to the q model since average q can be approximated to a market-to-book ratio). That said, there are gains to be able to compare our results with the rest of the literature that scales by market equity. To allow for such comparability, we report results using both scaling variables.

There is one other important measurement issue. Since industry effects are so important, we compute industry-adjusted R&D ratios by subtracting the industry median in that year (based on two-digit SIC codes). As described earlier, the adjustment ties the empirical procedures to industrial organization models in which productive R&D is necessary to compete within an industry.

2.5 Aggregates

Before proceeding to asset pricing tests, it is useful to get a perspective on the magnitude of corporate R&D, the time series variation thereof, and the importance of imputation. Figure 1 shows aggregate R&D scaled by aggregate book equity over time. The graph shows this ratio using the original Compustat data, post-supplementary data collection, and after imputation.

Three salient features of the data stand out. First, the supplementary data collection is most influential in the early part of the sample period, roughly pre-1998, where collecting all the information in financial statements seems to be important. Second, imputation is relevant over the entire period but especially in the latter part of the time series. Finally, in terms of levels, aggregate R&D as a fraction of aggregate book equity rises from about 2.5% in 1975 to almost 7.0% in 2021.

3 Does R&D Forecast Future Profitability and Investment?

We estimate annual cross-sectional regressions predicting future profitability ($Y_{t+\tau}$) and investment ($dA_{t+\tau}/A_{t+\tau-1}$) in the spirit of Fama and French (2006). The regressions predict profitability or investment up to ten years ahead ($\tau = 1, \ldots, 10$). For the profitability regressions, we use both operating profitability ($OP_{t+\tau}$) and cash-based operating profitability ($CP_{t+\tau}$). Following Aharoni, Grundy, and Zeng (2013), the regressions are estimated at the firm-level rather than on a per-share basis, and the scaling variable for future profitability is book equity in the same year ($B_{t+\tau}$). Using a contemporaneous scaling variable ensures that profitability growth is not baked into the regressions.[9] The control variables include past profitability (Y_t), a separate indicator for firms with negative profitability ($Y_t < 0$), annual stock returns ($R1Y_t$), the logarithm of market equity ($\log M_t$), the logarithm of book-to-market ratios ($\log B_t/M_t$), accruals relative to book equity for firms with positive and negative accruals ($+AC_t/B_t$ and $-AC_t/B_t$, respectively), and investment (dA_t/A_{t-1}). Our interest is in the ability of industry-adjusted R&D scaled by book or market equity ($R\&D_t/B_t$ or $R\&D_t/M_t$) to predict future profitability and/or investment. All explanatory variables are measured in year t.

Table 2 reports average coefficients and Newey-West t-statistics from the regressions. To prevent information overload, we only display results for forecasts out to five years (i.e., $\tau = 1, \ldots, 5$). Panel A1 of Table 2 uses the full cross-section of stocks and deflates R&D by book equity. The first set of columns contain profitability regressions that predict $OP_{t+\tau}/B_{t+\tau}$. The coefficients on the standard explanatory variables mirror the results reported in Aharoni, Grundy, and Zeng (2013) and Fama and French (2006). For example, firms with low book-to-market ratios have higher future profitability, and past profitability is a strong predictor of future profitability. Given these well-known patterns, we do not dwell on these variables but instead focus on R&D. For OP regressions, the coefficient on $R\&D_t/B_t$ is reliably positive for up to three years in the future, with t-statistics of 4.63, 3.45, and 2.88 corresponding to $\tau = 1, 2$, and 3, respectively. The second set of columns shows results for predicting $CP_{t+\tau}/B_{t+\tau}$. As with OP regressions, $R\&D_t/B_t$ is positively related to cash profitability in the following three years with only marginally smaller t-statistics.[10]

To gauge economic significance, we compute the implied change in profitability for firms in the 10th versus the 90th percentile of industry-adjusted R&D. That spread is roughly 0.20, which multiplied by the coefficient on R&D for $\tau = 1$ in the OP or CP regression, implies a difference in future profitability of about 2.6%. There are two ways to put that into perspective: (a) it represents about 10% of the profitability of the median firm, (b) past profitability, the most important predictor of future profitability, is associated with a difference of profitability of about 20%. The effect of R&D is, therefore, about 13% of the marginal impact of past profitability.[11] In our view, these quantitative effects are both

[9]Fama and French (2006) use book equity in year t as the scalar for all future horizons but as they point out, for $\tau > 1$ the dependent variable becomes a mix of profitability and profitability growth. Our interest is in the effect of R&D in year t on profitability in each subsequent year rather than cumulatively. We can report, however, that scaling by book equity in year t results in identical inferences with respect to R&D.

[10]From untabulated regressions, we can report that the difference in the average adjusted R^2 of regressions with and without R&D are positive in every specification. The implication is that incremental explanatory power of R&D does not come at the expense of other variables, especially lagged profitability.

[11]In comparable one-year regressions, Aharoni, Grundy, and Zeng (2013) report that the effect of past profitability between the 10th and 90th percentile on future profitability is 22.3%, quite close to our estimate.

plausible and economically meaningful.

The last set of columns contain investment regressions. As is typical in these types of regressions, investment is negatively related to book-to-market ratios and accruals. More importantly from our perspective, the coefficients on R&D bounce around. In the first year ($\tau = 1$), the coefficient is negative but the t-statistic of 2.02 is hardly overwhelming. For $\tau = 2, 3$, and 5, the t-statistics are well below 2.00; the t-statistic for $\tau = 4$ is positive (2.17), but it too does not inspire confidence. Our view of the regressions is that there is no robust evidence that R&D in year t is associated with future investment growth.

Panel A2 of Table 2 replicates the above regressions using R&D scaled by market equity, $R\&D_t/M_t$. The coefficient on R&D is positively related to future OP and CP for all five years. In fact, we can report that for OP, t-statistics remain above two up to six years in the future, and for CP, up to nine years. We recompute economic magnitudes in the same manner as above. The spread in R&D is about 0.08, which multiplied by the (larger) coefficient on R&D for $\tau = 1$ in the OP regression, also generates a difference in profitability of about 2%. Since the coefficient on R&D in the CP regression is also higher, the implied effect is also larger, about 3%. Interestingly, there is a negative effect of R&D on investment for $\tau = 1$ and 2, but beyond that, we observe no statistical significance.

Separate regressions for large and small firms are useful for two reasons. First, the majority of aggregate R&D in the corporate sector is done by large firms. In our data, 78% of aggregate R&D is done by large firms. Second, large firms are only 20% of the total number of firms but represent 90% of the aggregate market capitalization. Since equal-weighted regressions give an inordinate amount of weight to small stocks, it is important that the ability of R&D to forecast future profitability not be restricted to small stocks. The large firm regressions are reported in Panels B1 and B2 of Table 2, respectively.[12]

Some interesting differences emerge. Scaling by book equity (Panel B1 of Table 2), there is no relations between R&D and future OP. But using cash profitability makes all the difference: R&D forecasts CP for all five years. In fact, we observe positive coefficients on R&D for all 10 years with little diminution in magnitudes over time. As before, there is no evidence that R&D forecasts future investment. Scaling by market equity (Panel B2 of Table 2) and using OP, R&D forecasts OP up to five years in the future. And using CP, the coefficients are reliably different from zero for up to 10 years in the future. If anything, the evidence that R&D forecasts future profitability is stronger in large stocks.

[12] To conserve space, we do not present regressions for small firms in a table. But we can report the following: the coefficients on R&D are positive using both OP and CP for up to three years, but not thereafter thereafter. With respect to investment, we continue to find no evidence that R&D is associated with increased in future investment.

In summary, there is persuasive evidence that current R&D forecasts future cash profitability.[13] This suggests that the channel by which R&D could show up in expected returns returns is through expected profitability. In contrast, there is no evidence that current R&D is related to future investment growth.

4 R&D and the Cross-Section of Returns

4.1 Portfolio Formation and Characteristics

We form five portfolios based on industry-adjusted R&D scaled by book or market equity at the end of each June, rebalancing annually. Since zero R&D is special, we isolate such firms in their own portfolio. The remainder firms are divided into quartiles based on NYSE breakpoints. The result is five portfolios that are, by construction, unequal in terms of the number of securities and the distribution of market capitalization.

Panels A and B of Table 3 display portfolio characteristics for portfolios sorted on R&D scaled by book and market equity, respectively. To compute characteristics that are ratios, we sum both the numerator and denominator of each ratio. For example, to compute the portfolio book-to-market ratio, we sum the total book value of stocks in a portfolio and divide by their total market value. This representation of portfolio aggregates, following Fama and French (1993), sidesteps issues related to weighting schema within a portfolio.

In Panel A of Table 3, the zero R&D portfolio contains an average of 1,971 firms, which represent about 40% of the aggregate market capitalization. The high R&D portfolio contains 773 firms, representing 18% of the aggregate market capitalization but 52% of the aggregate value of R&D. The intermediate portfolios (labelled Low, Q2, and Q3) are also well diversified and contain reasonable fractions of the aggregate market capitalization and R&D. Using market equity as a scaling variable (Panel B of Table 3) does not make a difference in terms of the number of stocks in each portfolio. There are, however, some differences in the percentage of the aggregate market capitalization in the high R&D portfolio. When scaling by book equity, this portfolio represents 18.3% of the aggregate market capitalization but scaling by market equity reduces it to 10.3%. In both cases, the high R&D portfolio represents the majority of aggregate R&D.

[13] Our interest in cash profitability is from an asset pricing perspective rather than the smoothing and matching of expenditures to profits that is of interest to the accounting profession. Thus we have nothing to say about accruals beyond the empirical observation that R&D forecasts cash profitability more so than operating profitability (which includes accruals). For interested readers, in unreported regressions analogous to those in Table 2, current R&D forecasts future accruals scaled by book equity. For large stocks (and scaling R&D by market equity), the forecasting power extends to at least five years.

The remainder of the columns show portfolio characteristics that are important for evaluating factor models. Book-to-market ratios vary systematically across the R&D portfolios in interesting ways. In Panel A of Table 3, high R&D firms are largely growth stocks while Zero R&D firms are mostly value stocks. When scaling by market equity, however, the high R&D portfolio consists largely of value stocks. This is consistent with evidence in Chan, Lakonishok, and Sougiannis (2001) who report that high R&D stocks have low past returns. Absent changes in book value, price declines make them value stocks, precisely what we observe; the average one year return ($R1Y$) for stocks in this portfolio is -0.8%, substantially lower than for other portfolios. Despite these price declines, the high R&D portfolio consists of highly profitable stocks, based on both on OP and CP. They also tend to be firms with low current investment, especially when compared with the low R&D portfolio. Perhaps most notably, firms in the high R&D portfolio have much larger (negative) levels of accruals. Indeed, accruals change monotonically across the portfolios offering the first hint of the importance of cash-based profitability in subsequent tests.

4.2 Evaluating the Umbrella Models

Table 4 evaluates the original five-factor and q models against the value-weighted excess returns of these portfolios. Panel A contains loadings and alphas from monthly regressions of excess portfolio returns of portfolios sorted on R&D scaled by book equity.

In the five-factor model, the loadings on HML and CMA align with the portfolio characteristics in Table 3: high (zero) R&D stocks are growth stocks that have negative (positive) loadings on HML, and are also low (high) investment stocks that have a positive (negative) loading on CMA. However, the loadings on RMW are exactly the opposite of what one would expect. The zero R&D portfolio has low operating profitability but a positive loading on RMW (0.17, t-statistic = 8.16). Equally damaging, the high R&D portfolio has a negative loading on RMW (-0.16, t-statistic = -5.41), despite consisting of highly profitable stocks. These miscues drive the (annualized) alphas, which are -1.74% (t-statistic = -3.17) for the zero R&D portfolio and 1.92% (t-statistic = 2.55) for the high R&D portfolio.

We also report loadings and alphas for three types of spread portfolios: high-minus-zero, high-minus-low, and high-minus-Q3 (the portfolio adjacent to the high R&D portfolio). The high-minus-zero R&D portfolio has an annualized alpha of 3.67% (t-statistic = 3.32). Because the low R&D portfolio has loadings that align with portfolio characteristics (and has a zero alpha), the high-minus-low R&D portfolio also has an alpha that is indistinguishable from zero. That is not the case, however, for the high-minus-Q3 portfolio. In fact, the alpha for that portfolio is 4.00% (t-statistic = 3.73).

The right section of Panel A of Table 4 shows results for the q model. Here the loadings on both the investment (I/A) and the profitability factor (ROE) are problematic. Stocks in the high R&D portfolio have low levels of investment, yet the loading on I/A is negative (-0.28, t-statistic $= -7.41$). Moreover, the portfolio consists of highly profitable stocks but the loading on ROE is also negative (-0.08, t-statistic $= -2.86$). The result is an alpha of 2.32% (t-statistic $= 2.70$), about 40 basis points higher than for the five-factor model. A similar problem occurs in the zero R&D portfolio, which has an alpha of -2.30% (compared to -1.74% for the five-factor model). As a result, the high-minus-zero R&D portfolio has an alpha of 4.62% (t-statistic $= 3.31$).

The original five-factor model fares even worse when portfolios are formed on R&D scaled by market equity (Panel B of Table 4). There are two sources of failure. First, despite the fact that the high R&D portfolio consists of value stocks, the loading on HML is negative (-0.09, t-statistic $= -2.49$). Second, as before, the zero R&D and high R&D portfolios have loadings that are the opposite of what one would expect. The combination is fatal from the perspective of the model because the alphas are even larger for book equity scaled portfolios. In the high R&D portfolio, the alpha rises from 1.92% when using book equity as a deflator, to 2.74% when using market equity. As a result, the high-minus-zero spread portfolio now has an alpha of 4.47% per annum.

Scaling by market equity also exacerbate the problems for the q model, albeit in complex ways. In the high R&D portfolio, the loading on I/A is no longer negative (it is now 0.02 with a t-statistic of 0.49), but the loading on ROE is even more negative (-0.25, t-statistic $= -7.81$). That drives the alpha even higher to 4.15% (compared to 2.32% in Panel A of Table 4). And, of course, this means that the high-minus-zero portfolio now has an alpha of 6.47% (t-statistic $= 4.60$).

One may be concerned that industry concentrations in portfolios could be driving inferences. The above tests are based on industry-adjusted measures of R&D so that is unlikely. It is still possible, however, that the zero R&D portfolio consists largely of stocks in industries that do not do any R&D. We note that our inferences are not merely driven by spreads in high-minus-zero portfolios but by the loadings and intercepts of the high R&D portfolios themselves. Moreover, the spreads in alphas between high R&D portfolios and the adjacent portfolio (Q3), which has a similar industry composition, remain large and statistically significant. In Panel A of Table 4 using portfolios sorted on R&D scaled by book equity, those spreads in alphas are 4.00% and 3.84% for the five-factor and q model, respectively. When portfolios are formed on R&D scaled by market equity, the equivalent spreads in alphas are even larger, at 4.10% and 4.73% respectively. And in all the above cases, the t-statistics are well above 3.00. In short, the failure of the original models cannot be ascribed to industry concentrations in constituent portfolios.

4.3 Resurrecting the Models

4.3.1 The Five-Factor Model

Although original versions of both models fail, there are reasons to believe that all may not be lost. For the five-factor model, the issue seems to be profitability. High industry-adjusted R&D stocks are profitable but have negative loadings on RMW, and zero R&D stocks have low profitability but positive loadings on RMW. But it is also true that high R&D stocks have considerably larger (negative) accruals. In fact, there is monotonic variation between accruals and sorts on R&D. Given this, and given the evidence in Ball, Gerakos, Linnainmaa, and Nikolaev (2016) and Fama and French (2018), it is natural to ask whether the five-factor model can be resurrected by replacing RMW with its cash-based counterpart, RMW_{CP}.

Table 5 does precisely that. The loadings on RMW_{CP} are remarkably different from those on RMW, especially for the extreme portfolios. When we scale R&D by book equity (Panel A), in the zero R&D portfolio, the loading on RMW in Table 4 was 0.17 (t-statistic = 8.16). The corresponding loading on RMW_{CP} in Table 5 is -0.10 (t-statistic = -3.31). Similarly, the high R&D portfolio had a loading of -0.16 (t-statistic = -5.41) on RMW in Table 4. In Table 5, the loading on RMW_{CP} is 0.11 (t-statistic = 2.62). As a result, the alphas of the zero and high R&D portfolios are now indistinguishable from zero. In the high-minus-zero, high-minus-low, and high-minus-Q3 spread portfolios, the alphas are also statistically indistinguishable from zero. And when we consider portfolios formed on R&D scaled by market equity (Panel B of Table 5), an almost identical story unfolds; the loadings on RMW_{CP} flip signs relative to RMW, and all alphas are indistinguishable from zero. Indeed, the improvement in model performance is quite remarkable.

4.3.2 The *q* Model

The situation with the q model is more complicated. Recall from Table 4 that the model faces issues on two fronts, the loadings on both the profitability factor (ROE) and the current investment factor (I/A) do not match their portfolio characteristics.

Since ROE is constructed from earnings before extraordinary items, it includes accruals, which results in counterfactual positive (negative) loadings on ROE for the zero (high) R&D portfolios. An obvious solution is to replace ROE with RMW_{CP}. Table 6 reports the results from such an amended q model. As with the correction to the five-factor model, the loadings on RMW_{CP} are quite difference from the loadings on ROE in Table 4. Using R&D scaled by book equity, the loading on the zero and high R&D portfolios were 0.09 and −0.08 respectively. In Panel A of Table 6, the corresponding loadings on RMW_{CP} are −0.18 and +0.18 respectively. There is a similar change in sign for portfolios formed on R&D scaled by market equity. These loadings drive alphas to zero: in every portfolio and regardless of the scaling variable, alphas from the amended q model are indistinguishable from zero.

A deeper conceptual issue has to do with the two period nature of the theoretical framework that drives the empirical q model. The model generates expected returns by focusing on marginal investment relative to current investment, holding expected profitability constant. But it stands to reason that R&D has implications for future profitability and potentially expected future investment. The multiperiod version of the model in Hou, Mo, Xue, and Zhang (2021) adds expected investment growth factor (EG), raising the hope that future investment growth can soak up the difference between realized returns and expected returns.[14]

We take these hopes to the data. Table 7 evaluates the R&D portfolios against the q^5 model that includes investment growth factor (EG). In book and market equity scaled portfolios, investment growth loads systematically across the R&D portfolios. In zero R&D portfolios, the loadings on EG are negative, and they increase monotonically to large and positive loadings in the high R&D portfolio. Consequently, the alphas of each of these portfolios are indistinguishable from zero (with one exception, that of the Q2 portfolio using book equity as a scalar). However, the resolution is incomplete. Even using the q^5-factor model, the counterfactual loadings on ROE remain. Brushing that aside, the monotonic loadings on EG do not line up with the evidence in Table 2 where there is no relation between current R&D and future investment. As a result, even though the aphas from the q^5 model are indistinguishable from zero, the economic drivers remain unclear.

An obvious solution is to amend the q^5 model by replacing ROE with RMW_{CP}. The results are dramatically different. Table 8 shows that the loadings on EG are driven down to zero for most of the R&D portfolios, matching the annual investment regressions in Table 2. The loadings on RMW_{CP} line up precisely with portfolio characteristics: the zero R&D portfolios have negative loadings and the high R&D portfolios have sharply positive loadings. It is these loading that cause alphas to be indistinguishable from zero for every portfolio. The implication is that the dominant and important feature of these models that explains the returns of R&D portfolios is expected cash profitability, not investment or investment

[14] In the augmented model, expected returns are:

$$R_{it+1} = \frac{X_{it+1} + (a/2)(I_{t+1}/A_{it+1})^2 + (1-\delta)[1 + a(I_{it+1}/A_{it+1})]}{1 + aI_{it}/A_{it}},$$

where the additional term $(1-\delta)(1 + aI_{it+1}/A_{it+1})/(1 + aI_{it}/A_{it})$ is the growth in marginal q and is proportional to expected investment-to-asset growth.

growth.

5 To Capitalize or Not to Capitalize?

US GAAP requires that R&D be expensed rather than capitalized because of the uncertainty around future benefits. A large accounting literature investigates the issue, starting with Lev and Sougiannis (1996) who estimate industry-specific capitalization rates. A more recent literature under the intangibles umbrella investigates the asset pricing implications of capitalization. Some studies focus exclusively on organizational capital (measured using SG&A), others focus on knowledge capital (principally R&D), and still others combine the two. The prototypical approach is to initialize capital stock, capitalize subsequent expenditures, and depreciate the accumulated capital stock using a fixed or industry-specific depreciation rate. Annual capitalized intangibles are used to adjust book values, which are then deployed in asset pricing tests.[15] The underlying idea is that capitalization of intangibles "fixes" the problems associated with the incomplete measurement of book value, thereby improving the performance of the value factor.

Our results thus far suggest that asset prices should anticipate expected future profitability associated with R&D. If that is the case, capitalization should be unnecessary so long as expected profitability is explicitly recognized as a determinant of expected returns. To reconcile our evidence with the intangibles literature, we follow the procedures in Park (2019) but focus entirely on R&D (i.e., we have nothing to say about organizational capital). We first estimate the initial stock of R&D capital at the start of the Compustat data for each firm, cumulating R&D expenditures going forward, and depreciating using the rates in Li and Hall (2016). [16] This generates an estimate of the capitalized value of R&D for each firm-year. We scale these capitalized values by market equity, industry-adjust, and form five portfolios based on the resulting ratio (iXRD/M). In doing so, as before, we isolate zero (capitalized) R&D in one portfolio.

To keep the presentation manageable, Table 9 only reports alphas and the loadings on profitability factor for the original and amended models. Original versions of both the five-factor and q models again generate large alphas and counterfactual loadings on RMW and

[15] A non-comprehensive list of papers that use variants of such an approach include, Arnott, Harvey, Kalesnik, and Linnainmaa (2021), Bongaerts, Kang, and van Dijk (2022), Eisfeldt, Kim, and Papanikolaou (2020), Eisfeldt and Papanikolaou (2013), Ewens, Peters, and Wang (2019), Gulen, Li, Peters, and Zekhnini (2021), Iqbal, Rajgopal, Srivastava, and Zhao (2022), Lev and Srivastava (2019), Kazemi (2021), and Park (2019).

[16] To allow for comparability with the intangibles literature, we do not use imputed R&D in calculating capitalized R&D.

ROE respectively. For example, the zero-R&D decile has annualized alphas of −2.51% in the five-factor model and −2.70% in the q model. And the high R&D decile has equivalent annualized alphas of 1.57% and 2.74%. As a result, the high-minus-zero spread portfolios have extraordinarily large alphas, 4.07 and 5.44% with t-statistics of 3.34 and 4.26 respectively. Prima facie, therefore, it appears that capitalizing R&D is informative with respect to future returns, even after controlling for RMW and ROE. The reason, as before, is the loadings on RMW and ROE. The zero R&D portfolio has low profitability stocks but positive loadings on RMW and ROE, and the high R&D portfolio has negative loadings on RMW and ROE, even though it contains profitable stocks.

For the five-factor model, replacing RMW with RMW_{CP} changes the picture substantially. The alphas of each of the portfolios are indistinguishable from zero, as are alphas of the spread portfolios. If we replace ROE with RMW_{CP} in the q model, a similar picture unfolds. The loadings on RMW_{CP} map more closely to portfolio composition, driving the alphas to zero. In other words, so long as one uses cash-based operating profitability in expected return models, capitalized R&D does not have anything to say about future returns.

6 Robustness and Replication

Given the extensive work on R&D in economics, finance, and accounting, we expect readers to be skeptical of the data, our methods, and our conclusions. In this section, we discuss the robustness of our results to key data and methodological issues. We also describe model performance in small and large stocks. The supplemental and imputed R&D data, as well as the Matlab source code to generate key results, will be put on public domain.

6.1 Sensitivity to Imputed Data

The imputation process addresses the non-randomness of R&D reporting. Nonetheless, some readers may be concerned that our results are purely an artifact of the imputation process. We do not believe that to be the case but we also perform our baseline tests on a sample that does not include any imputed data. Table 10 contains results for original and amended models. To manage space, we suppress loadings MKT, SMB, HML, CMA, and I/A, and only report alphas and loadings on ROE, RMW and RMW_{CP}. We report results for a portfolio of missing (i.e., non-imputed) R&D firms, as well as the zero through high R&D quintiles.

With non-imputed data, both the missing R&D and zero R&D portfolios have negative alphas when using the original five-factor or q model. When we scale R&D by book equity

(Panel A of Table 10), the missing R&D portfolio has alphas of −1.35% and −2.06% respectively. The zero R&D portfolio alphas are −2.45% for both models. The high R&D portfolio still has large positive alphas, 2.27% for the five-factor model and 2.81% for the q model. Scaling R&D by market equity, the equivalent alphas are even larger. For the five-factor model, the high-minus-zero R&D spread is 4.75% when using book equity, rising to 5.71% with market equity. Replacing RMW with RMW_{CP} in both models, the loadings on profitability change sign for the missing, zero and high R&D portfolios. With this replacement, regardless of the scaling variable or the factor model, in every single case, the alphas are indistinguishable from zero.

6.2 Industry Imputation and Adjustment

Thus far, we use two-digit SIC codes to impute and industry-adjust scaled R&D. This ensures that in any given year, the median industry R&D measure is based on a reasonably large number of firms. The cost of doing so is precision in the definition of the industry that the firm competes in. To alleviate concerns arising from these measurement effects, we reproduce results when both the imputation and adjustment process uses three-digit SIC codes.

Table 11 contains profitability loadings and alphas for portfolios constructed using three digit SIC codes. The format of the table is exactly the same as prior tables to allow for back-to-back comparisons, but we (selectively) highlight the lack of differences. In Table 4, using book equity as the scaling variable, for the five-factor model the alphas for the high-minus-zero, high-minus-low and high-minus-Q3 portfolios were 3.67%, 1.56% and 4.00% respectively. In Table 11, the equivalent alphas are 3.90%, 2.75% and 3.21% respectively. Using the q model, alphas for the same portfolios based on two-digit SIC codes were 4.62%, 1.57%, and 3.84%. Using three-digit SIC codes, those alphas are now 4.77%, 2.77% and 3.84% respectively. If we choose to look at R&D scaled by market equity, a similar pattern emerges. Tightening industry adjustments, therefore, results in precisely the same inferences with respect to model performance.

6.3 Firm Size

Our tests deploy the entire cross-section of securities to maximize power. Because we always use value-weighted returns in test portfolios, small firms do not have an inordinate influence on the results.[17] However, since a large fraction of aggregate R&D is conducted by large

[17] It is precisely for this reason that we do not estimate standard Fama and MacBeth (1973) regressions that equal-weight all observations. We do, however, estimate value-weighted Fama and MacBeth (1973)

firms, it is useful and interesting to examine the pricing of R&D within small and large firms. To do so, we first sort firms on market capitalization at the end of June, classifying stocks into small and large groups based on NYSE breakpoints. Within each size group, we then create three portfolios corresponding to zero, low, and high industry-adjusted R&D.

Before describing the tests on these 2×3 double sorts, it is useful to understand the composition and characteristics of the portfolios. Panels A and B of Table 12 provide portfolio information using book and market equity as scalars. In Panel A, for small stocks, the zero, low and high R&D portfolios have 1,538, 756, and 987 stocks respectively. These correspond to 4.8%, 2.2% and 2.6% of the aggregate market. The low and high R&D portfolios within this group represent 4.0% and 18.0% of the aggregate R&D. In large stocks, the zero, low and high R&D portfolios have, on average, 438, 210 and 213 stocks. These represent 35.7%, 23.4%, and 31.2% of the aggregate market capitalization. The low and high R&D portfolios in this group represent 23.3% and 54.3% of the aggregate R&D.

These statistics paint a simple picture: large stocks represent 90% of the aggregate market capitalization and 80% of the aggregate R&D. Thus, while we focus our attention on large stocks, it is important to recognize that the high R&D group in small stocks is not be ignored; they still represent 20% of the aggregate R&D. The characteristics of each of the portfolios within small and large stocks line up in a very similar manner to that in Table 3. When the scaling variable is book equity, the high R&D portfolio is composed of largely profitable growth stocks. Accruals are much more negative in the high R&D portfolio in large stocks, a distinction that becomes important when comparing loadings on RMW with RMW_{CP}. When the scaling variable is market equity, as before, past returns for high R&D firms are poor regardless of firm size, so that these are value stocks.

Panels C and D of Table 12 presents loadings on RMW, ROE, RMW_{CP}, and alphas for original and amended factor models. In small stocks, there is a monotonic pattern between R&D levels and alphas. This is true in portfolios scaled by book and market equity. For example, when the scalar is book (market) equity, the high R&D portfolio has a five-factor alpha of 1.40% (2.70%) with a t-statistic of 2.32 (3.95). Even though this portfolio only represents 2.6% of the aggregate market cap, it should not be discounted because it contains 18% of the aggregate value of R&D. Given these alphas, and the negative alphas in the zero R&D portfolios, the high-minus-zero portfolios in Panels C and D have large returns with high t-statistics. The q model faces similar problems although the magnitudes are even larger. For example, the book (market) equity scaled high R&D portfolio has an alpha of

regressions. We do not explicitly report the results in a table because they duplicate the results in portfolio sorts – after controlling for cash profitability, the coefficients on R&D scaled by book or market equity are statistically indistinguishable from zero.

2.50% (4.18%), so that the high-minus-zero spreads portfolios also have larger alphas. If we replace RMW and ROE with RMW_{CP}, the loadings on profitability shift and in almost every case, the alphas become indistinguishable from zero. In short, the same set of problems that plagued factor models in the full cross-section of stocks appear in small stocks, and the same solution clears them up.

The case for large stocks is similar. When R&D is scaled by book equity, the high-minus-zero portfolios continue to generate positive alphas using both the five-factor and q models (2.31% and 3.33% respectively). When scaling by market equity, the corresponding alphas of the same portfolios are smaller, 1.37% and 2.96% with t-statistics of 1.44 and 2.81 respectively. But as before, replacing RMW and ROE with RMW_{CP} drives down the alphas so that they are indistinguishable from zero.

7 Loose Ends

Our results are likely to leave some readers with unresolved questions, both with respect to some of our existing tests as well as avenues of investigation that we do not pursue. In this section, we attempt to resolve such loose ends.

7.1 R&D and Future Revenue

As described earlier, firms engage in both product R&D to increase future revenues, and/or process R&D to reduce costs. Industry and firm size effects are important for the relative magnitudes of these types of R&D (see, for example, Athey and Schmutzler (1995) and Cohen and Klepper (1996)). FASB Accounting Standards Codification (ASC) 730 provides precise definitions and guidance for both types of R&D, but does not mandate separate disclosure, so we do not attempt to separate the effects of the two.

We can, however, assess whether R&D is related to future revenues. To do so, we estimate annual regressions of a form similar to Table 2, replacing future profits with future revenue. We do not tabulate the results but they are precisely what one would expect: current R&D expenditures are positively related to future revenues (scaled by book equity). Using all stocks, the coefficients on R&D for $\tau = 1$, 2, 3 are 0.69, 0.39, and 0.53, with t-statistics of 3.03, 2.08, and 2.54 respectively. In large (small) stocks, the positive coefficients extend out to five (three) years.

7.2 Other Non-Capitalized Investment

The economic logic that we employ says that non-capitalized investment with uncertain payoffs should be reflected in future profits, and to the extent that prices correctly anticipate future profits, portfolios based on such investment should not have excess returns. Conceptually, this argument applies not just to R&D but to all non-capitalized investment, including advertising and investment in human capital. Given that, one could ask why we do not study other such expenditures in the same framework. In our judgement, the bad data problem is difficult to surmount. SG&A clubs together a variety of sundry expenditures. Some of these expenditures are pure ongoing expenses (e.g., maintenance expenses, wages, etc.), while others reflect investment expenditures with uncertain payoffs. Given our inability to untangle the two, we elect not to go down this road.[18]

8 Conclusion

We study the asset pricing implications of corporate R&D expenditures. Selective and strategic non-reporting of R&D can introduce bias and noise into asset pricing tests. We hand-collect supplementary data and systematically impute R&D to non-reporting firms based on micro-economic fundamentals to avoid these issues.

Our asset pricing tests tell a simple yet powerful economic story. Current R&D expenditures forecast future firm-level profitability, at least three years, and sometimes as far as ten years, into the future. This establishes the channel by which R&D should show up in asset prices—expectations of future profitability. We then ask whether factor models that incorporate expected profitability can price portfolios formed on R&D. The original versions of both the five-factor and q pricing models are unable to do so, principally because their profitability factors do not account for accruals. Using cash-based operating profitability, however, cleans up the models and eviscerates pricing errors. Finally, we ask whether capitalizing R&D is necessary to reflect intangible investment in book values. The answer is no, because non-capitalized R&D expenditures are already reflected in expected profitability, which in turn, is reflected in asset prices.

[18] Eisfeldt and Papanikolaou (2013), Lev and Radhakrishnan (2005), Peters and Taylor (2017), and others capitalize all or some portion of SG&A to attempt to measure the intangible capital associated with such expenditures. Ewens, Peters, and Wang (2019) exploit prices paid for intangible assets in acquistions, as do Deng and Lev (2006) who focus on in-process R&D (IPRD).

Appendices

A Variable Definitions

Variable	Definition
Pref Stock ($Pref$)	Redemption value (PSTKRV); liquidation value (PSTKL); or par value (PSTK), in that order. Set to zero if all missing.
Book value (B)	SEQ + TXDITC $-Pref$; or CEQ + TXDITC $-Pref$; or AT − LT + TXDITC $-Pref$; in that order. TXDITC set to zero if missing. TXDITC set to zero after 1993. Follows Fama and French (1993).
Total assets (A)	AT.
SG&A (SGA)	XSGA − XRD. XSGA to zero if the data footnote XSGA_DC takes the value 4, which Compustat classifies as a combined item. XRD includes supplementary data but excludes post-imputation R&D expenditures (see Appendix B).
Operating profitability (OP)	REVT − COGS $-SGA$.
Cash-based operating profitability (CP)	$OP - \Delta$[RECT + INVT + XPP − (DRC + DRLT) − AP − XACC]. All missing values in the second term set to zero. Follows balance sheet method of Ball, Gerakos, Linnainmaa, and Nikolaev (2016).
Accruals (AC)	Δ[(ACT − CHE) − (LCT − DLC − TXP)] − DP. All missing values set to zero. Follows Sloan (1996). Hribar and Collins (2002) recommend calculating accruals from cash flows as −(RECCH + INVCH + APALCH + TXACH + AOLOCH + DPC). This alternate calculation has no impact on our results.

All accounting information is assumed to be known six months after the fiscal year-end.

B Treatment of Missing R&D Data

B.1 Compustat Reporting and Conventions

Compustat's data process consists of three steps. First, R&D for all firm-years is assigned a null value. Null values are then replaced with valid R&D expenditures reported in 10K statements, Annual Reports, Income Statements, and Notes to Financial statements. Second, when Compustat deems that R&D expense is "insignificant," "immaterial," or "less than 1% of sales," the null value is left in place but a separate data item indicates the immateriality (Compustat code `XRD_DC` = 8).[19] Third, Compustat makes industry adjustments as follows. For single segment firms, Compustat assumes that R&D is equal to zero (not null) for firms in the following industries: Wholesale, Retail, Casinos, Hotels, Leasing Companies, Restaurants, Hospitals, and Healthcare Providers. For mining firms, Compustat does not report R&D (even if the firm does report R&D) because reported R&D typically pools pure R&D with non-R&D expenses (e.g., technology expenditures).

B.2 Supplemental Data

We organize supplemental data collection around three use cases in which Compustat records null values. For each case, we provide an example below and then describe the data collection process.

1. **Interior Cases**: This occurs when R&D is missing in the interior of a firm's time series with valid data at the beginning and end of the series. For example, `permno` = 10026 contains valid R&D from 1986 to 1988, is missing data between 1989 and 1994, and then again contains valid data from 1995 onward. For such cases, we employ a two-step process, sequentially.

   (a) For each interior year, we download all available 10Ks and Annual Reports from the Mergent database. We convert scanned pdfs to machine readable versions and search for the term "research" in each document. If R&D is reported, we record the dollar amount. If the financial statement indicates that R&D was incurred but was immaterial, we record it as zero. If no R&D is reported but one of four requirements for data imputation are met, we flag the firm-year accordingly.[20]

   (b) The Mergent database is incomplete. It does not always contain 10Ks and Annual Reports for the population of firm-years, particularly early in the time series. For such firm-years, we resort to a manual data collection coordinated by the Business Librarian of the Library of Congress (LOC). For the pre-1983 period, we match firm-years to microfiche numbers using the digitized tables maintained by the LOC. After 1983, we match firm-years to microfiche numbers using the physical

[19] A separate footnote code indicates whether R&D includes engineering expense (`XRD_FN` = BG), customer or government sponsored expense (`XRD_FN` = BF), in-process, acquired, or purchased R&D (`XRD_FN` = BW), or some combination of the above (`XRD_FN` = BV).

[20] Section B.3 describes these requirements.

copies of the Q-Data Index provided by the LOC. We then read each financial statement, record R&D as incurred, zero if the financial statement reports R&D being as immaterial, or generate a flag for data imputation.

2. **End-point Cases**: This occurs when R&D is missing at the beginning or end of the time series with zero reported in the remainder of the series. For example, `permno` = 10028 has missing R&D from 1986 to 1988, and zero thereafter. For such cases (and ensuring that `XRD_DC` $\neq 8$), we programmatically backfill or forward-fill zeroes as appropriate.

3. **Full Period Cases**: This occurs when R&D is missing for the entire time series. An example is `permno` = 10032, which is missing R&D for the entire period from 1986 to 2021, but for which the materiality code `XRD_DC` is only equal to 8 for two of those years (2010 and 2011). Since manual data collection for the entire time series would be prohibitively expensive, we follow a three-step process to optimize the collection.

   (a) We do not collect data for microcap firms (below the 20th percentile in the distribution of market capitalization at the end of each December), Utilities, and Financials.

   (b) We download financial statements only for three years: the first year after the start of the time series ($t + 1$), one year before the last year of the time series ($T - 1$), and the midpoint of those two end-points.[21] If financial statements are not available for all three points, we collect what is available.

   (c) As before, we convert financial statements to machine readable pdfs, scan for the term "research," and record reported R&D, or a data imputation flag.

B.3 Imputation

Our imputation process relies on information gathered directly from financial statements. In reading financial statements, we flag a firm-year if one or more of the following four conditions are met.

1. If the firm employs research personnel or has physical research facilities. Annual reports often discuss the employment of research scientists, the location of laboratories, and other such markers that demonstrate that a firm conducts its own R&D. For example, International Multifoods (`permno` = 53129) identifies a vice-president of R&D in 2012. Similarly, Potlatch (`permno` = 49744) describes the establishment of a research center in 2000.

2. If the firm has R&D tax credits, typically reported in Notes to Financial Statements.

[21] There is considerable persistence in the reporting of R&D. If a firm deems it immaterial in year t, it very likely deems it immaterial in year $t+1$. Similarly, if a firm provides some indication that it does R&D but does not report it in year t, the financial statement in year $t+1$ often contains very similar language.

3. If the annual report states that research is important to the current business or to growth in the business. For example, Pepsico's annual report states that research is important to its packaged and processed food business. Similarly, Nike (`permno` = 57665) frequently describes research as being important to its future growth.

4. If the firm has patents, as reported in Notes to Financial Statements or in other parts of the annual report.

When one or more of the above conditions are met, we impute R&D as follows. We first compute the median R&D/Sales ratio for small and large stocks for each industry-year (we use the NYSE median as a size breakpoint to define small and large stocks). For industry demarcations, we use both two and three digit SIC codes. We then replace missing values with the above-calculated median. If the median is not calculable, we replace missing values with zero, mirroring the literature described in Section 2.2. When none of the above conditions are met, we follow the same procedure except that we require a minimum number of firms to calculate the median (10 and 5 firms for small and large stocks respectively).

References

Agarwal, Rajshree, and Michael Gort, 2002, Firm and Product Life Cycles and Firm Survival, *American Economic Review* 92, 184–190.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, 2005, Competition and Innovation: An Inverted-U Relationship, *Quarterly Journal of Economics* 120, 701–728.

Aharoni, Gil, Bruce Grundy, and Qi Zeng, 2013, Stock Returns and the Miller Modigliani Valuation Formula: Revisiting the Fama French Analysis, *Journal of Financial Economics* 110, 347–357.

Arnott, Robert D., Campbell R. Harvey, Vitali Kalesnik, and Juhani T. Linnainmaa, 2021, Reports of Value's Death May be Greatly Exaggerated, *Financial Analysts Journal* 77, 44–67.

Athey, Susan, and Armin Schmutzler, 1995, Product and Process Flexibility in an Innovative Environment, *RAND Journal of Economics* 26, 557–574.

Ball, Ray, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri V. Nikolaev, 2016, Accruals, Cash Flows, and Operating Profitability in the Cross Section of Stock Returns, *Journal of Financial Economics* 121, 28–45.

Barillas, Francisco, and Jay Shanken, 2017, Which Alpha? *Review of Financial Studies* 30, 1316–1338.

Bongaerts, Dion, Xiaowei Kang, and Mathijs van Dijk, 2022, The Intangibles Premium: Risk or Mispricing? *Working paper https://ssrn.com/abstract=3927990*.

Bryzgalova, Svetlana, Sven Lerner, Martin Lettau, and Markus Pelger, 2022, Missing Financial Data, *Working paper https://ssrn.com/abstract=4106794*.

Chan, Louis K.C., Josef Lakonishok, and Theodore Sougiannis, 2001, The Stock Market Valuation of Research and Development Expenditures, *Journal of Finance* 56, 2431–2456.

Chen, Andrew Y., and Jack McCoy, 2022, Missing Values and the Dimensionality of Expected Returns, *Working paper arXiv:2207.13071*.

Cochrane, John H., 1991, Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations, *Journal of Finance* 46, 209–237.

Cohen, Lauren, Karl Diether, and Christopher Malloy, 2013, Misvaluing Innovation, *Review of Financial Studies* 26, 635–666.

Cohen, Wesley M., and Steven Klepper, 1996, Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D, *Review of Economics and Statistics* 78, 232–243.

Corrado, Carol, Jonathan Haskel, Cecilia Jona-Lasinio, and Massimiliano Iommi, 2022, Intangible Capital and Modern Economies, *Journal of Economic Perspectives* 36, 3–28.

Deng, Zhen, and Baruch Lev, 2006, In-process R&D: To Capitalize or Expense? *Journal of Engineering and Technology Management* 23, 18–32.

Donelson, Dain C., and Robert J. Resutek, 2012, The Effect of R&D on Future Returns and Earnings Forecasts, *Review of Accounting Studies* 17, 848–876.

Dong, Ming, David Hirshleifer, and Siew Hong Teoh, 2021, Misvaluation and Corporate Inventiveness, *Journal of Financial and Quantitative Analysis* 56, 2605–2633.

Eberhart, Allan C., William F. Maxwell, and Akhtar R. Siddique, 2004, An Examination of Long-term Abnormal Stock Returns and Operating Performance Following R&D Iincreases, *Journal of Finance* 59, 623–650.

Eisfeldt, Andrea L., Edward T. Kim, and Dimitris Papanikolaou, 2020, Intangible Value, *Critical Finance Review forthcoming.*

Eisfeldt, Andrea L., and Dimitris Papanikolaou, 2013, Organization Capital and the Cross-Section of Expected Returns, *Journal of Finance* 68, 1365–1406.

Ewens, Michael, Ryan H. Peters, and Sean Wang, 2019, Measuring Intangible Capital with Market Prices, *Working paper https://www.nber.org/papers/w25960.*

Fama, Eugene F., and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2006, Profitability, Investment and Average Returns, *Journal of Financial Economics* 82, 491–518.

Fama, Eugene F., and Kenneth R. French, 2015, A Five-Factor Asset Pricing Model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and Kenneth R. French, 2016, Dissecting Anomalies with a Five-Factor Model, *Review of Financial Studies* 29, 69–103.

Fama, Eugene F., and Kenneth R. French, 2018, Choosing Factors, *Journal of Financial Economics* 128, 234–252.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, Return, and Equilibrium: Empirical Tests, *Journal of Political Economy* 81, 607–636.

Freyberger, Joachim, Björn Höppner, Andreas Neuhierl, and Michael Weber, 2021, Missing Data in Asset Pricing Panels, *Working paper https://papers.ssrn.com/abstract=3932438.*

Graham, Benjamin, and Spencer Barrett Meredith, 1937, *The Interpretation of Financial Statements* vol. 4. (Harper New York).

Gulen, Huseyin, Dongmei Li, Ryan H. Peters, and Morad Zekhnini, 2021, Intangible Capital in Factor Models, *Working paper https://papers.ssrn.com/abstract=3725002*.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, 2013, Innovative Efficiency and Stock Returns, *Journal of Financial Economics* 107, 632–654.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An Augmented q-Factor Model with Expected Growth, *Review of Finance* 25, 1–41.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting Anomalies: An Investment Approach, *Review of Financial Studies* 28, 650–705.

Hribar, Paul, and Daniel W. Collins, 2002, Errors in Estimating Accruals: Implications for Empirical Research, *Journal of Accounting Research* 40, 105–134.

Iqbal, Aneel, Shivaram Rajgopal, Anup Srivastava, and Rong Zhao, 2022, Value of Internally Generated Intangible Capital, *Working paper https://papers.ssrn.com/abstract=3917998*.

Kahle, Kathleen M., and Ralph A. Walkling, 1996, The Impact of Industry Classifications on Financial Research, *Journal of Financial and Quantitative Analysis* 31, 309–335.

Kapons, Martin M., 2020, R&D Expenses, R&D Capitalization, the Book-to-Price Ratio and the Cross-Section of Returns, *Working paper*.

Kazemi, Maziar, 2021, Intangible Investment, Displacement Risk, and the Value Discount, *Working paper https://papers.ssrn.com/abstract=3988890*.

Koh, Ping-Sheng, and David M. Reeb, 2015, Missing R&D, *Journal of Accounting and Economics* 60, 73–94.

Lev, Baruch, and Suresh Radhakrishnan, 2005, The Valuation of Organization Capital, in Carol Corrado, John Haltiwanger, and Daniel Sichel, eds.: *Measuring Capital in the New Economy* (University of Chicago Press,).

Lev, Baruch, and Theodore Sougiannis, 1996, The Capitalization, Amortization, and Value-relevance of R&D, *Journal of Accounting and Economics* 21, 107–138.

Lev, Baruch, and Anup Srivastava, 2019, Explaining the Recent Failure of Value Investing, *Working paper https://papers.ssrn.com/abstract=3442539*.

Li, Dongmei, 2011, Financial Constraints, R&D Investment, and Stock Returns, *Review of Financial Studies* 24, 2974–3007.

Li, Wendy C.Y., and Bronwyn H. Hall, 2016, Depreciation of Business R&D Capital, *Working paper https://www.nber.org/papers/w22473*.

McVay, Sarah Elizabeth, 2006, Earnings Management Using Classification Shifting: An Examination of Core Earnings and Special Items, *Accounting Review* 81, 501–531.

Miller, Merton H., and Franco Modigliani, 1961, Dividend Policy, Growth, and the Valuation of Shares, *Journal of Business* 34, 411–433.

Novy-Marx, Robert, 2013, The Other Side of Value: The Gross Profitability Premium, *Journal of Financial Economics* 108, 1–28.

Oswald, Dennis, Ana Simpson, and Paul Zarowin, 2022, Capitalization vs. Expensing and the Behavior of R&D Expenditures, *Review of Accounting Studies* 27, 1199–1232.

Park, Hyuna, 2019, An Intangible-Adjusted Book-to-Market Ratio Still Predicts Stock Returns, *Critical Finance Review* 25, 207–236.

Penman, Stephen H., 2009, Accounting for Intangible Assets: There is also an Income Statement, *Abacus* 45, 358–371.

Peters, Ryan H., and Lucian A. Taylor, 2017, Intangible Capital and the Investment-q Relation, *Journal of Financial Economics* 123, 251–272.

Resutek, Robert J., 2022, Is R&D Really That Special? A Fixed-Cost Explanation for the Empirical Patterns of R&D Firms, *Contemporary Accounting Research* 39, 721–749.

Schumpeter, Joseph A., 1943, *Capitalism, Socialism and Democracy.* (New York: Routledge).

Sloan, Richard G., 1996, Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *Accounting Review* 3, 289–315.

Sun, Estelle Y., 2021, The Differential Role of R&D and SG&A for Earnings Management and Stock Price Manipulation, *Contemporary Accounting Research* 38, 242–275.

Tirole, Jean, 1988, *The Theory of Industrial Organization.* (MIT press).

Figure 1: Aggregate R&D

This figure show the ratio of aggregate R&D scaled by aggregate book equity over time. The line labelled "Compustat" only uses original R&D data derived directly from Compustat. The line labelled "Supplementary" adds additional hand-collected R&D data. The line labelled "Imputed" adds data imputed via methods described in Appendix B.

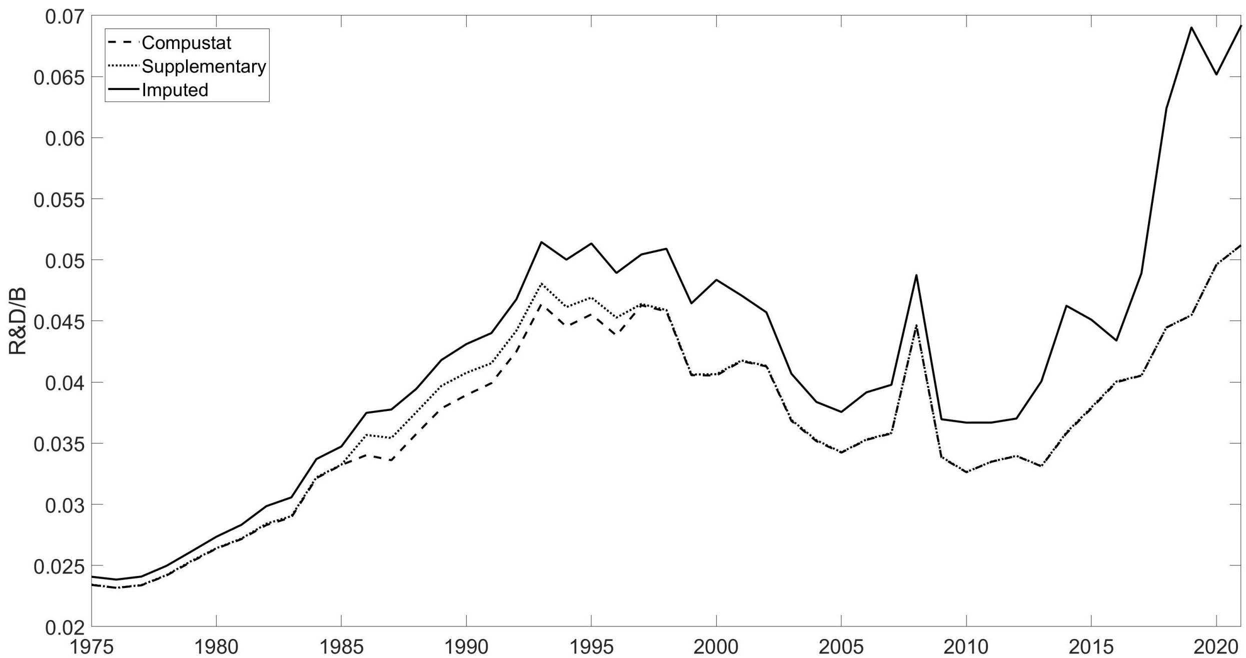

Table 1: Industry Distribution of R&D Data

For each two-digit SIC code with an average of more than ten firms, the table reports the average number of firms with valid R&D reported by Computat, the number of firms after supplementary data collection, and the number of firms after data imputation. Appendix B provides details of the supplementary data collection and imputation process.

SIC	Name	Compustat	Supplementary	Imputed
13	Oil and Gas Extraction	24	87	169
15	Building Construction General Contractors	3	19	37
20	Food and Kindred Products	39	70	109
22	Textile Mill Products	12	22	35
23	Apparel	7	25	46
24	Lumber and Wood Products	6	19	27
25	Furniture and Fixtures	17	23	31
26	Paper and Allied Products	22	34	45
27	Printing, Publishing, and Allied Industries	11	39	65
28	Chemicals and Allied Products	293	308	325
29	Petroleum Refining and Related Industries	12	21	28
30	Rubber and Miscellaneous Plastics Products	35	43	56
32	Stone, Clay, Glass, and Concrete Products	16	25	30
33	Primary Metal Industries	27	41	63
34	Fabricated Metal Products	50	66	85
35	Industrial Machinery and Computer Equipment	257	276	294
36	Electronic and Other Electrical Equipment	287	305	333
37	Transportation Equipment	73	83	100
38	Measuring, Analyzing, and Controlling Instruments	263	270	278
39	Miscellaneous Manufacturing Industries	29	38	49
42	Motor Freight Transportation and Warehousing	1	14	33
45	Transportation by Air	1	17	31
48	Communications	22	56	114
49	Electric, Gas, and Sanitary Services	7	23	182
50	Wholesale Trade: Durable Goods	59	89	118
51	Wholesale Trade: Non-durable Goods	28	48	66
53	General Merchandise Stores	30	37	40
54	Food Stores	19	29	33
56	Apparel and Accessory Stores	36	40	43
58	Eating and Drinking Places	54	69	73
59	Miscellaneous Retail	51	65	77
60	Depository Institutions	1	3	419
61	Non-depository Credit Institutions	2	7	45
62	Brokers, Dealers, Exchanges, and Services	5	11	59
63	Insurance Carriers	3	12	118
65	Real Estate	7	18	44
67	Holding and Other Investment Offices	27	34	65
73	Business Services	253	298	402
78	Motion Pictures	5	12	30
79	Amusement and Recreation Services	18	27	39
80	Health Services	38	54	80
87	Engineering, Accounting, and Management Services	38	52	86
99	Nonclassifiable Establishments	21	24	38

Table 2: Annual Regressions to Predict Profitability and Investment

We run annual cross-sectional regressions in which the dependent variable is $OP_{t+\tau}/B_{t+\tau}$, $CP_{t+\tau}/B_{t+\tau}$, or $dA_{t+\tau}/A_{t+\tau-1}$ where $\tau = 1, \ldots, 5$. OP is operating profitability, CP is cash-based operating profitability, and dA/A is asset growth. The independent variables are measured in year t. Y is OP when the dependent variable is $OP_{t+\tau}/B_{t+\tau}$ and CP otherwise. $R1Y$ is the lagged one-year stock return, $\log M$ is log market capitalization, $\log B/M$ is log book-to-market ratio, $-AC/B$ is the ratio of accruals to book equity when accruals are negative (and zero otherwise), $+AC/B$ is the ratio of accruals to book equity when accruals are positive (and zero otherwise), dA/A is the percentage growth in total assets. Please see text and Appendix A for exact definitions of OP, CP, AC, B, and A. Panels A1/B1 use $R\&D/B$ (the ratio of R&D to book equity) and Panels A2/B2 use $R\&D/M$ (the ratio of R&D to market equity) as the final independent variable. We industry adjust $R\&D/B$ and $R\&D/M$ using 2-digit SIC codes. All independent and dependent variables are winsorized each year at 1%/99% level before running regressions. We report the average number of stocks and the average $\overline{R}^2$ in the last two rows of each panel. Panels A1/A2 includes all stocks while Panels B1/B2 restricts the regressions to large stocks (defined as those that have market capitalization above the median NYSE market capitalization at the end of the year). The table shows average slopes with t-statistics in parenthesis. The sample period is 1975 to 2021.

	$OP_{t+\tau}/B_{t+\tau}$					$CP_{t+\tau}/B_{t+\tau}$					$dA_{t+\tau}/A_{t+\tau-1}$				
$\tau \rightarrow$	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Panel A1: All stocks, using R&D to book equity															
cnst	0.042	0.048	0.050	0.056	0.057	0.063	0.071	0.072	0.082	0.086	0.117	0.134	0.137	0.137	0.135
	(4.34)	(4.38)	(3.90)	(4.58)	(5.13)	(3.37)	(3.54)	(4.01)	(4.54)	(4.69)	(9.15)	(9.06)	(9.47)	(9.66)	(9.00)
Y_t/B_t	0.369	0.286	0.250	0.205	0.196	0.209	0.156	0.142	0.094	0.094	−0.017	−0.011	−0.008	−0.011	−0.003
	(18.67)	(18.93)	(15.61)	(12.55)	(11.98)	(13.69)	(9.59)	(10.62)	(5.96)	(8.51)	(−4.48)	(−2.80)	(−2.11)	(−2.80)	(−0.87)
$Y_t < 0$	−0.199	−0.146	−0.112	−0.108	−0.094	−0.148	−0.124	−0.093	−0.097	−0.091	−0.027	−0.007	0.005	0.010	0.016
	(−12.63)	(−8.41)	(−6.83)	(−6.00)	(−4.81)	(−10.48)	(−7.38)	(−6.37)	(−6.23)	(−6.65)	(−3.70)	(−1.07)	(0.80)	(1.25)	(2.57)
$R1Y_t$	0.001	−0.006	−0.015	−0.012	−0.016	−0.006	−0.002	−0.010	0.004	−0.009	0.072	0.022	0.009	0.004	−0.001
	(0.28)	(−1.28)	(−3.32)	(−3.07)	(−3.81)	(−1.20)	(−0.27)	(−1.59)	(0.76)	(−1.20)	(13.68)	(5.40)	(2.61)	(1.03)	(−0.29)
$\log M_t$	0.020	0.025	0.027	0.029	0.030	0.028	0.031	0.032	0.033	0.032	−0.008	−0.010	−0.010	−0.010	−0.010
	(13.28)	(14.72)	(14.52)	(16.33)	(16.77)	(12.51)	(12.17)	(12.41)	(13.06)	(12.39)	(−4.47)	(−5.13)	(−5.39)	(−5.53)	(−5.46)
$\log B_t/M_t$	−0.030	−0.015	−0.008	−0.007	−0.004	−0.004	0.005	0.005	0.004	−0.001	−0.095	−0.073	−0.060	−0.048	−0.041
	(−5.68)	(−3.13)	(−1.54)	(−1.20)	(−0.75)	(−0.62)	(0.78)	(0.81)	(0.61)	(−0.10)	(−20.00)	(−16.55)	(−14.46)	(−12.39)	(−10.00)
$-AC_t/B_t$	−0.132	−0.093	−0.077	−0.069	−0.063	−0.045	−0.021	−0.035	−0.017	−0.013	0.046	0.029	0.019	−0.001	0.000
	(−12.06)	(−8.12)	(−6.86)	(−5.58)	(−5.37)	(−2.29)	(−1.24)	(−1.70)	(−0.83)	(−0.95)	(7.69)	(6.06)	(4.23)	(−0.11)	(0.04)
$+AC_t/B_t$	0.048	0.047	0.042	0.043	0.048	0.255	0.198	0.122	0.147	0.128	−0.009	−0.007	−0.005	−0.003	−0.004
	(4.90)	(4.17)	(3.80)	(3.20)	(3.47)	(8.01)	(5.21)	(3.54)	(4.51)	(3.42)	(−0.88)	(−0.85)	(−0.67)	(−0.37)	(−0.51)
dA_t/A_{t-1}	−0.017	−0.014	−0.009	−0.012	−0.012	−0.005	0.015	0.022	0.009	0.013	0.081	0.039	0.017	0.007	0.004
	(−2.53)	(−1.86)	(−1.28)	(−1.20)	(−0.93)	(−0.51)	(1.37)	(1.72)	(0.63)	(0.78)	(8.37)	(4.61)	(3.02)	(1.10)	(0.52)
$R\&D_t/B_t$	0.135	0.109	0.086	0.044	0.039	0.131	0.133	0.077	0.076	0.038	−0.031	0.001	0.032	0.051	0.022
	(4.63)	(3.45)	(2.88)	(1.28)	(1.27)	(3.16)	(2.95)	(2.47)	(1.51)	(0.87)	(−2.02)	(0.05)	(1.44)	(2.17)	(0.91)
#stocks	3,825	3,548	3,296	3,069	2,866	3,825	3,548	3,296	3,069	2,866	3,827	3,544	3,290	3,063	2,859
$\overline{R}^2$	22.22	12.98	9.16	7.20	6.07	9.01	5.76	4.17	3.42	2.99	11.04	5.15	3.37	2.75	2.32

	$OP_{t+\tau}/B_{t+\tau}$					$CP_{t+\tau}/B_{t+\tau}$					$dA_{t+\tau}/A_{t+\tau-1}$				
$\tau \rightarrow$	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Panel A2: All stocks, using R&D to market equity															
cnst	0.038	0.043	0.044	0.052	0.054	0.054	0.062	0.062	0.076	0.082	0.124	0.136	0.137	0.138	0.135
	(3.96)	(3.94)	(3.47)	(4.36)	(4.98)	(2.97)	(3.13)	(3.50)	(4.27)	(4.47)	(10.28)	(9.70)	(9.89)	(9.91)	(9.35)
Y_t/B_t	0.368	0.284	0.247	0.203	0.194	0.208	0.155	0.141	0.093	0.093	−0.016	−0.011	−0.008	−0.010	−0.003
	(18.45)	(18.94)	(15.57)	(12.51)	(11.96)	(13.68)	(9.51)	(10.57)	(5.83)	(8.50)	(−4.23)	(−2.65)	(−2.11)	(−2.69)	(−0.73)
$Y_t < 0$	−0.198	−0.146	−0.112	−0.109	−0.095	−0.146	−0.122	−0.091	−0.097	−0.091	−0.027	−0.008	0.006	0.010	0.016
	(−12.49)	(−8.45)	(−6.87)	(−6.08)	(−4.82)	(−10.40)	(−7.32)	(−6.25)	(−6.20)	(−6.59)	(−3.88)	(−1.09)	(0.82)	(1.30)	(2.64)
$R1Y_t$	0.001	−0.005	−0.014	−0.011	−0.015	−0.004	0.000	−0.007	0.006	−0.006	0.071	0.022	0.008	0.003	−0.001
	(0.24)	(−1.03)	(−3.13)	(−2.65)	(−3.36)	(−0.81)	(−0.01)	(−1.16)	(1.18)	(−0.89)	(13.68)	(5.41)	(2.51)	(0.76)	(−0.25)
$\log M_t$	0.020	0.025	0.028	0.030	0.030	0.028	0.032	0.033	0.033	0.033	−0.009	−0.010	−0.010	−0.010	−0.010
	(12.97)	(14.86)	(14.65)	(16.30)	(16.56)	(12.98)	(12.42)	(12.68)	(13.38)	(12.59)	(−4.93)	(−5.40)	(−5.56)	(−5.71)	(−5.70)
$\log B_t/M_t$	−0.038	−0.021	−0.013	−0.007	−0.005	−0.013	−0.002	0.001	0.002	−0.001	−0.091	−0.072	−0.061	−0.050	−0.042
	(−7.87)	(−4.67)	(−2.67)	(−1.48)	(−0.97)	(−1.84)	(−0.34)	(0.11)	(0.26)	(−0.12)	(−18.09)	(−14.57)	(−12.30)	(−11.02)	(−9.51)
$-AC_t/B_t$	−0.135	−0.094	−0.078	−0.067	−0.061	−0.048	−0.023	−0.034	−0.015	−0.010	0.048	0.030	0.018	−0.003	0.000
	(−12.71)	(−8.49)	(−7.12)	(−5.80)	(−5.25)	(−2.43)	(−1.39)	(−1.65)	(−0.78)	(−0.72)	(8.29)	(6.88)	(4.35)	(−0.53)	(−0.02)
$+AC_t/B_t$	0.051	0.050	0.044	0.042	0.047	0.256	0.200	0.123	0.145	0.126	−0.010	−0.007	−0.003	−0.001	−0.003
	(5.21)	(4.56)	(4.07)	(3.24)	(3.49)	(8.13)	(5.29)	(3.56)	(4.52)	(3.39)	(−1.06)	(−0.93)	(−0.49)	(−0.11)	(−0.39)
dA_t/A_{t-1}	−0.017	−0.013	−0.008	−0.010	−0.011	−0.005	0.016	0.024	0.011	0.015	0.080	0.039	0.016	0.006	0.004
	(−2.69)	(−1.73)	(−1.17)	(−1.03)	(−0.83)	(−0.56)	(1.49)	(1.86)	(0.74)	(0.90)	(8.29)	(4.56)	(2.91)	(0.99)	(0.48)
$R\&D_t/M_t$	0.238	0.241	0.242	0.129	0.109	0.389	0.341	0.330	0.225	0.167	−0.185	−0.074	−0.004	0.019	0.026
	(5.56)	(5.03)	(4.77)	(2.23)	(2.08)	(5.35)	(4.23)	(4.77)	(2.49)	(1.97)	(−6.99)	(−2.14)	(−0.10)	(0.47)	(0.72)
#stocks	3,825	3,548	3,296	3,069	2,866	3,825	3,548	3,296	3,069	2,866	3,827	3,544	3,290	3,063	2,859
$\overline{R}^2$	22.17	12.99	9.19	7.20	6.08	9.06	5.80	4.22	3.41	2.99	11.11	5.14	3.34	2.69	2.23

	$OP_{t+\tau}/B_{t+\tau}$					$CP_{t+\tau}/B_{t+\tau}$					$dA_{t+\tau}/A_{t+\tau-1}$				
$\tau \rightarrow$	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Panel B1: Large stocks, using R&D to book equity															
cnst	−0.010	−0.014	−0.037	−0.067	−0.061	0.067	0.078	0.058	0.032	0.056	0.209	0.160	0.123	0.118	0.125
	(−0.49)	(−0.54)	(−1.32)	(−2.14)	(−1.94)	(2.75)	(2.49)	(1.68)	(0.76)	(1.52)	(13.83)	(11.04)	(9.24)	(8.79)	(8.00)
Y_t/B_t	0.638	0.508	0.468	0.395	0.379	0.280	0.196	0.184	0.129	0.116	−0.027	−0.011	−0.001	−0.006	−0.008
	(17.12)	(13.01)	(10.12)	(7.30)	(7.39)	(6.88)	(3.93)	(4.15)	(2.88)	(3.39)	(−4.59)	(−1.53)	(−0.27)	(−1.24)	(−1.83)
$Y_t < 0$	0.012	−0.019	−0.019	−0.016	0.005	0.000	−0.013	0.003	−0.023	−0.001	0.007	−0.010	0.008	−0.013	0.002
	(0.22)	(−0.32)	(−0.22)	(−0.20)	(0.07)	(−0.01)	(−0.39)	(0.08)	(−0.62)	(−0.04)	(0.67)	(−1.39)	(0.83)	(−1.99)	(0.32)
$R1Y_t$	−0.009	−0.010	−0.037	−0.038	−0.048	−0.043	−0.045	−0.055	−0.030	−0.047	0.076	0.045	0.028	0.018	0.015
	(−0.88)	(−1.07)	(−3.65)	(−2.93)	(−4.59)	(−3.02)	(−3.16)	(−4.37)	(−2.05)	(−2.97)	(9.68)	(6.60)	(4.39)	(3.21)	(1.77)
$\log M_t$	0.015	0.022	0.026	0.033	0.034	0.017	0.022	0.026	0.031	0.031	−0.020	−0.013	−0.008	−0.007	−0.008
	(5.54)	(6.32)	(7.26)	(8.68)	(7.97)	(6.04)	(6.07)	(5.78)	(6.16)	(5.74)	(−10.55)	(−8.32)	(−5.91)	(−4.32)	(−3.98)
$\log B_t/M_t$	−0.039	−0.040	−0.041	−0.046	−0.047	−0.100	−0.087	−0.083	−0.082	−0.086	−0.086	−0.066	−0.050	−0.044	−0.041
	(−4.40)	(−3.78)	(−3.46)	(−3.05)	(−3.72)	(−11.19)	(−8.42)	(−7.76)	(−6.72)	(−8.41)	(−17.01)	(−16.69)	(−14.71)	(−13.60)	(−15.13)
$-AC_t/B_t$	0.016	0.022	−0.012	−0.006	−0.020	0.018	0.033	0.040	0.045	0.077	0.025	0.011	0.009	0.002	0.000
	(0.52)	(0.65)	(−0.40)	(−0.13)	(−0.62)	(0.38)	(0.64)	(0.70)	(0.74)	(1.39)	(2.78)	(1.51)	(1.34)	(0.20)	(−0.05)
$+AC_t/B_t$	−0.013	0.018	0.008	0.038	−0.002	0.135	0.146	0.104	0.078	−0.026	−0.029	0.003	0.006	−0.001	0.000
	(−0.38)	(0.64)	(0.28)	(1.31)	(−0.07)	(1.87)	(2.02)	(1.63)	(1.24)	(−0.39)	(−2.91)	(0.35)	(0.77)	(−0.15)	(−0.01)
dA_t/A_{t-1}	−0.017	−0.049	−0.049	−0.030	−0.030	−0.048	−0.045	−0.047	−0.020	−0.042	0.093	0.059	0.042	0.022	0.017
	(−1.75)	(−4.00)	(−4.22)	(−1.69)	(−1.71)	(−2.58)	(−2.74)	(−2.42)	(−0.82)	(−1.69)	(7.86)	(7.08)	(5.96)	(2.43)	(1.75)
$R\&D_t/B_t$	−0.018	−0.041	0.015	0.008	0.056	0.363	0.262	0.281	0.328	0.403	0.044	0.017	0.021	0.043	0.040
	(−0.24)	(−0.53)	(0.20)	(0.11)	(0.98)	(2.94)	(2.22)	(2.15)	(2.58)	(4.26)	(1.62)	(0.60)	(0.70)	(1.47)	(1.48)
#stocks	864	830	798	766	736	864	830	798	766	736	864	829	797	765	736
$\overline{R}^2$	43.52	26.57	20.20	15.52	12.85	18.55	12.53	9.05	6.89	5.97	13.71	7.64	4.73	3.29	3.06

	$OP_{t+\tau}/B_{t+\tau}$					$CP_{t+\tau}/B_{t+\tau}$					$dA_{t+\tau}/A_{t+\tau-1}$				
$\tau \rightarrow$	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Panel B2: Large stocks, using R&D to market equity															
cnst	−0.010	−0.011	−0.032	−0.064	−0.058	0.058	0.074	0.054	0.029	0.051	0.208	0.158	0.123	0.116	0.123
	(−0.45)	(−0.40)	(−1.17)	(−2.16)	(−1.90)	(2.45)	(2.44)	(1.61)	(0.74)	(1.41)	(13.62)	(11.13)	(9.19)	(9.23)	(8.18)
Y_t/B_t	0.625	0.491	0.461	0.388	0.376	0.281	0.192	0.186	0.128	0.120	−0.024	−0.010	0.000	−0.004	−0.006
	(17.73)	(12.66)	(10.25)	(7.21)	(7.48)	(7.24)	(3.89)	(4.33)	(2.86)	(3.59)	(−4.46)	(−1.48)	(−0.03)	(−0.79)	(−1.61)
$Y_t < 0$	−0.003	−0.044	−0.043	−0.021	0.007	0.000	−0.016	0.002	−0.023	0.003	0.010	−0.008	0.008	−0.011	0.003
	(−0.05)	(−0.75)	(−0.50)	(−0.27)	(0.11)	(0.01)	(−0.49)	(0.05)	(−0.64)	(0.09)	(0.92)	(−1.23)	(0.89)	(−1.56)	(0.42)
$R1Y_t$	−0.008	−0.007	−0.036	−0.038	−0.043	−0.041	−0.042	−0.052	−0.028	−0.040	0.076	0.045	0.027	0.017	0.014
	(−0.78)	(−0.77)	(−3.44)	(−2.78)	(−4.26)	(−2.85)	(−3.11)	(−4.08)	(−1.89)	(−2.59)	(9.48)	(6.65)	(4.26)	(3.17)	(1.67)
$\log M_t$	0.015	0.022	0.026	0.033	0.034	0.016	0.021	0.026	0.030	0.030	−0.020	−0.013	−0.008	−0.007	−0.008
	(5.61)	(6.35)	(7.26)	(8.77)	(8.03)	(5.92)	(6.18)	(5.77)	(6.26)	(5.89)	(−10.32)	(−8.12)	(−6.02)	(−4.65)	(−4.12)
$\log B_t/M_t$	−0.043	−0.045	−0.044	−0.048	−0.051	−0.113	−0.099	−0.092	−0.093	−0.099	−0.086	−0.065	−0.050	−0.044	−0.041
	(−5.06)	(−4.25)	(−3.77)	(−3.13)	(−4.06)	(−12.40)	(−9.33)	(−7.83)	(−7.11)	(−9.16)	(−16.23)	(−16.36)	(−14.49)	(−13.44)	(−14.75)
$-AC_t/B_t$	0.017	0.028	0.007	−0.002	−0.018	0.020	0.035	0.056	0.044	0.072	0.023	0.011	0.010	0.000	−0.001
	(0.56)	(0.84)	(0.24)	(−0.05)	(−0.56)	(0.44)	(0.66)	(0.97)	(0.71)	(1.30)	(2.56)	(1.55)	(1.40)	(−0.03)	(−0.11)
$+AC_t/B_t$	−0.022	0.017	−0.002	0.036	0.004	0.126	0.140	0.093	0.077	−0.017	−0.025	0.004	0.007	0.000	0.003
	(−0.65)	(0.59)	(−0.05)	(1.20)	(0.11)	(1.77)	(1.94)	(1.54)	(1.22)	(−0.25)	(−2.57)	(0.41)	(0.83)	(0.02)	(0.34)
dA_t/A_{t-1}	−0.017	−0.049	−0.047	−0.028	−0.030	−0.045	−0.044	−0.043	−0.016	−0.040	0.093	0.058	0.043	0.022	0.017
	(−1.72)	(−4.12)	(−4.13)	(−1.61)	(−1.67)	(−2.42)	(−2.62)	(−2.31)	(−0.66)	(−1.63)	(7.78)	(7.05)	(5.99)	(2.51)	(1.78)
$R\&D_t/M_t$	0.336	0.354	0.479	0.452	0.518	1.076	0.806	0.812	0.898	0.948	−0.101	−0.080	−0.053	−0.033	−0.005
	(2.10)	(2.04)	(2.77)	(2.47)	(2.98)	(4.67)	(3.68)	(3.30)	(4.06)	(3.78)	(−2.16)	(−1.44)	(−0.88)	(−0.51)	(−0.07)
#stocks	864	830	798	766	736	864	830	798	766	736	864	829	797	765	736
$\overline{R}^2$	43.12	26.39	19.99	15.39	12.87	18.32	12.50	8.87	6.81	6.16	13.52	7.56	4.64	3.19	3.03

Table 3: Characteristics of Portfolios sorted on R&D

We sort stocks at the end of each June based on their R&D to book equity ratio or their R&D to market equity ratio. We industry adjust $R\&D/B$ and $R\&D/M$ using 2-digit SIC codes before sorting. Stocks with $R\&D/B = 0$ and $R\&D/M = 0$ are placed in one group and the remaining stocks are sorted into quartiles based on breakpoints from NYSE stocks only. These quartiles are labeled Low, Q2, Q3, and High. The portfolios are value-weighted and rebalanced once every year at the end of June. This table presents formation period characteristics of these portfolios. #stocks is the average number of stocks, %ME is the percentage of aggregate market capitalization, $\%R\&D$ is the percentage of aggregate R&D, B/M is the ratio of book equity to market equity, dA/A is percentage growth of total assets, OP/B is the ratio of operating profitability to book equity, CP/B is the ratio of cash profitability to book equity, AC/B is the ratio of accruals to book equity, and $R1Y$ is the last year excess stock return (in percent). Please see text and Appendix A for exact definitions of OP, CP, AC, and B. Portfolio accounting characteristics are calculated on aggregate basis. For example, B/M is the ratio of the sum of book values to the sum of market values of stocks in the portfolio. $R1Y$ is calculated as the value-weighted across stocks in the portfolio. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	#stocks	$\%ME$	$\%R\&D$	B/M	dA/A	OP/B	CP/B	AC/B	$R1Y$
	Panel A: Sorts on $R\&D/B$								
Zero	1,971	40.0	0.0	0.631	0.097	0.332	0.306	−0.071	8.8
Low	542	8.7	10.5	0.449	0.119	0.344	0.296	−0.063	8.2
Q2	459	13.7	14.5	0.488	0.093	0.378	0.367	−0.082	10.5
Q3	396	18.1	19.8	0.497	0.068	0.389	0.379	−0.099	8.0
High	773	18.3	52.4	0.369	0.077	0.565	0.536	−0.125	9.1
	Panel B: Sorts on $R\&D/M$								
Zero	2,019	40.8	0.0	0.618	0.097	0.340	0.315	−0.072	8.8
Low	603	15.6	13.1	0.289	0.137	0.434	0.388	−0.068	16.3
Q2	442	16.1	18.2	0.403	0.091	0.433	0.418	−0.084	10.6
Q3	408	17.3	21.4	0.508	0.074	0.400	0.386	−0.104	7.0
High	792	10.3	44.1	0.663	0.057	0.486	0.467	−0.142	−0.8

Table 4: Original Factor Model Regressions for R&D Sorted Portfolios

We sort stocks at the end of each June based on their R&D to book equity ratio in Panel A and their R&D to market equity ratio in Panel B as in Table 3. R&D is calculated as described in the text. We industry adjust $R\&D/B$ and $R\&D/M$ using 2-digit SIC codes before sorting. Stocks with $R\&D/B = 0$ and $R\&D/M = 0$ and are placed in one group and the remaining stocks are sorted into quartiles based on NYSE breakpoints. These quartiles are labeled Low, Q2, Q3, and High. The portfolios are value-weighted and rebalanced once every year at the end of June. Various subpanels report alphas and loadings from various factor models. The five-factor models is Fama and French (2015) model where the factors are Mkt, SMB, HML, RMW, and CMA . The q model is the four-factor model of Hou, Xue, and Zhang (2015) where the factors are Mkt, ME, I/A, and ROE. All alphas are reported in annualized percent. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	Five-factor model						q model				
	Alpha	Mkt	SMB	HML	RMW	CMA	Alpha	Mkt	ME	I/A	ROE
	Panel A: Sorts on $R\&D/B$										
Zero	−1.74	1.01	0.01	0.34	0.17	−0.10	−2.30	1.02	0.00	0.35	0.09
	(−3.17)	(92.67)	(0.47)	(17.31)	(8.16)	(−3.04)	(−3.25)	(74.71)	(0.12)	(11.39)	(4.13)
Low	0.36	1.08	0.00	−0.12	−0.12	−0.20	0.75	1.08	−0.02	−0.34	−0.09
	(0.40)	(60.62)	(0.01)	(−3.64)	(−3.49)	(−3.91)	(0.82)	(60.85)	(−0.83)	(−8.53)	(−3.10)
Q2	3.96	0.97	−0.11	−0.17	−0.16	−0.21	4.66	0.97	−0.13	−0.41	−0.14
	(3.86)	(47.55)	(−3.46)	(−4.52)	(−4.17)	(−3.55)	(4.42)	(47.93)	(−4.55)	(−9.09)	(−4.05)
Q3	−2.07	0.97	−0.09	−0.07	0.16	0.32	−1.52	0.95	−0.12	0.23	0.05
	(−3.11)	(73.21)	(−4.26)	(−2.91)	(6.42)	(8.19)	(−2.16)	(70.54)	(−6.02)	(7.61)	(2.44)
High	1.92	1.04	0.00	−0.31	−0.16	0.13	2.32	1.03	0.03	−0.28	−0.08
	(2.55)	(69.94)	(0.12)	(−11.33)	(−5.41)	(2.91)	(2.70)	(62.33)	(1.28)	(−7.41)	(−2.86)
High−Zero	3.67	0.03	−0.01	−0.65	−0.33	0.22	4.62	0.01	0.03	−0.62	−0.17
	(3.32)	(1.57)	(−0.15)	(−16.36)	(−7.76)	(3.50)	(3.31)	(0.54)	(0.73)	(−10.35)	(−3.86)
High−Low	1.56	−0.03	0.00	−0.19	−0.04	0.33	1.57	−0.05	0.05	0.07	0.01
	(1.36)	(−1.40)	(0.07)	(−4.59)	(−0.82)	(4.95)	(1.31)	(−2.17)	(1.57)	(1.26)	(0.34)
High−Q3	4.00	0.08	0.09	−0.24	−0.32	−0.19	3.84	0.08	0.15	−0.51	−0.13
	(3.73)	(3.59)	(2.73)	(−6.15)	(−7.79)	(−3.05)	(3.29)	(3.53)	(4.57)	(−10.05)	(−3.58)
	Panel A: Sorts on $R\&D/M$										
Zero	−1.73	1.01	0.00	0.33	0.17	−0.09	−2.32	1.02	0.00	0.34	0.10
	(−3.21)	(94.65)	(0.29)	(16.92)	(8.13)	(−2.83)	(−3.37)	(77.16)	(0.05)	(11.51)	(4.39)
Low	2.41	1.06	−0.05	−0.38	−0.12	−0.20	2.73	1.07	−0.06	−0.62	−0.04
	(2.75)	(61.07)	(−1.96)	(−12.03)	(−3.62)	(−3.98)	(2.70)	(54.78)	(−2.12)	(−14.13)	(−1.27)
Q2	0.07	0.97	0.00	−0.11	0.08	0.04	0.11	0.97	−0.01	−0.08	0.07
	(0.09)	(61.49)	(−0.13)	(−3.90)	(2.63)	(0.86)	(0.13)	(61.65)	(−0.50)	(−2.16)	(2.55)
Q3	−1.36	0.94	−0.01	−0.07	0.16	0.38	−0.58	0.92	−0.05	0.27	0.03
	(−1.93)	(66.99)	(−0.44)	(−2.81)	(5.93)	(9.21)	(−0.77)	(63.14)	(−2.27)	(8.18)	(1.24)
High	2.74	1.08	0.16	−0.09	−0.22	0.28	4.15	1.04	0.16	0.02	−0.25
	(2.77)	(55.08)	(5.27)	(−2.49)	(−5.88)	(4.91)	(4.12)	(53.65)	(5.70)	(0.49)	(−7.81)
High−Zero	4.47	0.06	0.15	−0.42	−0.39	0.37	6.47	0.02	0.16	−0.32	−0.35
	(3.49)	(2.52)	(3.93)	(−9.03)	(−7.94)	(4.97)	(4.60)	(0.72)	(4.05)	(−5.27)	(−7.73)
High−Low	0.32	0.02	0.21	0.29	−0.10	0.48	1.42	−0.03	0.22	0.64	−0.21
	(0.27)	(0.67)	(5.77)	(6.75)	(−2.19)	(6.96)	(1.09)	(−1.14)	(6.06)	(11.39)	(−5.06)
High−Q3	4.10	0.14	0.17	−0.02	−0.38	−0.10	4.73	0.12	0.21	−0.24	−0.28
	(3.39)	(5.77)	(4.56)	(−0.39)	(−8.26)	(−1.38)	(3.84)	(5.21)	(6.05)	(−4.61)	(−7.15)

Table 5: Amended Five-Factor Model Regressions for R&D Sorted Portfolios

We sort stocks at the end of each June based on their R&D to book equity ratio in Panel A and their R&D to market equity ratio in Panel B as in Table 3. R&D is calculated as described in the text. We industry adjust $R\&D/B$ and $R\&D/M$ using 2-digit SIC codes before sorting. Stocks with $R\&D/B = 0$ and $R\&D/M = 0$ and are placed in one group and the remaining stocks are sorted into quartiles based on breakpoints from NYSE stocks only. These quartiles are labeled Low, Q2, Q3, and High. The portfolios are value-weighted and rebalanced once every year at the end of June. The amended five-factor model replaces the RMW factor of Fama and French (2015) with RMW_{CP}, the cash-based operating profitability factor of Ball, Gerakos, Linnainmaa, and Nikolaev (2016)). All alphas are reported in annualized percent. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	Alpha	Mkt	SMB	HML	RMW_{CP}	CMA
	Panel A: Sorts on $R\&D/B$					
Zero	−0.10	0.99	−0.06	0.36	−0.10	−0.12
	(−0.17)	(86.83)	(−3.50)	(17.37)	(−3.31)	(−3.65)
Low	0.44	1.08	0.02	−0.15	−0.11	−0.19
	(0.47)	(60.49)	(0.70)	(−4.76)	(−2.23)	(−3.58)
Q2	2.28	0.99	−0.04	−0.18	0.11	−0.19
	(2.10)	(47.97)	(−1.33)	(−4.86)	(2.08)	(−3.13)
Q3	−1.29	0.96	−0.13	−0.04	0.02	0.29
	(−1.79)	(69.82)	(−6.48)	(−1.48)	(0.44)	(7.35)
High	0.33	1.06	0.07	−0.32	0.11	0.15
	(0.42)	(69.83)	(2.92)	(−11.62)	(2.62)	(3.37)
High−Zero	0.44	0.07	0.12	−0.68	0.21	0.27
	(0.36)	(3.13)	(3.70)	(−16.47)	(3.41)	(4.08)
High−Low	−0.11	−0.02	0.05	−0.17	0.21	0.33
	(−0.09)	(−0.78)	(1.42)	(−4.08)	(3.53)	(5.10)
High−Q3	1.62	0.11	0.20	−0.29	0.09	−0.14
	(1.38)	(4.76)	(5.96)	(−7.01)	(1.51)	(−2.21)
	Panel B: Sorts on $R\&D/M$					
Zero	−0.13	0.99	−0.06	0.35	−0.10	−0.11
	(−0.23)	(88.75)	(−3.66)	(17.00)	(−3.29)	(−3.45)
Low	0.83	1.08	0.01	−0.38	0.13	−0.18
	(0.91)	(61.82)	(0.20)	(−12.10)	(2.84)	(−3.64)
Q2	−0.44	0.97	0.00	−0.08	0.14	0.03
	(−0.53)	(61.84)	(−0.19)	(−2.72)	(3.29)	(0.62)
Q3	−0.36	0.92	−0.06	−0.04	−0.02	0.35
	(−0.47)	(64.09)	(−2.79)	(−1.67)	(−0.47)	(8.42)
High	1.41	1.09	0.22	−0.13	0.02	0.31
	(1.33)	(54.40)	(7.55)	(−3.49)	(0.30)	(5.33)
High−Zero	1.54	0.10	0.28	−0.48	0.11	0.42
	(1.09)	(3.77)	(7.21)	(−9.74)	(1.61)	(5.46)
High−Low	0.57	0.02	0.22	0.26	−0.12	0.50
	(0.46)	(0.74)	(6.25)	(5.92)	(−1.83)	(7.19)
High−Q3	1.77	0.17	0.28	−0.08	0.03	−0.04
	(1.32)	(6.71)	(7.57)	(−1.82)	(0.51)	(−0.56)

Table 6: Amended q Factor Model Regressions for R&D Sorted Portfolios

We sort stocks at the end of each June based on their R&D to book equity ratio in Panel A and their R&D to market equity ratio in Panel B as in Table 3. R&D is calculated as described in the text. We industry adjust $R\&D/B$ and $R\&D/M$ using 2-digit SIC codes before sorting. Stocks with $R\&D/B = 0$ and $R\&D/M = 0$ and are placed in one group and the remaining stocks are sorted into quartiles based on breakpoints from NYSE stocks only. These quartiles are labeled Low, Q2, Q3, and High. The portfolios are value-weighted and rebalanced once every year at the end of June. Various subpanels report alpha and loadings from various factor models. The amended q model replaces the ROE factor of Hou, Xue, and Zhang (2015) with RMW_{CP}, the cash-based operating profitability factor of Ball, Gerakos, Linnainmaa, and Nikolaev (2016)). All alphas are reported in annualized percent. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	Alpha	Mkt	ME	I/A	RMW_{CP}
	Panel A: Sorts on $R\&D/B$				
Zero	−0.29	1.00	−0.04	0.32	−0.18
	(−0.40)	(73.95)	(−2.25)	(10.51)	(−5.02)
Low	0.54	1.09	−0.01	−0.35	−0.07
	(0.56)	(60.82)	(−0.41)	(−8.62)	(−1.56)
Q2	2.50	0.99	−0.08	−0.39	0.15
	(2.29)	(48.71)	(−2.76)	(−8.45)	(2.81)
Q3	−1.25	0.95	−0.13	0.23	0.02
	(−1.72)	(70.02)	(−6.48)	(7.58)	(0.70)
High	0.40	1.05	0.07	−0.25	0.18
	(0.45)	(64.07)	(3.13)	(−6.69)	(4.25)
High−Zero	0.68	0.05	0.12	−0.57	0.36
	(0.48)	(1.94)	(3.08)	(−9.48)	(5.18)
High−Low	−0.15	−0.04	0.09	0.10	0.26
	(−0.12)	(−1.62)	(2.60)	(1.96)	(4.32)
High−Q3	1.65	0.10	0.20	−0.48	0.16
	(1.36)	(4.48)	(6.18)	(−9.43)	(2.67)
	Panel B: Sorts on $R\&D/M$				
Zero	−0.33	1.00	−0.05	0.32	−0.17
	(−0.47)	(76.21)	(−2.37)	(10.60)	(−4.96)
Low	0.75	1.09	−0.02	−0.59	0.23
	(0.73)	(56.76)	(−0.64)	(−13.44)	(4.66)
Q2	−0.49	0.97	0.00	−0.05	0.16
	(−0.59)	(62.41)	(−0.19)	(−1.52)	(3.91)
Q3	−0.22	0.91	−0.06	0.26	−0.02
	(−0.28)	(62.90)	(−2.71)	(7.98)	(−0.46)
High	2.12	1.06	0.22	0.02	0.00
	(1.94)	(52.19)	(7.39)	(0.48)	(0.06)
High−Zero	2.45	0.06	0.27	−0.29	0.17
	(1.61)	(2.21)	(6.43)	(−4.58)	(2.35)
High−Low	1.37	−0.02	0.24	0.61	−0.23
	(1.01)	(−0.85)	(6.45)	(10.56)	(−3.48)
High−Q3	2.34	0.15	0.28	−0.24	0.02
	(1.76)	(6.14)	(7.69)	(−4.29)	(0.32)

Table 7: q^5 Factor Model Regressions for R&D Sorted Portfolios

We sort stocks at the end of each June based on their R&D to book equity ratio in Panel A and their R&D to market equity ratio in Panel B as in Table 3. R&D is calculated as described in the text. We industry adjust $R\&D/B$ and $R\&D/M$ using 2-digit SIC codes before sorting. Stocks with $R\&D/B = 0$ and $R\&D/M = 0$ and are placed in one group and the remaining stocks are sorted into quartiles based on breakpoints from NYSE stocks only. These quartiles are labeled Low, Q2, Q3, and High. The portfolios are value-weighted and rebalanced once every year at the end of June. Various subpanels report alpha and loadings from various factor models. q^5 model is the factor model of Hou, Mo, Xue, and Zhang (2021) where the factors are Mkt, ME, I/A, ROE, and EG. All alphas are reported in annualized percent. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	Alpha	Mkt	ME	I/A	ROE	EG
	Panel A: Sorts on $R\&D/B$					
Zero	−0.81	1.00	−0.02	0.36	0.16	−0.19
	(−1.09)	(72.34)	(−0.87)	(12.03)	(6.40)	(−5.45)
Low	1.57	1.07	−0.03	−0.33	−0.05	−0.10
	(1.59)	(58.22)	(−1.23)	(−8.39)	(−1.57)	(−2.25)
Q2	2.65	1.00	−0.11	−0.43	−0.23	0.26
	(2.38)	(48.46)	(−3.67)	(−9.59)	(−6.04)	(4.94)
Q3	−1.42	0.95	−0.12	0.23	0.06	−0.01
	(−1.88)	(67.80)	(−5.98)	(7.61)	(2.29)	(−0.35)
High	0.36	1.06	0.06	−0.29	−0.17	0.25
	(0.40)	(63.46)	(2.38)	(−8.02)	(−5.52)	(5.93)
High−Zero	1.17	0.06	0.07	−0.65	−0.33	0.44
	(0.81)	(2.28)	(1.92)	(−11.14)	(−6.70)	(6.48)
High−Low	−1.21	−0.01	0.09	0.04	−0.12	0.36
	(−0.97)	(−0.53)	(2.69)	(0.89)	(−2.73)	(6.07)
High−Q3	1.78	0.11	0.18	−0.52	−0.23	0.26
	(1.44)	(4.68)	(5.40)	(−10.51)	(−5.42)	(4.54)
	Panel A: Sorts on $R\&D/M$					
Zero	−0.84	1.00	−0.02	0.35	0.16	−0.19
	(−1.17)	(74.82)	(−0.97)	(12.17)	(6.69)	(−5.59)
Low	1.78	1.08	−0.05	−0.63	−0.09	0.12
	(1.64)	(53.63)	(−1.66)	(−14.32)	(−2.30)	(2.39)
Q2	−1.01	0.98	0.00	−0.08	0.01	0.14
	(−1.16)	(60.92)	(0.14)	(−2.41)	(0.48)	(3.50)
Q3	−0.60	0.92	−0.05	0.27	0.03	0.00
	(−0.74)	(60.78)	(−2.21)	(8.15)	(1.04)	(0.07)
High	1.44	1.08	0.20	0.00	−0.38	0.35
	(1.39)	(55.85)	(7.14)	(0.03)	(−10.61)	(7.10)
High−Zero	2.28	0.08	0.22	−0.35	−0.54	0.54
	(1.59)	(2.86)	(5.65)	(−6.08)	(−11.03)	(7.94)
High−Low	−0.34	0.00	0.24	0.63	−0.29	0.22
	(−0.24)	(−0.18)	(6.65)	(11.24)	(−6.15)	(3.45)
High−Q3	2.05	0.16	0.24	−0.26	−0.40	0.34
	(1.59)	(6.69)	(7.14)	(−5.11)	(−9.20)	(5.68)

Table 8: Amended q^5 Model Regressions for R&D Sorted Portfolios

We sort stocks at the end of each June based on their R&D to book equity ratio in Panel A and their R&D to market equity ratio in Panel B as in Table 3. R&D is calculated as described in the text. We industry adjust $R\&D/B$ and $R\&D/M$ using 2-digit SIC codes before sorting. Stocks with $R\&D/B = 0$ and $R\&D/M = 0$ and are placed in one group and the remaining stocks are sorted into quartiles based on breakpoints from NYSE stocks only. These quartiles are labeled Low, Q2, Q3, and High. The portfolios are value-weighted and rebalanced once every year at the end of June. Various subpanels report alpha and loadings from various factor models. Amended q^5 model replaces the ROE factor of Hou, Mo, Xue, and Zhang (2021) with RMW_{CP}, the cash-based operating profitability factor of Ball, Gerakos, Linnainmaa, and Nikolaev (2016)). All alphas are reported in annualized percent. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	Alpha	Mkt	ME	I/A	RMW_{CP}	EG
	Panel A: Sorts on $R\&D/B$					
Zero	−0.37	1.00	−0.04	0.32	−0.19	0.01
	(−0.48)	(71.10)	(−2.15)	(10.27)	(−4.33)	(0.34)
Low	1.49	1.07	−0.03	−0.33	0.03	−0.16
	(1.49)	(58.04)	(−0.97)	(−7.99)	(0.49)	(−3.16)
Q2	2.27	1.00	−0.08	−0.40	0.13	0.04
	(1.98)	(46.97)	(−2.59)	(−8.43)	(1.93)	(0.68)
Q3	−1.40	0.95	−0.13	0.23	0.01	0.02
	(−1.83)	(67.42)	(−6.26)	(7.35)	(0.22)	(0.63)
High	0.00	1.06	0.08	−0.26	0.14	0.07
	(−0.00)	(62.03)	(3.34)	(−6.85)	(2.69)	(1.44)
High−Zero	0.36	0.06	0.12	−0.58	0.33	0.05
	(0.24)	(2.06)	(3.16)	(−9.45)	(3.87)	(0.72)
High−Low	−1.49	−0.01	0.11	0.07	0.11	0.22
	(−1.19)	(−0.59)	(3.22)	(1.32)	(1.58)	(3.57)
High−Q3	1.39	0.11	0.21	−0.49	0.13	0.04
	(1.10)	(4.49)	(6.20)	(−9.40)	(1.82)	(0.67)
	Panel A: Sorts on $R\&D/M$					
Zero	−0.41	1.00	−0.04	0.32	−0.18	0.01
	(−0.55)	(73.26)	(−2.27)	(10.37)	(−4.28)	(0.34)
Low	1.21	1.08	−0.03	−0.57	0.28	−0.08
	(1.12)	(54.19)	(−0.89)	(−12.99)	(4.66)	(−1.44)
Q2	−1.16	0.98	0.01	−0.07	0.09	0.11
	(−1.33)	(60.97)	(0.27)	(−1.96)	(1.79)	(2.55)
Q3	−0.50	0.92	−0.05	0.26	−0.05	0.05
	(−0.61)	(60.78)	(−2.47)	(7.66)	(−1.03)	(1.14)
High	1.34	1.08	0.23	0.00	−0.08	0.13
	(1.17)	(50.97)	(7.71)	(0.07)	(−1.23)	(2.26)
High−Zero	1.75	0.07	0.28	−0.31	0.10	0.11
	(1.10)	(2.54)	(6.59)	(−4.77)	(1.11)	(1.47)
High−Low	0.13	0.00	0.26	0.58	−0.36	0.20
	(0.09)	(−0.01)	(6.91)	(9.94)	(−4.54)	(2.92)
High−Q3	1.84	0.16	0.29	−0.25	−0.03	0.08
	(1.32)	(6.23)	(7.78)	(−4.43)	(−0.41)	(1.19)

Table 9: Profitability Loadings and Alphas from Factor Models on Portfolios Formed on Capitalized R&D Scaled by Market Equity

We sort stocks at the end of each June based on their capitalized R&D to market equity ratio. We capitalize R&D following the procedure in Park (2019). We use only the R&D as reported by Compustat or using supplementary information as described in the text. We industry adjust capitalized R&D to market equity ratio using 2-digit SIC codes before sorting. Stocks with zero capitalized R&D are placed in one group and the remaining stocks are sorted into quartiles based on breakpoints from NYSE stocks only. These quartiles are labeled Low, Q2, Q3, and High. The portfolios are value-weighted and rebalanced once every year at the end of June. Various subpanels report alpha and selected loadings from various factor models. The factors in five-factor model are Mkt, SMB, HML, RMW, and CMA. The amended five-factor model replaces the RMW factor with RMW_{CP}. The factors in q model are Mkt, ME, I/A, and ROE. The amended q model replaces the ROE factor with RMW_{CP}. All alphas are reported in annualized percent. We report loadings on only RMW, ROE, or RMW_{CP} factor, wherever applicable. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	Five-Factor		Amended Five-Factor		q		Amended q	
	Alpha	RMW	Alpha	RMW_{CP}	Alpha	ROE	Alpha	RMW_{CP}
Zero	−2.51	0.39	−0.68	0.05	−2.70	0.23	−0.92	0.01
	(−3.13)	(12.76)	(−0.71)	(0.95)	(−3.07)	(8.17)	(−0.96)	(0.18)
Low	1.26	−0.06	−0.21	0.17	1.20	0.03	−0.40	0.26
	(1.44)	(−1.63)	(−0.23)	(3.67)	(1.21)	(0.86)	(−0.40)	(5.25)
Q2	1.11	0.02	0.11	0.16	1.54	0.03	0.29	0.21
	(1.36)	(0.60)	(0.13)	(3.75)	(1.74)	(1.09)	(0.32)	(4.81)
Q3	−1.29	0.17	−0.18	−0.03	−0.43	0.02	0.01	−0.04
	(−1.87)	(6.29)	(−0.24)	(−0.77)	(−0.59)	(0.98)	(0.01)	(−0.98)
High	1.57	−0.11	1.02	−0.01	2.74	−0.17	1.57	−0.03
	(1.82)	(−3.29)	(1.12)	(−0.14)	(3.11)	(−6.14)	(1.68)	(−0.66)
High−Zero	4.07	−0.50	1.70	−0.05	5.44	−0.40	2.50	−0.04
	(3.34)	(−10.71)	(1.22)	(−0.74)	(4.26)	(−9.87)	(1.75)	(−0.56)
High−Low	0.31	−0.05	1.23	−0.18	1.54	−0.20	1.97	−0.29
	(0.24)	(−1.11)	(0.94)	(−2.65)	(1.12)	(−4.56)	(1.39)	(−4.12)
High−Q3	2.86	−0.28	1.20	0.02	3.17	−0.19	1.57	0.01
	(2.51)	(−6.29)	(0.98)	(0.36)	(2.73)	(−5.26)	(1.28)	(0.10)

Table 10: Alphas and Selected Loadings from Portfolio Sorted on Non-imputed R&D

We sort stocks at the end of each June based on their R&D to book equity ratio in Panel A and their R&D to market equity ratio in Panel B as in Table 3. R&D is calculated as described in the text, except that we do not replace missing values with imputed values. Stocks with missing values of R&D are placed in a separate portfolio labelled Missing. Various subpanels report alpha and selected loadings from various factor models. The factors in five-factor model are Mkt, SMB, HML, RMW, and CMA. The amended five-factor model replaces the RMW factor with RMW_{CP}. The factors in q model are Mkt, ME, I/A, and ROE. The amended q model replaces the ROE factor with RMW_{CP}. All alphas are reported in annualized percent. We report loadings on only RMW, ROE, or RMW_{CP} factor, wherever applicable. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

	Five-Factor		Amended Five-Factor		q		Amended q	
	Alpha	RMW	Alpha	RMW_{CP}	Alpha	ROE	Alpha	RMW_{CP}
	Panel A: Sorts on $R\&D/B$							
Missing	−1.35	0.06	0.01	−0.15	−2.06	0.03	−0.13	−0.24
	(−2.26)	(2.47)	(0.01)	(−4.87)	(−2.64)	(1.29)	(−0.16)	(−6.20)
Zero	−2.45	0.41	−0.65	0.06	−2.45	0.22	−0.84	0.02
	(−3.40)	(14.64)	(−0.73)	(1.31)	(−3.00)	(8.37)	(−0.94)	(0.43)
Low	−0.44	−0.02	0.10	−0.09	−0.52	0.03	0.00	−0.04
	(−0.45)	(−0.48)	(0.10)	(−1.79)	(−0.51)	(0.90)	(0.00)	(−0.80)
Q2	3.62	−0.10	2.52	0.08	4.39	−0.11	2.79	0.10
	(3.48)	(−2.56)	(2.30)	(1.43)	(4.17)	(−3.33)	(2.55)	(1.86)
Q3	−0.59	0.07	−0.56	0.05	−0.32	0.03	−0.63	0.08
	(−0.78)	(2.41)	(−0.72)	(1.29)	(−0.42)	(1.27)	(−0.81)	(2.04)
High	2.27	−0.21	0.44	0.10	2.81	−0.12	0.58	0.17
	(2.86)	(−6.72)	(0.52)	(2.35)	(3.14)	(−4.31)	(0.63)	(3.86)
High−Zero	4.72	−0.61	1.09	0.04	5.26	−0.34	1.42	0.16
	(3.82)	(−12.87)	(0.74)	(0.58)	(3.68)	(−7.48)	(0.92)	(2.07)
High−Low	2.71	−0.19	0.34	0.19	3.33	−0.15	0.58	0.22
	(2.14)	(−3.84)	(0.26)	(2.89)	(2.58)	(−3.70)	(0.44)	(3.30)
High−Q3	2.85	−0.27	1.01	0.05	3.13	−0.15	1.22	0.10
	(2.50)	(−6.24)	(0.82)	(0.81)	(2.65)	(−4.08)	(0.99)	(1.59)
	Panel A: Sorts on $R\&D/M$							
Missing	−1.38	0.06	−0.03	−0.15	−2.13	0.03	−0.16	−0.24
	(−2.25)	(2.52)	(−0.05)	(−4.63)	(−2.63)	(1.33)	(−0.20)	(−6.01)
Zero	−2.62	0.41	−0.87	0.07	−2.65	0.23	−1.06	0.03
	(−3.70)	(14.89)	(−1.00)	(1.52)	(−3.31)	(8.85)	(−1.20)	(0.72)
Low	1.25	−0.02	0.20	0.14	1.14	0.06	−0.09	0.24
	(1.37)	(−0.50)	(0.22)	(2.88)	(1.11)	(1.75)	(−0.09)	(4.67)
Q2	2.60	−0.12	0.58	0.20	2.75	−0.02	0.74	0.26
	(2.76)	(−3.19)	(0.60)	(4.04)	(2.70)	(−0.63)	(0.73)	(5.18)
Q3	−0.73	0.07	0.43	−0.11	0.16	−0.05	0.55	−0.11
	(−0.98)	(2.41)	(0.56)	(−2.91)	(0.21)	(−1.96)	(0.71)	(−2.85)
High	3.09	−0.27	0.97	0.09	4.47	−0.28	1.68	0.07
	(2.97)	(−6.81)	(0.86)	(1.62)	(4.24)	(−8.40)	(1.46)	(1.31)
High−Zero	5.71	−0.68	1.84	0.03	7.12	−0.51	2.74	0.04
	(4.19)	(−12.93)	(1.13)	(0.31)	(4.91)	(−11.00)	(1.66)	(0.53)
High−Low	1.84	−0.26	0.76	−0.05	3.33	−0.34	1.77	−0.16
	(1.38)	(−4.97)	(0.54)	(−0.63)	(2.38)	(−7.61)	(1.18)	(−2.22)
High−Q3	3.82	−0.34	0.53	0.21	4.30	−0.23	1.13	0.18
	(2.95)	(−6.84)	(0.38)	(2.91)	(3.27)	(−5.60)	(0.82)	(2.69)

Table 11: Alphas and Selected Loadings for Portfolios formed on 3-digit SIC code-imputed R&D

We sort stocks at the end of each June based on their R&D to book equity ratio in Panel A and their R&D to market equity ratio in Panel B as in Table 3. R&D is calculated as described in the text, except that we imputed missing values of R&D and industry adjust R&D using 3-digit SIC codes (instead of 2-digit SIC codes). Various subpanels report alpha and selected loadings from various factor models. The factors in the five-factor model are Mkt, SMB, HML, RMW, and CMA. The amended five-factor model replaces the RMW factor with RMW_{CP}. The factors in q model are Mkt, ME, I/A, and ROE. The amended q model replaces the ROE factor with RMW_{CP}. All alphas are reported in annualized percent. We report loadings on only RMW, ROE, or RMW_{CP} factor, wherever applicable. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021

	Five-Factor		Amended Five-Factor		q		Amended q	
	Alpha	RMW	Alpha	RMW_{CP}	Alpha	ROE	Alpha	RMW_{CP}
	Panel A: Sorts on $R\&D/B$							
Zero	−1.82	0.19	−0.22	−0.08	−2.37	0.10	−0.41	−0.16
	(−3.33)	(8.93)	(−0.36)	(−2.74)	(−3.38)	(4.64)	(−0.57)	(−4.50)
Low	−0.67	−0.06	0.17	−0.17	−0.37	−0.05	0.15	−0.14
	(−0.74)	(−1.82)	(0.19)	(−3.63)	(−0.40)	(−1.86)	(0.16)	(−2.91)
Q2	3.93	−0.28	1.89	0.08	4.96	−0.22	2.46	0.10
	(3.56)	(−6.54)	(1.58)	(1.26)	(4.44)	(−6.30)	(2.07)	(1.71)
Q3	−1.13	0.12	−0.34	−0.02	−0.71	0.04	−0.43	0.00
	(−1.58)	(4.28)	(−0.45)	(−0.56)	(−0.96)	(1.64)	(−0.56)	(0.10)
High	2.08	−0.13	0.37	0.14	2.40	−0.06	0.36	0.22
	(2.79)	(−4.68)	(0.48)	(3.55)	(2.83)	(−2.36)	(0.42)	(5.14)
High−Zero	3.90	−0.32	0.59	0.23	4.77	−0.17	0.77	0.38
	(3.54)	(−7.60)	(0.49)	(3.73)	(3.45)	(−3.79)	(0.55)	(5.45)
High−Low	2.75	−0.07	0.20	0.31	2.77	−0.01	0.20	0.35
	(2.38)	(−1.60)	(0.17)	(5.24)	(2.31)	(−0.23)	(0.17)	(6.01)
High−Q3	3.21	−0.25	0.71	0.16	3.11	−0.10	0.79	0.21
	(2.89)	(−5.91)	(0.60)	(2.70)	(2.65)	(−2.74)	(0.65)	(3.60)
	Panel A: Sorts on $R\&D/M$							
Zero	−1.77	0.18	−0.19	−0.09	−2.31	0.10	−0.38	−0.16
	(−3.30)	(8.55)	(−0.33)	(−2.97)	(−3.40)	(4.57)	(−0.55)	(−4.67)
Low	1.39	−0.09	0.05	0.12	1.58	−0.01	−0.07	0.22
	(1.47)	(−2.46)	(0.05)	(2.47)	(1.49)	(−0.26)	(−0.06)	(4.24)
Q2	1.90	−0.02	0.63	0.17	2.00	0.01	0.59	0.21
	(2.25)	(−0.58)	(0.72)	(3.84)	(2.27)	(0.23)	(0.66)	(4.74)
Q3	−0.88	0.12	0.00	−0.04	−0.15	0.01	0.15	−0.03
	(−1.27)	(4.41)	(0.01)	(−0.94)	(−0.20)	(0.34)	(0.20)	(−0.90)
High	2.85	−0.23	1.21	0.05	4.20	−0.24	1.95	0.04
	(2.93)	(−6.18)	(1.16)	(1.03)	(4.20)	(−7.56)	(1.80)	(0.85)
High−Zero	4.62	−0.41	1.40	0.14	6.51	−0.34	2.34	0.20
	(3.61)	(−8.27)	(0.99)	(2.01)	(4.60)	(−7.54)	(1.54)	(2.75)
High−Low	1.46	−0.14	1.16	−0.07	2.62	−0.23	2.02	−0.18
	(1.16)	(−2.93)	(0.88)	(−1.03)	(1.96)	(−5.44)	(1.43)	(−2.58)
High−Q3	3.73	−0.35	1.21	0.09	4.35	−0.25	1.80	0.08
	(3.23)	(−7.85)	(0.95)	(1.40)	(3.69)	(−6.63)	(1.43)	(1.27)

Table 12: Attributes, Characteristics, Alphas, and Selected Loadings on Portfolios Sorted by Size and R&D

We sort stocks at the end of each June based on their size and R&D to book equity ratio in Panels A/C and their size and R&D to market equity ratio in Panels B/D as in Table 3. R&D is calculated as described in the text. In each panel, we first sort stocks into two groups based on size (with median breakpoints from NYSE only), and then sort stocks within each group based on R&D. Panels A and B report characteristics of these double-sorted portfolios as that for single-sorted portfolios in Table 3. Panels C and D report alpha and selected loadings from various factor models. The factors in five-factor model are Mkt, SMB, HML, RMW, and CMA. The amended five-factor model replaces the RMW factor with $\mathrm{RMW_{CP}}$. The factors in q model are Mkt, ME, I/A, and ROE. The amended q model replaces the ROE factor with $\mathrm{RMW_{CP}}$. All alphas are reported in annualized percent. We report loadings on only RMW, ROE, or $\mathrm{RMW_{CP}}$ factor, wherever applicable. t-statistics are reported in parenthesis below alphas/loadings. The sample includes all stocks with positive sales at portfolio formation. The sample period is 1975 to 2021.

		#stocks	$\%ME$	$\%R\&D$	B/M	dA/A	OP/B	CP/B	AC/B	$R1Y$
		Panel A: Sorts on $R\&D/B$								
Small	Zero	1,528	4.8	0.0	0.787	0.087	0.276	0.286	−0.082	1.8
	Low	756	2.2	4.0	0.644	0.087	0.266	0.230	−0.049	−0.1
	High	987	2.6	18.0	0.537	0.063	0.324	0.299	−0.075	−1.7
Large	Zero	438	35.7	0.0	0.614	0.100	0.342	0.308	−0.068	9.7
	Low	210	23.4	23.3	0.477	0.104	0.384	0.362	−0.086	10.5
	High	213	31.2	54.3	0.413	0.074	0.494	0.473	−0.114	9.7
		Panel B: Sorts on $R\&D/M$								
Small	Zero	1,564	4.8	0.0	0.764	0.082	0.291	0.302	−0.104	1.2
	Low	836	2.7	3.7	0.483	0.094	0.298	0.257	−0.049	7.3
	High	971	2.1	18.7	0.681	0.046	0.333	0.313	−0.092	−10.1
Large	Zero	455	36.0	0.0	0.603	0.100	0.348	0.315	−0.068	9.8
	Low	237	27.4	27.0	0.335	0.112	0.452	0.427	−0.083	14.3
	High	201	27.1	50.3	0.550	0.070	0.449	0.432	−0.121	5.9

		Five-Factor		Amended Five-Factor		q		Amended q	
		Alpha	RMW	Alpha	RMW_{CP}	Alpha	ROE	Alpha	RMW_{CP}
		Panel C: Sorts on $R\&D/B$							
Small	Zero	−1.93	0.32	−0.16	−0.01	−1.85	0.07	0.10	−0.19
		(−4.05)	(17.37)	(−0.25)	(−0.17)	(−2.20)	(2.68)	(0.12)	(−4.63)
	Low	−0.08	−0.09	0.04	−0.09	0.65	−0.17	0.38	−0.16
		(−0.15)	(−4.34)	(0.06)	(−3.04)	(1.07)	(−9.02)	(0.58)	(−4.96)
	High	1.40	−0.37	−0.04	−0.08	2.50	−0.35	0.25	−0.08
		(2.32)	(−16.02)	(−0.06)	(−2.21)	(3.05)	(−13.59)	(0.25)	(−1.75)
	High−Zero	3.33	−0.69	0.11	−0.08	4.34	−0.42	0.15	0.11
		(3.44)	(−18.56)	(0.09)	(−1.22)	(2.94)	(−9.06)	(0.09)	(1.39)
Large	Zero	−1.72	0.15	−0.15	−0.11	−2.34	0.09	−0.39	−0.18
		(−2.93)	(6.49)	(−0.25)	(−3.48)	(−3.18)	(3.75)	(−0.52)	(−4.77)
	Low	2.34	−0.09	1.31	0.08	2.85	−0.07	1.48	0.11
		(3.33)	(−3.20)	(1.78)	(2.16)	(4.00)	(−3.09)	(2.01)	(3.17)
	High	0.59	−0.04	−0.17	0.08	0.99	−0.02	−0.11	0.14
		(1.07)	(−1.67)	(−0.30)	(2.81)	(1.61)	(−0.84)	(−0.18)	(4.49)
	High−Zero	2.31	−0.18	−0.01	0.19	3.33	−0.10	0.27	0.31
		(2.33)	(−4.77)	(−0.01)	(3.64)	(2.75)	(−2.71)	(0.22)	(5.21)
		Panel D: Sorts on $R\&D/M$							
Small	Zero	−1.92	0.31	−0.09	−0.02	−1.76	0.06	0.18	−0.21
		(−4.04)	(16.93)	(−0.14)	(−0.69)	(−2.10)	(2.12)	(0.21)	(−5.05)
	Low	−0.69	−0.16	−0.63	−0.13	−0.14	−0.18	−0.49	−0.16
		(−1.42)	(−8.49)	(−1.20)	(−4.99)	(−0.23)	(−9.10)	(−0.72)	(−4.72)
	High	2.70	−0.41	0.86	−0.06	4.18	−0.43	1.44	−0.10
		(3.95)	(−15.74)	(1.00)	(−1.37)	(5.29)	(−16.94)	(1.44)	(−1.96)
	High−Zero	4.63	−0.72	0.95	−0.04	5.94	−0.48	1.26	0.11
		(4.35)	(−17.69)	(0.68)	(−0.54)	(4.08)	(−10.42)	(0.77)	(1.42)
Large	Zero	−1.71	0.14	−0.19	−0.11	−2.36	0.09	−0.44	−0.17
		(−2.98)	(6.53)	(−0.31)	(−3.42)	(−3.32)	(4.06)	(−0.60)	(−4.66)
	Low	2.42	−0.08	0.74	0.18	2.49	0.01	0.68	0.26
		(3.82)	(−3.20)	(1.14)	(5.49)	(3.39)	(0.29)	(0.94)	(7.45)
	High	−0.34	0.10	0.23	0.00	0.59	−0.02	0.51	−0.01
		(−0.58)	(4.48)	(0.36)	(−0.07)	(0.94)	(−0.93)	(0.78)	(−0.28)
	High−Zero	1.37	−0.04	0.41	0.10	2.96	−0.11	0.95	0.16
		(1.44)	(−1.17)	(0.42)	(2.09)	(2.81)	(−3.31)	(0.87)	(2.97)